ECOLOGICAL GENOMICS OF NEMATODE RESPONSES TO DIFFERENT BACTERIAL ENVIRONMENTS

by

JOSEPH COOLON

B.S., Kansas State University, 2003

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2008

Abstract

Determining the genetic mechanisms involved in organismal response to environmental change is essential for understanding the effects of anthropogenic disturbance. The composition of the bacterial-feeding nematode community is an excellent biological indicator of disturbance, particularly in grassland ecosystems. We have previously shown that grassland soil nematodes are responsive to perturbations in the field including the addition of nitrogen fertilizer. We are interested in how this perturbation affects the microbial community and downstream effects on the next trophic level, the bacterial-feeding nematodes. To determine the effects of disturbance on soil bacterial communities we used massively parallel sequencing and found that chronic nitrogen addition on tallgrass prairie significantly impacts overall bacterial community diversity and the abundance of specific bacterial taxa. Because native soil nematodes lack well developed genomic tools, we employed *Caenorhabditis elegans* as a model for native soil nematode taxa and used transcriptional profiling to identify 204 candidate genes regulated in response to altered bacterial diets isolated from grassland soils. To biologically validate our results we used mutations that inactivate 21 of the identified genes and showed that most contribute to fitness or lifespan in a given bacterial environment. Although these bacteria may not be natural *C. elegans* food sources, this study aimed to show how changes in food source, as can occur in environmental disturbance, has large effects on gene expression and those genes whose expression are affected, contribute to fitness. Furthermore, we identified new functions for genes of unknown function as well as previously well-characterized genes, demonstrating the utility of this approach to further describe *C. elegans* genome. We also investigated the function of previously well-characterized *C. elegans* defense pathways in our grassland soil bacterial environments and found that some are environment specific. Additionally, we found that cuticular collagen genes are important for lifespan, and appear to function downstream of known defense pathways. Overall, our results suggest that anthropogenic disturbance in grasslands alters the most basal components of the soil food web, bacteria and bacterial-feeding nematodes through the genes they possess and how they are expressed, and resultant bottom-up effects could have profound consequences on ecosystem health and function.

ECOLOGICAL GENOMICS OF NEMATODE RESPONSES TO DIFFERENT BACTERIAL ENVIRONMENTS

by

JOSEPH COOLON

B.S., Kansas State University, 2003

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2008

Approved by:

Major Professor
Dr. Michael A. Herman

Copyright

Abstract

Determining the genetic mechanisms involved in organismal response to environmental change is essential for understanding the effects of anthropogenic disturbance. The composition of the bacterial-feeding nematode community is an excellent biological indicator of disturbance, particularly in grassland ecosystems. We have previously shown that grassland soil nematodes are responsive to perturbations in the field including the addition of nitrogen fertilizer. We are interested in how this perturbation affects the microbial community and downstream effects on the next trophic level, the bacterial-feeding nematodes. To determine the effects of disturbance on soil bacterial communities we used massively parallel sequencing and found that chronic nitrogen addition on tallgrass prairie significantly impacts overall bacterial community diversity and the abundance of specific bacterial taxa. Because native soil nematodes lack well developed genomic tools, we employed *Caenorhabditis elegans* as a model for native soil nematode taxa and used transcriptional profiling to identify 204 candidate genes regulated in response to altered bacterial diets isolated from grassland soils. To biologically validate our results we used mutations that inactivate 21 of the identified genes and showed that most contribute to fitness or lifespan in a given bacterial environment. Although these bacteria may not be natural *C. elegans* food sources, this study aimed to show how changes in food source, as can occur in environmental disturbance, has large effects on gene expression and those genes whose expression are affected, contribute to fitness. Furthermore, we identified new functions for genes of unknown function as well as previously well-characterized genes, demonstrating the utility of this approach to further describe *C. elegans* genome. We also investigated the function of previously well-characterized *C. elegans* defense pathways in our grassland soil bacterial environments and found that some are environment specific. Additionally, we found that cuticular collagen genes are important for lifespan, and appear to function downstream of known defense pathways. Overall, our results suggest that anthropogenic disturbance in grasslands alters the most basal components of the soil food web, bacteria and bacterial-feeding nematodes through the genes they possess and how they are expressed, and resultant bottom-up effects could have profound consequences on ecosystem health and function.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to take this opportunity to thank my advisor Mike Herman who has always pushed me to do my best. His willingness to take on a guy that was a bit rough around the edges in the beginning has changed my life. I will never forget his words that have most inspired me "just sit down and be brilliant" that I'm sure will go through my head in the future when I have students.

I would also like to thank my supervisory committee: John Blair, Mark Ungerer, and Karen Garrett for their comments and ideas to make my project as good as possible. Also, thanks to Tim Todd, Ted Morgan, Loretta Johnson and Brett Sandercock for countless discussions and help. Especially thanks to Ken Jones for everything he did and contributed as I developed as a scientist.

My experience at K-State would not have been complete without the members of the Clem, Hancock, Hirt, Brown and Morgan labs. My countless discussions with Bart Bryant, Lance Thurlow and Sherry Miller have kept me sane and always cheer me up. Thanks to the faculty, staff and students of the Division of Biology for endless support. Teaching would not have been the same without Robbie Bear. Thanks to the Konza Prairie LTER for use of their site.

Dedication

I would like to dedicate this dissertation to a few people in different areas of my life. First, without the inspiration of my seventh and eigth grade science teacher Carol Williamson I would not have entered the sciences as a career. Carol was the best teacher of my life and showed me that science was more that just a job, but that it was a way of thinking about the world that surrounds us. She showed me that science was fun and interesting through her enthusiasm and love for science, and to her I am forever greatful.

Secondly, without the endless support of family and friends I don't know how long I would have made it in graduate school. I especially need to mention my parents Jim and Lyn Coolon, for letting me dream, always supporting my every decision and my siblings, Matt and KT Coolon, for their love and support.

Most importantly, I dedicate this to my wife. Rosemary, your patience, support and endless efforts to get me to relax have kept me going and I just wanted to take this chance to thank you for all you have done for me and continue to do for me daily.

CHAPTER 1 - *Caenorhabditis elegans* as a model system

Introduction

Caenorhabditis elegans as a system

Caenorhabditis elegans is one of the best-studied genetic model organisms. Fourty years ago Sydney Brenner selected *C. elegans* and developed it as a genetic system to facilitate genetic analysis of development of the nervous system (Brenner 1974, Wood 1988). *Caenorhabditis elegans* feeds primarily on microbes, and in particular bacteria, but can also be cultured axenically on defined chemical media in liquid culture (Szewczyk et al. 2003). Since its isolation it has been maintained on agar plates and fed almost exclusively *Escherichia coli* strain OP50 in the laboratory. Under ideal conditions embryonic development occurs during a 12 hour period within the egg, proceeds through 4 larval developmental stages (L1-L4) with each larval stage separated by molting of the outer cuticle and and leads to the formation of a reproductive adult. In times of stress or caloric restriction *C. elegans* will proceed from the second larval stage (L2) to an alternative developmental life stage, the dauer larvae. Dauer larvae are resistant to desiccation and freezing and it is likely this diapause stage facilitates an over-wintering strategy in times of stress in the wild. Because dauer larvae are resistant to very low temperatures, *C. elegans* can be frozen almost indefinitely in liquid nitrogen which allows for maintenance of large numbers of mutant stocks with minor manipulation.

Caenorhabditis elegans typically exists as a self-fertile hermaphrodite. Males arise at a very low rate as a result of meiotic non-disjunction and can mate with hermaphrodites. Hermaphrodites will typically produce ~300 self-progeny when grown under standard laboratory conditions (20°C, on nematode growth media (NGM) feeding on *E. coli* OP50) (See Chapter 3) and around 1,000 progeny when mated to a male. Development from a fertilized embryo to a reproductive adult occurs over a period of 3.5 days at 20°C. Hermaphrodites are comprised of 959 somatic nuclei (males have 1,031) when fully grown and development is characterized by invariant cell lineages. A fate map of every cell division in *C. elegans* was constructed which in part led to a Nobel Prize in 2002 (Sulston et al. 1983). Lifespan of wild type *C. elegans* is typically two weeks in duration with around 50% dying by six days due to the pathogenic effects of its food source (Tenor and Aballay 2008) (See Chapter 3).

Caenorhabditis elegans has received much attention for the properties that make it an excellent model organism. These characteristics include a fast generation time, small size (~1mm), ease of culture, transparent body, and small genome (Wood 1988). The *C. elegans* genome consists of approximately 20,000 genes with around 15% of genes transcribed as operons (Blumenthal et al. 2002). This means some genes are co-transcribed as a single large messenger RNA comprised of multiple genes. *C. elegans* was the first completely sequenced animal genome (and second eukaryote) in 1998 (Consortium 1998). Due to the efforts of the *C. elegans* community, thousands of genes have been discovered and their functions determined. *Caenorhabditis elegans* has been firmly

established as a powerful model in the post-genomic era through the development of genetic and genomic tools including gene chips that can be used to look at the activity of all the genes in its genome simultaneously.

RNA mediated interference (RNAi) was discovered in *C. elegans* (Fire et al. 1998), and occurs when double stranded RNA is administered through injection into the gonad (as well as by feeding and soaking with dsRNA). Using this technique, genes can be disabled through targeted knockdown of transcript levels. This has facilitated many reverse genetic analyses, which can be used to target most of the genes in the genome. Through strength of the research community, *C. elegans* has quickly become one of the most dominant systems used in many fields including molecular biology, developmental genetics and neurobiology, and recently as a model for human innate immunity.

CHAPTER 2 - Land use impacts on soil bacterial communities of the tallgrass prairie ecosystem

Abstract

Land use change is an important aspect of global change in grasslands of the Central Plains. Management of native grasslands and conversion of grasslands to agroecosystems can alter soil microbial community composition and structure, resulting in decreased diversity and/or disruption of biogeochemical processes. However, due to the high diversity of soil microbes, very little is known about specific taxonomic changes underlying microbial community-level responses. We used massively parallel sequencing coupled with unique barcoded primer constructs to assess the effects of selected land-use practices on soil bacterial communities. The data show that chronic nitrogen enrichment of native tallgrass prairie, conversion of prairie to agriculture, as well as changes in tillage practices, significantly impact overall community diversity. Other than the loss of rare species, however, it appears that community composition is remarkably resistant to the disturbance. This resistance may be a function of spatial heterogeneity or could simply be that land use change consists of multiple drivers that, while acting to change ecosystem function, can mask the effects on specific taxa. To our knowledge, this study is the first of its kind to use massively parallel sequencing to look at treatment effects in replicated field studies and determine effects on microbial community diversity at multiple taxonomic levels, as well as to identify individual taxon responses to disturbance.

Introduction

Understanding the effects of anthropogenic disturbance on natural systems is crucial for predicting the long-term consequences of human induced changes in the Earth's ecosystems. Determining the responses of microbial communities is of prime interest as they are largely responsible for energy and nutrient transformations, and are the most basal components of soil and aquatic food webs exerting widespread downstream effects on ecosystems. Recent studies have shown that microbial communities are sensitive to land use in various ecosystem types. For example, wholesale land use conversions, such as deforestation (Baath et al. 1995, Pietikainen and Fritze 1995, Holmes and Zak 1999) or tilling for agriculture (Frey et al. 1999, Allison et al. 2005), cause shifts in the bacterial community resulting in decreased overall diversity and/or disruption of biogeochemical processes, ultimately leading to alterations in ecosystem function. Even seemingly less intrusive soil perturbations such as burning of aboveground biomass (Vazquez et al. 1993, Baath et al. 1995, Pietikainen and Fritze 1995) or nitrogen amendments (Sarathchandra et al. 2001, Marschner et al. 2003, Bittman et al. 2005) can cause shifts in the microbial community. In the tallgrass prairie, where the predominant historic land use change was conversion of grasslands to agriculture (Samson and Knopf 1994a), disturbances such as tilling (Groffman et al. 1993, Sotomayor and Rice 1996), burning (Groffman et al. 1993, Ajwa et al. 1999, Dell et al. 2005), fertilization (Dell and Rice 2005) and irrigation (Williams and Rice 2007) elicit biogeochemical changes in the soil, presumably due to changes in microbial community composition or activity. Given that grassland plant communities (Blair 1997, Turner et al. 1997, Collins et al. 1998, Knapp et al. 2002), vertebrates (Shochat et al. 2005, Patten et al. 2006) and invertebrates (Blair et al. 1996, Todd 1996, Todd et al. 1999, Blair et al.

2000a, Callaham et al. 2003, Jones et al. 2006b) in native tallgrass prairie have differential responses to prairie perturbations, it is reasonable to assume that the tallgrass prairie soil microbial community would also respond to these changes. Previous research in the tallgrass prairie has inferred changes in the soil microbial community based on direct assessments of biogeochemical changes or through changes in frequency of microbial biomarkers in the soil (Groffman et al. 1993, Sotomayor and Rice 1996, Ajwa et al. 1999, Dell and Rice 2005, Dell et al. 2005, Williams and Rice 2007). However, very little is known about the changes in the microbial community that produced these responses. Moreover, studies of microbial community change are often limited in scope and replication, as surveying the thousands of bacterial species in a natural soil environment has been impossible until recent advances in sequencing technology.

Here we have taken advantage of 454 massively parallel sequencing (MPS) technologies to sequence bacterial 16S DNA amplified from the soil of selected ongoing ecological studies on the Konza Prairie Long-Term Ecological Research (LTER) site. By using barcoded primers and a new method to assign Operational Taxonomic Units (OTUs), we determined frequency of occurrence of thousands of bacterial taxa across multiple replicated treatments and field experiments simultaneously. To our knowledge, this study is the first of its kind to use mass parallel sequencing to look at treatment effects in replicated field studies and determine effects on microbial community diversity at multiple taxonomic levels, as well as to identify individual taxon responses to disturbance.

Methods

Study design, soil sampling and DNA extraction

In coordination with the Konza Prairie LTER Program and the Konza Prairie Biological Station, a 3487-hectare area of native tallgrass prairie in the Flint Hills of northeastern Kansas, we utilized three existing field sites. The first site consists of eight large plots (50 m x 25 m) subdivided into a total of 64, 12.5 m x 12.5 m subplots. Treatments were arranged in a split-strip plot design with four replications per treatment combination. Burning treatment was assigned at the whole plot level and mowing and fertilization (nitrogen and phosphorus) additions were imposed as strip-plot treatments. Burning was performed in April of each year and fertilizer treatments were applied in late May to early June. In the carbon sequestration experiment, each of the four replicates (6 m x 6 m plots) per treatment were arranged such that treatments had four plots in a row to facilitate tillage and yearly sorghum planting. For the natural prairie to agriculture experiment, each of three replicates, a 3 m x 3 m grid was placed 50 m either side of the old field boundary, within the old field and the adjacent prairie, respectively. Within these six plots, soil cores were taken from the nine 1 m^2 sub-plots.

For each of the 94 plots, two 2.5 cm diameter 10 cm deep cores were collected, pooled, and homogenized. From each composite sample, genomic DNA was extracted from a 10 g subsample using a standard soil DNA extraction kit (UltraClean Mega Soil DNA Kit, MoBio, Carlsbad, CA). After extraction, the DNA was diluted to ~5ng/_l and stored at -80Y C.

16S rDNA amplification for massively parallel sequencing

PCR reactions [1x Amplitaq Gold polymerase reaction buffer (Applied Biosystems, Foster City, CA), 3.75 mM $MgCl_2$, 200 μM dNTP, 0.5 μM each forward and reverse $_b$MPS 16S primers, one unit Amplitaq Gold LD polymerase, and 5ng soil extracted DNA] were run for 25 cycles at 95°C for 1 minute, 55°C for 1 minute, 74°C for 1 minute on an iCycler IQ real-time thermocycler (Bio-Rad Laboratories, Hercules, CA). PCR reactions were spiked with SYBR Green and run on a realtime thermocycler to assure that the PCR was stopped during log-phase, a necessity to assure accurate quantitation during sequencing. PCR amplification was done in triplicate, individual reactions pooled, and cleaned using an AMPure PCR cleanup kit (Agencourt Bioscience, Beverly, MA). We utilized a single MPS run with four blocks to reduce the number of unique primer combinations needed. We randomly assigned 128 DNA samples (94 from the current study and 34 samples available for additional studies) to one of 32 different $_b$MPS primers in one of four sequencing blocks and amplified the V3 region of the 16S rRNA gene independently using the appropriate $_b$MPS primers. For each of the four sequencing blocks, 100 ng of each differentially barcoded PCR product were pooled into respective sequencing pools and sequenced by 454 Life Sciences (Branford, CT).

Bioinformatics and OTU designation

Raw 454 sequences were searched for the occurrence of the primer barcode immediately preceding the 16S primer sequence within each sequence. Where found, the barcode was removed and the corresponding plot designation was incorporated into the sequence name. Sequences were removed if they did not contain a valid primer or

barcode sequence, contained more than one ambiguous base, or were shorter than 85 bp or longer than 125 bp. For the remaining sequences CAP3 (Huang and Madan 1999) was used to align sequences at each of 18 sequence identity levels (80% to 98%) using a minimal overlap of 75 bp.

Rarefaction curves

We calculated OTU accumulation curves (i.e., rarefaction curves) at the extremes of our sequence sampling range (80% and 98% sequence identity, respectively) for our main treatment effects. Within each experiment, OTU designations from the CAP3 results were pooled for each treatment. Within treatments, the pooled collection of sequences was randomly re-sampled at increasing intervals of 100 (e.g., 100 sequences, 200 sequences, etc.) and the number of OTUs identified within each subsample noted. Rarefaction curves were made by fitting a line through three replicate samplings at each of the sampling intervals.

Diversity indices

Overall taxonomic richness (S) was calculated by summing the number of OTUs, including singlets, which occurred within each plot. Simpson's Dominance ($\sum p_i^2$), Simpson's Diversity ($1/\sum p_i^2$), and Shannon's diversity ($e\sum p_i(\ln(p_i))$ were calculated for each plot, where p_i is the frequency of occurrence of each OTU. Evenness was calculated as the ratio of Shannon's Diversity and richness ($e\sum p_i(\ln(p_i)/S$). A final index of diversity, Fisher's alpha log-series (Fisher et al. 1943) was calculated by iterating the equation $S/N=[(1-x)/x][-\ln(1-x)]$, where S is richness and N is the total number of

sequences within the plot. Once x was solved, the diversity index alpha (α) was calculated as N(1-x)/x. Differences in diversity across treatments were analyzed using the MIXED procedure in SAS (SAS Institute Inc., Cary, NC).

OTU frequency

Using 98% sequence identity, a student's t-test was used to identify OTUs whose frequencies of occurrence were either too rare or too variable to provide accurate inferences on their distribution. Where OTU abundances were significantly greater than zero ($P < 0.05$), either across the entirety of an experiment or within specific treatments, the REG procedure in SAS (SAS Institute Inc., Cary, NC) was used to determine the overall effect of treatment across OTUs. A $\log_{10}$ transformation was performed on OTU response ($freq_{disturbed}$-$freq_{undistrubed}$) to reduce heterogeneity of variances. Additionally, the UNIVARIATE procedure is SAS, was used to analyze deviations from normality and to derive 1% and 5% quantiles. Where OTU responses fell within the upper or lower 5% or 1% quantile, the collection of sequences within each OTU were used to produce a consensus sequence. The taxonomic identities of those consenses were then explored via BLAST to the Ribosomal Database Project (http://rdp.cme.msu.edu/).

Results

Study design

We sampled soil bacterial communities in three separate experiments at the Konza Prairie Biological Station (KPBS) that model various disturbances associated with land-use conversion. The first was initiated in 1987 to assess above- and below-ground

responses of native tallgrass prairie communities to fire and nutrient addition treatments. The treatments examined were: burning (annually burned vs. unburned); ammonium nitrate addition (10g N/m^2 annually vs. no addition); and superphosphate addition (1g P/m^2 annually vs. no addition; see Methods for details). Thirty-two plots consisting of the factorial treatments of burning, nitrogen, and phosphorus additions were sampled (n = 4 per treatment combination). The second experiment was established in 2004 in an agricultural field that had been predominately used for tilled wheat (*Triticum aestivum*) production for more than 20 years. Two treatments were chosen to represent management practices that have historically resulted in different belowground carbon storage levels: (1) continuous grain sorghum (*Sorghum bicolor*) under no-tillage; and (2) continuous grain sorghum under conventional tillage. Lastly, we sampled plots in an agricultural field adjacent to native tallgrass prairie, both of which are on similar soils. This agricultural field has been in cultivation for more than 50 years since its conversion from native tallgrass prairie. The effects of historic land use conversion (native prairie vs. agriculture) were analyzed using three pairs of replicate plots (See Methods for details).

Sequencing, Bioinformatics, and OTU frequency

From each of the 94 plots sampled, DNA extracts of soil bacterial communities were amplified using universal bacterial primers (U341F and U533R, Watanabe et al. 2001) designed to amplify the third variable region (V3) of the 16S rRNA gene. These primer constructs (Table 2.1) also directly incorporated the 454 sequencing primers (Margulies et al. 2005) and a unique 5-bp sequence of DNA between the sequencing primer and the reverse 16S primer. These barcoded primer constructs ($_b$MPS primers)

were specifically designed to allow randomized sequencing from a mixture of multiple PCR products. From the 263,593 raw sequences generated, 197,608 sequences (85-125bp) were available after removing aberrant, short and/or long sequences. Of the remaining sequences, 147,924 sequences were derived from the 94 DNA samples used in this study and 49,684 sequences were generated for additional studies. A sequence assembly program, CAP3 (Huang and Madan 1999), was used to align the 147,924 sequences and assign operational taxonomic units (OTUs) at each of 18 sequence identity levels (80% to 98%). At each level of sequence identity, sequences were parsed by plot and used to calculate the frequency of occurrence of all OTUs for each of the 94 plots. There was a noted increase in the number of OTUs as the percent sequence identity increased from 80-98%, (Figure 2.1). This indicates that assembling sequences using CAP3 produced OTUs that follow expectations of biological complexity, with OTUs generated at different levels of sequence identity being of different taxonomic resolutions. To further assess the quality of the data and to determine whether the sampling of the microbial biomass was sufficient, we calculated rarefaction curves for the cumulative plots within each treatment. Figure 2.2 illustrates that OTUs generated at 80% sequence identity reached saturation indicating appropriate sampling intensity at lower sequence identity levels, while those generated at 98% sequence did not reach saturation.

Changes in community structure and overall diversity

From the OTU frequencies generated at each sequence identity level (80-98%, respectively), overall taxonomic richness, dominance, diversity, and evenness were calculated for each plot and analyzed across treatments using analysis of variance in SAS

statistical software (SAS Statistical Institute Inc., Cary, North Carolina). Significant changes in the overall microbial community structure due to treatment effects were observed in all three experiments (Table 2.2). Of the main effects, burning and phosphorous addition had only minor effects on microbial diversity and were removed from further analyses (data not shown). For the remainder of the treatments, ANOVA using OTUs assembled iteratively across the range of 80-98% sequence identities enabled assessment of the overall significance of the change in diversity across treatments, as well as the interaction of treatments and levels of sequence identity. For most diversity indices and treatments, a significant treatment effect occurred over most sequence identity levels, and there were significant interactions between treatment and sequence identity level (Table 2.2). A significant main effect was interpreted to mean that the treatment had similar effects on bacterial taxa no matter what sequence identity level was used to cluster the sequences, while a significant interaction between main effect and sequence identity level indicates that the treatment effect changed when OTUs became more taxonomically inclusive or exclusive (Table 2.2, Figures 2.3, 2.4).

To better illustrate the effect of treatments on soil bacterial diversity, means of various community indices and associated *p*-values were plotted (Figures 2.3, 2.4) at each of the 18 sequence identity levels investigated (80%-98%). The general shapes of the curves (e.g., increasing diversity, richness, and evenness and decreasing dominance as sequence identity increases) indicate that varying sequence identity level can provide insight into how different levels of biological organization respond to disturbance. Interestingly, while the ANOVA can assess the change in diversity across varying levels of sequence identity (Table 2.2), the diversity graphs define at what level of sequence

identity shifts in the community structure occur. For example, the effect of nitrogen enrichment on richness was consistently significant (dotted line, $p < 0.10$) across all levels of sequence identity tested, while for prairie vs. agriculture and effects of tilling, differences in richness only become significant at around 90% SIL. This suggests that in the chronic nitrogen addition experiment there were higher order taxonomic changes in the microbial community (Figure 2.3a), while changes in community structure in response to the other perturbations occurred at lower (finer) levels of resolution (Figure 2.3b-c). Interestingly, a major shift in significance can be seen in dominance and diversity starting at 92-94% sequence identity (Figure 2.3). This shift in significance, seen in several of our analyses (Figures 2.3, 2.4), suggests an important shift in taxonomic hierarchy (i.e., the change from family to genus, or genus to species) and alludes to the possibility that the interpretation of treatment effects on community structure may change depending on the level of taxonomy assessed.

Treatment effects on specific taxa

As the analysis of diversity indicates that the main effects of nitrogen addition, tilling, and prairie conversion have significant impacts on microbial communities, we focus the remaining analyses to determine how these effects acted on individual taxa to affect the community changes observed. To determine taxon specific effects, we used linear regression to analyze changes in OTU frequency of the taxa whose abundance was statistically greater than zero using T-tests at the 98% sequence identity (i.e., "microbial species", Konstantinidis et al. 2006). Taxa frequencies in disturbed and undisturbed plots were positively correlated ($p < 0.0001$) in all three experiments, with r^2 values of 0.24 for

tillage, 0.54 for nitrogen addition, and 0.78 for prairie conversion (Figure 2.5). This suggests that the frequency of most taxa remained largely consistent between treatments and controls. Additionally, regression analysis detected frequency-dependent changes in community structure in each case, with slope values deviating from the 1:1 line ($p < 0.0001$). The decrease in slope from the 1:1 line indicates that, in addition to loss of rare species (which are not included in Figure 2.5), treatments had a significant effect on the distribution of taxa abundance. The slope decrease could be attributed to disturbance causing an increase in the taxa of intermediate abundance, a decrease in the most abundant taxa or some combination. To investigate this further, we used a univariate anaysis to determine the distribution of treatment responses and to identify outliers that were most responsive to each disturbance. This analysis indicated that, while all three treatments reduced the global mean frequency of all taxa ($p \leq 0.05$), the data for nitrogen enrichment was positively skewed ($p \leq 0.01$) due to a small number of taxa that were strongly increased in this disturbance. An outlier test was used to make preliminary determinations of the identity of the responsive taxa. Positive and negative response groups were determined using 95% and 99% quantiles (black filled circles, and open circles, respectively, Figure 2.5). A full accounting of responsive taxa is provided in Tables 9.1, 9.2 and 9.3. These lists are notable for the small percentage of taxa that are represented across all disturbances. Thirteen taxa are represented in both the nitrogen enrichment and prairie conversion lists, while only one is represented across all three lists. The responses of the former group, represented primarily by Acidobacteria, were positively correlated ($r = 0.75$, $p = 0.003$) for nitrogen enrichment and prairie conversion. Similar groups (e.g. families and genera) were represented among both positive and

negative responders to disturbance, with the exception of tillage, which was associated with large declines primarily in the Xiphinematobacteriaceae (Table 9.2). In the case of N enrichment, positive and negative responding Acidobacteria were primarily from different clades (primarily Gps1- 3 and Gps4-6, respectively).

Discussion

Reaping the benefits of massively parallel sequencing

The unprecedented volume of sequences produced by MPS technologies provides unparalleled opportunities for microbial ecology (Schuster 2008). While several studies have employed MPS to define microbial diversity (e.g., Leininger et al. 2006, Sogin et al. 2006, Booijink et al. 2007, Hansen et al. 2007, Konstantinidis and Tiedje 2007, e.g., Roesch et al. 2007, Trosvik et al. 2007, Lauro and Bartlett 2008), we have developed bioinformatic methods that enable OTU designation across sampled plots, extending this approach to allow examination of microbial responses to a broad range of environmental perturbations. Specifically, we used sequence ID tags within the PCR primers (Table 2.1) to multiplex many samples within one run while maintaining the integrity of the analysis. Secondly we used a sequence assembly program (CAP3) to assign OTUs across the entirety of the project allowing frequencies of occurrence to be analyzed in expansive plot designs. Using replicated field plots and statistical analysis we have shown reproducible microbial responses. Massively parallel sequencing promises to be a valuable approach for microbial assessment in diverse systems as it is the only method that has the power to identify the vast number of bacterial species in a cost effective and

timely fashion. This work is unique among those employing MPS as we focus on the change in diversity rather than its overall quantification.

Implications of diversity change

Our results show that specific environmental perturbations (i.e., nitrogen input and alternate tilling practices) can greatly impact microbial communities. These disturbances have a significant effect on the microbial community by reducing species richness and diversity (Table 2.2, Figures 2.3, 2.4, 2.5). Nitrogen input was also observed to have a significant impact on the dominance of a few select taxa (Table 2.2, Figure 2.5). As nitrogen was applied as ammonium nitrate, it is reasonable to assume that groups of bacteria that can utilize either ammonium and/or nitrate as a resource (Hooper et al. 1997) would respond to this treatment. Of the responsive bacteria, *Nitrospira* and *Rhodoplanes*, *Bradyrhizobium, Sphingomonas*, (Table 9.1) are known nitrophiles (Hiraishi and Ueda 1994, Juretschko et al. 1998, Purkhold et al. 2000, Dionisi et al. 2002, Kaneko et al. 2002, Regan et al. 2002, Xie and Yokota 2006) and might be expected to respond to ammonium nitrate addition. However, it has been shown that many nitrifiers grow heterotrophically and therefore may not be responding directly to the addition to nitrogen. We found that a similar group of nitrifiers are positively responding to the addition of nitrogen and negatively responding to conversion to agriculture. At first glance this is counterintuitive, however taking the storage of belowground carbon into consideration, as well as the heterotrophic nature of these bacteria a new hypothesis can be generated. We propose that there is a complete systems change, where the disturbances (e.g. nitrogen addition) are causing a whole suite of changes to the

belowground environment including levels of carbon, quantity and species of roots, to name a few, that through indirect effects are causing minor but important changes to the bacterial community. The exact nature of the bacterial community response to disturbance will require further dissection to elucidate the mechanisms driving the responses.

Land use change, microbial diversity, and community function

The current study shows that conversion of native prairie to an agricultural system, as well as the specific effects of fertilization and tilling, significantly impacts the diversity of the microbial community. However, with the exception of nitrogen additions, the loss of diversity was shared more or less equitably across all taxonomic groups in our study. We further observed that the comparison expected to show the greatest difference in microbial community structure (praire vs. agricultural conversion) actually displayed the least divergence. The most surprising aspect of this was the massive number of factors that are different between prairie and agriculture should be having a larger effect. These differences include differences in nitrogen inputs, the practice of tillage, watering, growth of monocultures of plants, and loss of carbon compared to natural prairie systems, and so on. However, the microbial community appears to be remarkably resistant to change from all of these environmental drivers. We found lower order taxonomic changes caused by the seemingly less invasive addition of nitrogen compared to prairie conversion. How do we reconcile these results with other studies (e.g., Groffman et al. 1993, e.g., Sotomayor and Rice 1996, Ajwa et al. 1999, Dell and Rice 2005, Dell et al. 2005, Williams and Rice 2007) that show significant changes in microbial community

function? We hypothesize that this could be due to multiple factors. Firstly, when faced with changes in multiple drivers, any change in abundance of a specific taxon caused by one driver might be counteracted by another driver. However a more likely cause is that the prairie conversion experiment had many times the sampling intensity than that of the other experiments discussed here. This increased sampling intensity could reveal another feature of the soil microbial community, large amounts of heterogeneity, which could mask the effects of the disturbance on diversity as well as specific taxa responses. Future studies on heterogeneity will be necessary to determine the contribution it makes in understanding soil microbial responses to environmental change.

Conclusions

Overall, while we detected changes in the bacterial communities in response to disturbance, it appears that the community as a whole is remarkably resistant to the types of disturbance tested here. In all cases, microbial community responses included relatively minor changes in taxon frequency leading primarily to the loss of rare taxa. This is surprising as it might be expected that there would be a wholesale change in the soil microbial community in response to disturbances as great as conversion of native tallgrass prairie to an agricultural system where the differences include vegetation type, tillage, nitrogen fertilizer addition, and lack of burning, to name a few. It could be that this resistance to change is a function of spatial heterogeneity in the soil, providing microsites suitable for a range of microbial taxa under many different conditions or simply that land use change consists of multiple drivers that, while acting to change

ecosystem function, can mask the effects on specific taxa. Using new high-throughput technologies across well-designed field experiments, we were able to detect the effects of individual ecological drivers. Additional studies that tease out the effects of multiple drivers might lead to a better understanding of the effects of land use change on microbial communities, perhaps enhancing our ability to predict the effects of these changes. As shifts in microbial communities are likely to play a large role in ecosystem responses in the face of environmental changes, such studies would be both timely and wise.

Acknowledgements

We wish to thank the many people of the Ecological Genomics Institute at Kansas State University for their insight and assistance during the planning and execution of this project. We also want to thank the Konza Prairie Biological Station and the Konza Prairie LTER program for the long-term maintenance of the research site and for allowing us to utilize their facilities. We specifically thank Ari Jumpponen and Dave Myrold for their assistance with the manuscript. This research was funded by the NSF EPSCoR grant EPS-0236913 to Michael Herman and Loretta Johnson.

Figures and Tables

Figure 2.1 Iterative alignment of sequences from 80% to 98% sequence identity.

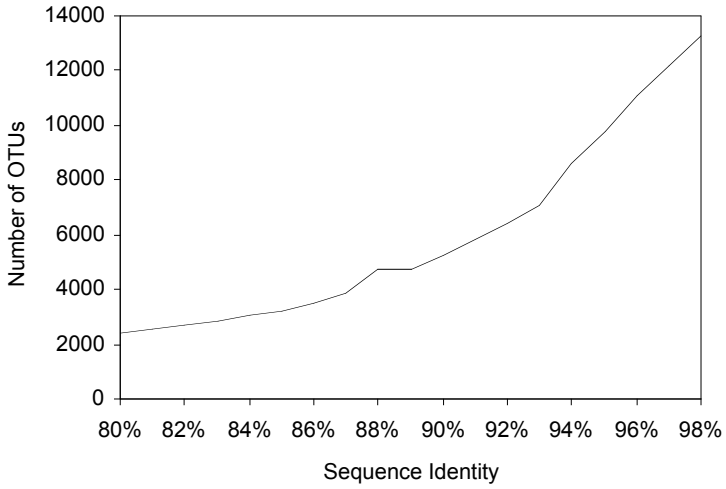


Figure 2.2 Rarefaction curves

Rarefaction curves were calculated independently for the cumulative treatments of (a) burned and unburned control (solid), burned and unburned nitrogen (dashed); (b) tilled (dashed) and untilled (solid); (c) prairie (solid) and agriculture (dashed). Curves were calculated at 98% and 80% sequence identity to illustrate the effect that changing sequence identity level has on OTU designation. Note, scales change between graphs.

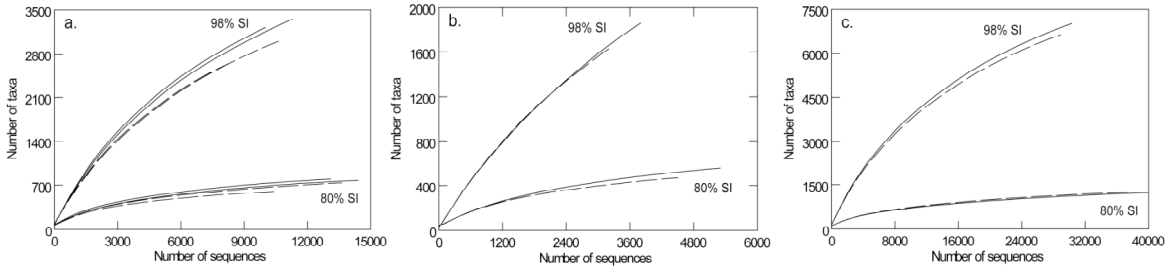
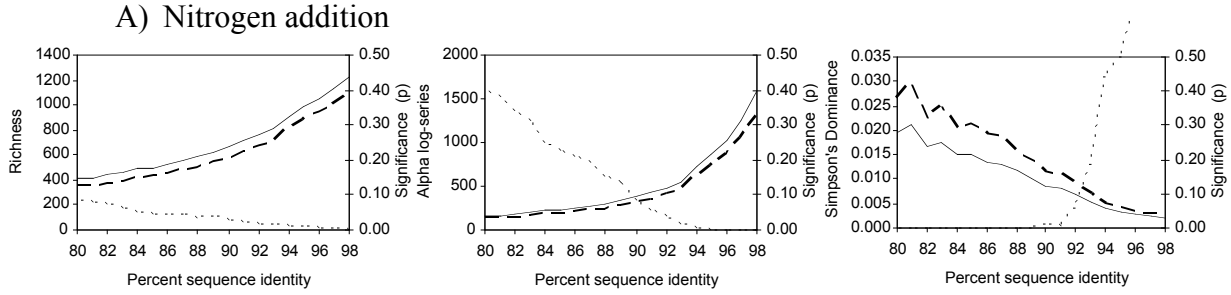


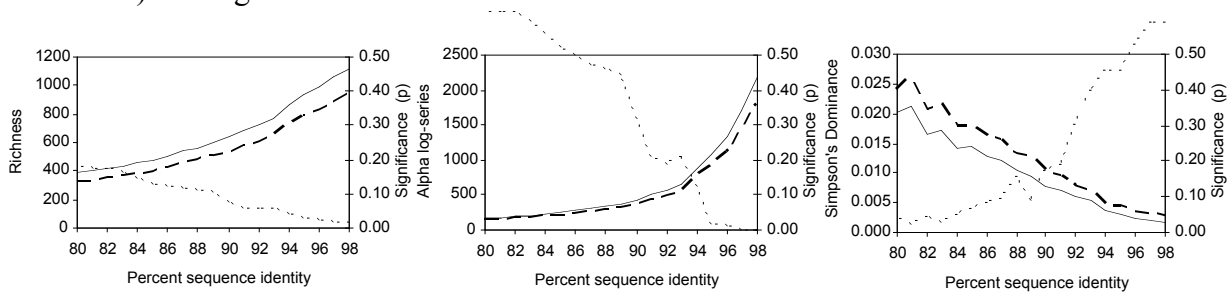
Figure 2.3 Diversity estimates for the effect of disturbance

Diversity estimates for the effect of nitrogen addition (a), tilling (b), and conversion of prairie to agriculture (c) were calculated independently and compiled respectively across each of 18 sequence identity levels (80-98%). The diversity estimate means for control (solid) and disturbed (dashed) treatments (nitrogen addition, tilling and agriculture, respectively) are shown. The change in the *p*-value of the diversity estimate (dotted) as sequence identity changes is also given.

A) Nitrogen addition



B) Tilling



C) Prairie conversion

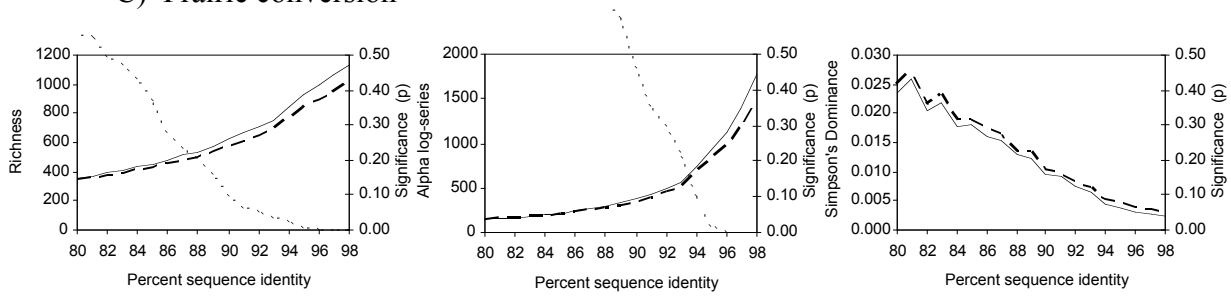
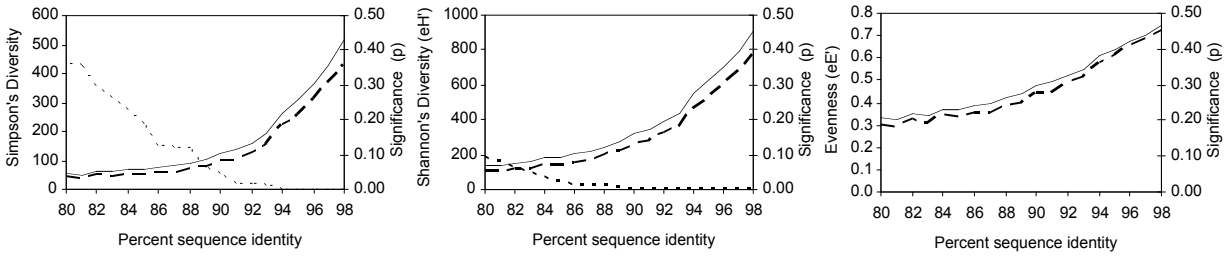


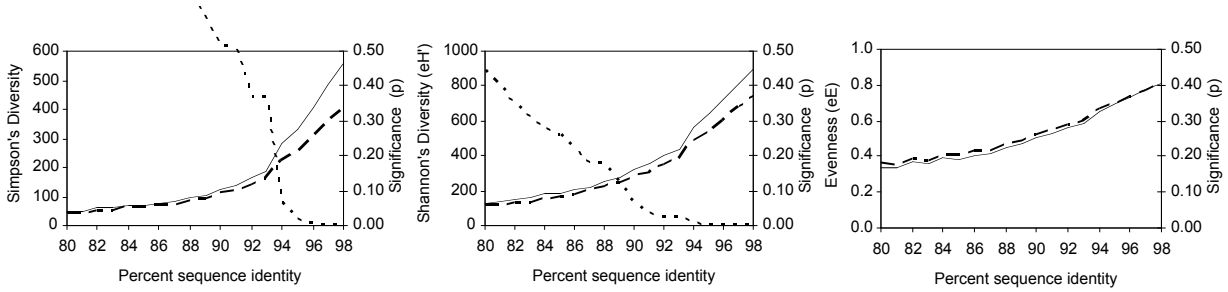
Figure 2.4 Additional diversity estimators

Shannon's and Simpson's diversity were calculated for the effect of nitrogen addition (a), tilling (b), and conversion of prairie to agriculture (c).

A) Nitrogen addition



B) Tilling



C) Prairie conversion

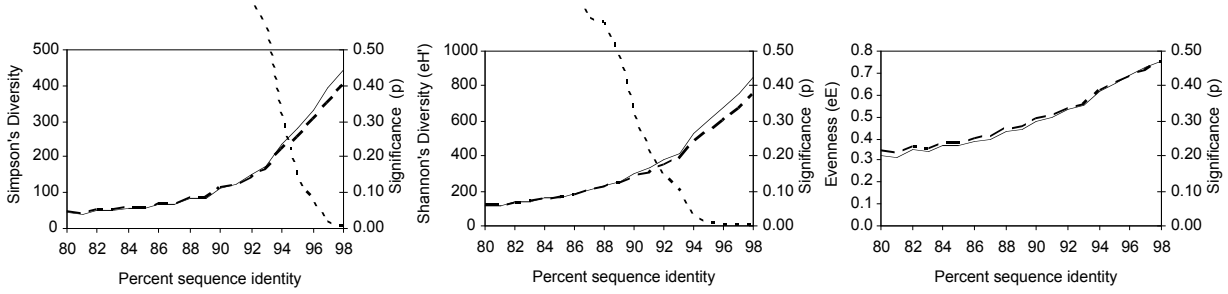


Figure 2.5 Taxon response to disturbance

Taxon response to disturbance ($freq_{disturbed}$ vs. $freq_{undisturbed}$) for (a) nitrogen enrichment [n = 1079, r^2 = 0.54], (b) tilling [n=142, r^2 = 0.24], and (c) conversion of prairie to agriculture [n=1964, r^2 = 0.78]. Comparison of the regression line (solid) to the theoretical 1:1 line (i.e., line of no response; dashed) yields a significant ($p < 0.0001$) negative response of slope to treatment in each study. The top 5% (black filled circles) and 1% (open circles) of the responding taxa are given in Tables 9.1-9.3.

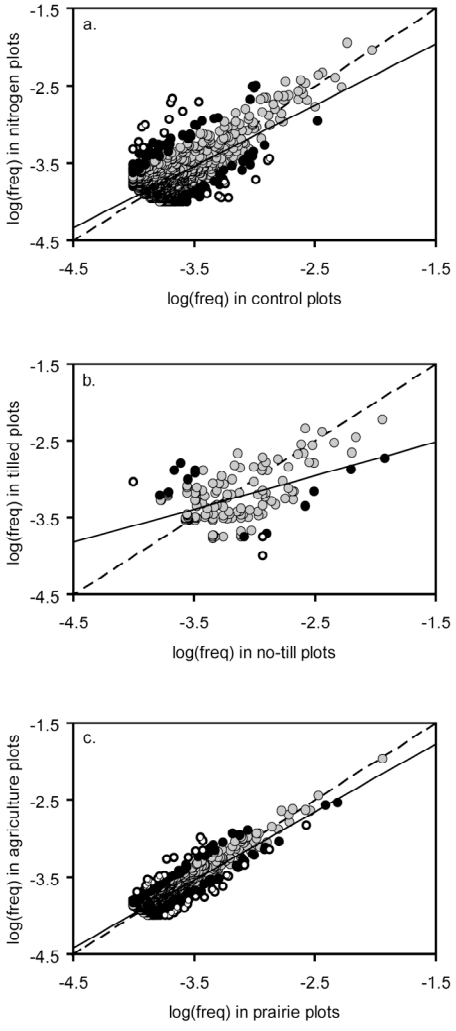


Table 2.1 Primers for barcoded massively parallel sequencing ($_b$MPS)

$_b$MPS primers were produced by adding unique barcode sequences (underlined) between the "A" sequencing primer of Margulies *et al.* (2005) and the reverse 16S primer U529R (**bold**) of Watanabe *et al.* (2001). As sequencing was done in only the reverse direction, no barcode was necessary within the "B" construct.

U341F-FC-B	GCCTTGCCAGCCCGCTCAG**CCTACGGGRSGCAGCAG**
U529R-FC-A3	GCCTCCCTCGCGCCATCAGACTCA**ACCGCGGCKGCTGGC**
U529R-FC-A9	GCCTCCCTCGCGCCATCAGAGCAG**ACCGCGGCKGCTGGC**
U529R-FC-A11	GCCTCCCTCGCGCCATCAGTATCA**ACCGCGGCKGCTGGC**
U529R-FC-A14	GCCTCCCTCGCGCCATCAGAGTAT**ACCGCGGCKGCTGGC**
U529R-FC-A16	GCCTCCCTCGCGCCATCAGCTACG**ACCGCGGCKGCTGGC**
U529R-FC-A18	GCCTCCCTCGCGCCATCAGACTAG**ACCGCGGCKGCTGGC**
U529R-FC-A20	GCCTCCCTCGCGCCATCAGTCTCT**ACCGCGGCKGCTGGC**
U529R-FC-A22	GCCTCCCTCGCGCCATCAGACTCG**ACCGCGGCKGCTGGC**
U529R-FC-A24	GCCTCCCTCGCGCCATCAGACTCT**ACCGCGGCKGCTGGC**
U529R-FC-A25	GCCTCCCTCGCGCCATCAGTGTCA**ACCGCGGCKGCTGGC**
U529R-FC-A27	GCCTCCCTCGCGCCATCAGCTACT**ACCGCGGCKGCTGGC**
U529R-FC-A29	GCCTCCCTCGCGCCATCAGAGCTG**ACCGCGGCKGCTGGC**
U529R-FC-A33	GCCTCCCTCGCGCCATCAGTGATG**ACCGCGGCKGCTGGC**
U529R-FC-A35	GCCTCCCTCGCGCCATCAGAGCGC**ACCGCGGCKGCTGGC**
U529R-FC-A40	GCCTCCCTCGCGCCATCAGTCACT**ACCGCGGCKGCTGGC**
U529R-FC-A42	GCCTCCCTCGCGCCATCAGTGTGC**ACCGCGGCKGCTGGC**
U529R-FC-A46	GCCTCCCTCGCGCCATCAGTACTA**ACCGCGGCKGCTGGC**
U529R-FC-A49	GCCTCCCTCGCGCCATCAGAGATG**ACCGCGGCKGCTGGC**
U529R-FC-A53	GCCTCCCTCGCGCCATCAGCGATG**ACCGCGGCKGCTGGC**
U529R-FC-A57	GCCTCCCTCGCGCCATCAGCTAGT**ACCGCGGCKGCTGGC**
U529R-FC-A64	GCCTCCCTCGCGCCATCAGTACGT**ACCGCGGCKGCTGGC**
U529R-FC-A66	GCCTCCCTCGCGCCATCAGTGTAG**ACCGCGGCKGCTGGC**
U529R-FC-A75	GCCTCCCTCGCGCCATCAGATATG**ACCGCGGCKGCTGGC**
U529R-FC-A76	GCCTCCCTCGCGCCATCAGTCACA**ACCGCGGCKGCTGGC**
U529R-FC-A77	GCCTCCCTCGCGCCATCAGTCTGA**ACCGCGGCKGCTGGC**
U529R-FC-A86	GCCTCCCTCGCGCCATCAGTACAG**ACCGCGGCKGCTGGC**
U529R-FC-A88	GCCTCCCTCGCGCCATCAGTACGA**ACCGCGGCKGCTGGC**
U529R-FC-A89	GCCTCCCTCGCGCCATCAGTGTAC**ACCGCGGCKGCTGGC**
U529R-FC-A90	GCCTCCCTCGCGCCATCAGATACG**ACCGCGGCKGCTGGC**
U529R-FC-A94	GCCTCCCTCGCGCCATCAGACTGA**ACCGCGGCKGCTGGC**
U529R-FC-A95	GCCTCCCTCGCGCCATCAGTGTGT**ACCGCGGCKGCTGGC**
U529R-FC-A96	GCCTCCCTCGCGCCATCAGATCTG**ACCGCGGCKGCTGGC**

Table 2.2 Analysis of variance of taxonomic diversity indices for microbial taxa

F-values are reported and significant main effects represent an overall change in diversity across all sequence identity levels (SIL, 80%-98%), whereas significant interaction effects indicate that the change in diversity was different when looking across sequence identity levels.

	Nitrogen addition (N)		Tilling (T)		Prairie Conversion (PC)	
	N	N x SIL	T	T x SIL	PC	PC x SIL
Richness	10.44 ***	0.46 NS	7.86 *	7.39 ***	51.41 ***	4.70 ***
Alpha log-series	29.34 ***	4.77 ***	7.69 NS	8.31 ***	33.40 *	1.17 NS
Simpson's Dominance	37.24 ***	10.14 ***	3.79 NS	1.96 *	1.49 NS	0.04 NS
Simpson's Diversity	21.07 **	3.30 ***	5.34 NS	7.23 ***	7.31 NS	1.53 NS
Shannon's Diversity	48.15 ***	2.56 ***	15.30 *	10.40 ***	47.77 ***	2.69 ***
Evenness	2.30 NS	0.29 NS	0.17 NS	2.10 **	0.11 NS	0.44 NS

* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$; NS $P > 0.05$

CHAPTER 3 - Genomic basis of nematode-bacteria interactions

Abstract

Determining the genetic mechanisms involved in organismal response to environmental change is essential for understanding the effects of anthropogenic disturbance. This can be challenging, as well developed genomic tools exist for only a few organisms. For example, while the composition of the bacterial-feeding soil nematode community is a biological indicator of disturbance, particularly in grassland ecosystems, the genes involved in the responses of bacterial-feeding soil nematodes to their bacterial environments or food sources are currently unknown. Here we employed *Caenorhabditis elegans* as a model for native soil nematode taxa and used transcriptional profiling to identify candidate genes regulated in response to diets of different bacteria isolated from grassland soils. A large portion of these candidate genes is predicted to effect metabolism and innate immunity. Using mutations that inactivate many of the identified genes, we showed that most contribute to fitness or lifespan in a given bacterial environment. Although these bacteria may not be natural food sources for *C. elegans*, this study demonstrated how changes in food source, as can occur in environmental disturbance, can have a large effect on gene expression, with important consequences for fitness. Furthermore, we identified functions for genes of previously unknown function and new functions for other well-characterized genes, demonstrating the utility of this approach to further describe the *C. elegans* genome. Overall our results suggest that soil nematode fitness in a given bacterial environment in part depends upon the regulation of

metabolic and defense functions that modulate trophic and pathogenic interactions with bacteria.

Introduction

The effects of human-induced environmental change are evident at multiple levels of biological organization. To date, most environmental change studies have focused on effects at the ecosystem, community and organismal levels. However, the ultimate controls of biological responses are located in the genome. Thus, genetic and genomic studies of organismal responses to environmental changes are necessary. Recent advances in genome analysis now make such analyses possible (Ungerer et al. 2008). In this study we begin to span this gap by studying aspects of environmental change using a genetically tractable model organism to identify genes involved in the response to environmental perturbations.

Human-induced changes to the abiotic environment include climatic shifts in temperature and rainfall, effects of pollution and changes in land use, such as conversion of natural landscapes to agriculture (Hannah 1995, Dobson 1997). Of these, the latter appears to be making the greatest impact (Foley 2005). This is particularly the case for grasslands (Samson and Knopf 1994b), which perform many essential ecosystem services, such as supplying clean water, recycling essential nutrients and preserving biodiversity (Daily 1997). In addition, grasslands are among the most endangered ecosystems on the planet, with replacement by agricultural systems altering both the above- and belowground communities (Baer et al. 2002). The tallgrass prairie of North America is an extreme example of such land-use conversion (Samson and Knopf 1994b) with only ~3% of the original prairie remaining (Knapp et al. 1998).

One of the most important members of belowground communities are nematodes, as they are among the most abundant invertebrates in soils, and are an important

component of the microfauna of grasslands (Curry 1994). Soil nematode species display a wide variety of feeding strategies (Freckman 1988), with microbial-feeding nematodes perhaps being the most important consumers of bacteria and fungi in many soil communities (Blair et al. 2000b, Yeates 2003) and their interactions with microbial decomposers affect ecosystem processes including decomposition and nutrient cycling (Freckman 1988, Coleman et al. 1991). In addition, nematodes are responsive to changing environmental conditions (Freckman and Ettema 1993, Todd 1996, Todd et al. 1999), making them ideal organisms to assess the potential impacts of global changes on soil communities (Yeates 2003).

Recent work has demonstrated that the soil nematode community in tallgrass prairie responds strongly to a variety of perturbations, such as nitrogen addition, increased soil moisture and different experimental fire regimes (Todd 1996, Todd et al. 1999, Jones et al. 2006c). These studies suggest that some nematode species are better adapted than others to the environmental changes caused by these disturbances. However, the environmental changes induced by these disturbances are complex, involving changes in the biotic environment that include microbes, competitors and predators, as well as changes in the abiotic chemical environment of the soil. In order to begin to sort out these interactions, we have focused on the responses of microbial-feeding nematodes to the microbial aspects of the biotic environment of grasslands on the Konza Prairie Biological Station. We recently demonstrated that grassland soil bacterial communities respond to various disturbance treatments with significant changes in species richness, dominance, diversity as well as species-specific responses (See Chapter 2). This differential bacterial community response, in conjunction with nematode food

preference (Shtonda and Avery 2006) could drive the observed changes in nematode community structure. Thus, an examination of the genomic response of nematodes to different bacterial environments may reveal the genetic basis of the observed nematode community response.

Here we assess nematode responses to bacteria that could be relevant to the observed changes in native soil nematode communities. Many groups have analyzed the transcriptional response of the genetically tractable nematode *Caenorhabditis elegans* to various, usually medically significant, bacteria (Mallo et al. 2002, Lee et al. 2006a, Troemel et al. 2006), in order to model human innate immunity (Gravato-Nobre and Hodgkin 2005, Sifri et al. 2005). While the high degree of evolutionary conservation allows *C. elegans* to be a good model for human innate immunity, it may be an even better model for soil nematode innate immunity. Therefore, we monitored the genomic response of *C. elegans* exposed to various grassland soil bacteria in the laboratory using transcriptional profiling to identify differentially expressed genes. In addition, we analyzed the functional significance of some of the differentially expressed genes by measuring fitness of mutant nematodes in the various bacterial environments. Our results demonstrate that certain genes are specifically induced in response to different bacteria and in many cases these genes function to contribute to nematode fitness and lifespan in different bacterial environments.

Materials and Methods

C. elegans and bacteria strains and maintenance

The following mutant strains were used: Bristol (*N2*), *cpi-1*(*ok1213*), *dpy-17*(*e1295*), *gei-7*(*ok531*), *mtl-2*(*gk125*), *dhs-28*(*ok450*), *Y57A10C.6*(*ok693*), *acdh-1*(*ok1489*), *rol-6*(*e187*), *ctl-1*(*ok1242*), *dpy-14*(*e188*), *fat-2*(*ok873*), *gld-1*(*op236*), *hsp-12.6*(*gk156*), *cey-2*(*ok902*), *cey-4*(*ok858*), *cyp-37A1*(*ok673*), *elo-5*(*gk182*), *pab-2*(*ok1851*), *sqt-2*(*sc108*), *F55F3.3*(*ok1758*), *C23H5.8*(*ok651*). Growth and maintenance conditions were as described (Brenner 1974, Sulston and Hodgkin 1988). Use of native soil bacteria was as for *E. coli* (OP50). Bacterial isolate 16S rDNA was sequenced to identify species and sequence is available at GenBank accession numbers: *Pseudomonas sp.*: EU704696, *Micrococcus luteus*: EU704697, *Bacillus megaterium*: EU704698.

Food preference and pathogenicity assays

Biased choice assays were performed as previously described (Shtonda and Avery 2006). Longevity assays were performed as previously described (Tan et al. 1999a, Tan and Ausubel 2000) and TD_{50} was calculated from survivorship curves as time to death for 50% of individuals in a population. Briefly, worms were synchronized by bleaching to collect eggs and hatched in M9 overnight. Worms were then grown to L4 on *E. coli* (OP50) to standardize test populations, and then transferred to the test bacteria as young adults (20 worms per plate) and were maintained at 25°C. Surviving worms were then re-plated daily and the fraction surviving was determined every 12 hours. Worms were determined to be dead when they no longer responded to touch with platinum wire. All pathogenicity tests were conducted in at least ten independent replicate experiments.

Life table analysis

Demographic measures were collected on individual worms in the four bacterial environments. Using life table analysis, lambda and other population parameters were calculated. Mutant functional tests were performed by plating eggs onto the test bacteria and then placing progeny from this generation onto the test bacteria, one L4 hermaphrodite (P_0 worm) per plate was incubated at 20°C with at least 10 replicates per treatment per strain. Re-plating of the original P_0 worm was done daily until death. Progeny per day was counted (age specific reproduction or m_x). Survival of the Po worm was monitored as well as the survival of all of the progeny from each reproductive period to determine age specific survival (l_x). Using life table analysis, intrinsic growth rate (R_o) was calculated as the sum of l_x times m_x ($\sum l_x m_x$). Generation time (T) was calculated by ($\sum l_x m_x$)/($\sum x l_x m_x$ where x = age class). Lambda (λ) was determined from R_o and T by calculating $\lambda = e^{(\ln Ro/T)}$, and λ was used as a measure of absolute fitness (Neal 2004). Replicate populations and subsequent life table calculations were used as replicates for statistical tests and each treatment by strain combination was repeated at least 10 times.

Statistical Analysis of functional tests

Hypothesis testing of *a priori* contrasts was done using the GLM procedure in SAS (SAS Institute Inc., Cary, North Carolina, USA). The model used for these tests is shown.

Model: $Y = \mu$ + genotype + bacteria + genotype X bacteria + error

With each hypothesis tested using test statements, and Y equal to any of the measured life history values (i. e. λ, T, R_o, TD_{50} etc.).

Microarray Hybridizations

Caenorhabditis elegans spotted oligonucleotide microarrays were obtained from the Genome Sequencing Center at Washington University in St. Louis. Extracted mRNA was made into cDNA using Genisphere 3DNA Array350 kits according to manufacture recommendations for use in a hybridization station (Genisphere Inc., Hatfield, Pennsylvania, USA). Microarray hybridizations were performed at the Kansas State University Gene Expression Facility using a Tecan 400 Hybridization station (Tecan Inc., Zurich, Switzerland). Indirect labeling of cDNA was used to prevent hybridization bias associated with direct labeling procedures (Manduchi et al. 2002). Hybridizations were carried out for 16 hours at 42° C according to manufacturer recommendations (Gensiphere Inc. and Tecan Inc.). Hybridized arrays were scanned with an Axon GenePix 4000B (MDS Analytical Technologies, Toronto, Canada) and data was collected using GenePix 6.0 software (MDS Analytical Technologies). Gridding and preprocessing were done manually to remove bad spots and dye blobs. Raw data files generated are MIAME compliant (Brazma et al. 2001) and available at Gene Expression Omnibus (GEO).

Microarray data analysis

Comparisons between treatment groups were made in a factorial design with all six pair-wise comparisons made to maximize the ability to detect differences between any

two treatments (Churchill and Oliver 2001). Comparisons were made with six biological replicates incorporating a dye swap every other replication to account for any potential dye bias associated with a particular fluorophore (i.e. Cy3 or Cy5) (Manduchi et al. 2002). Data was analyzed as in Wolfinger *et al*. (2001) using SAS statistical software (SAS Institute Inc., Cary, North Carolina, USA) where we conducted a two-step mixed model analysis of variance to account for all possible sources of variance. This two-step ANOVA was performed using the MIXED procedure in SAS, with the model for the first stage below and Y = background subtracted raw intensity from the raw data files generated by GenePix 6.0 (MDS Analytical Technologies).

Stage 1 model:
$\log_2 Y = \mu + \text{array} + \text{dye} + \text{array X dye} + \text{error}$

Where residuals, termed Relative Fluorescence Intensities (RFI) from stage 1 serve as the input for stage 2.

Stage 2 model:
$\text{RFI} = \mu + \text{array} + \text{dye} + \text{treatment} + \text{error}$

We used the false discovery rate (FDR) q to address the multiple testing problem (Storey and Tibshirani 2003). q statistics were calculated in Q-VALUE and using the significance threshold $q < 0.01$ we removed those genes that did not respond to bacterial environment in contrasts. Volcano plots were made in JMP 5.0 software (SAS Institute Inc., Cary, North Carolina, USA). An example of our SAS code can be found at (www.k-state.edu/hermanlab/SASCODE).

Gene classification

Identified genes were assigned Gene Ontology (GO) terms in batch with WormMart (www.wormbase.org/biomart/martview/) and manually grouped by similar function incorporating when possible new annotations found in recent literature. Clustering of genes according to the *C. elegans* co-expression mountains was performed in batch (http://workhorse.stanford.edu/cgi-bin/plot/plotcoordinates.pl).

Results

C. elegans response to soil bacteria

Modeling a more ecologically relevant environment for free-living soil nematodes in the laboratory, we isolated bacteria from grassland prairie soils at the Konza Prairie Biological Station for use as *C. elegans* food sources and environments. Soil bacterial titers identified *Micococcus luteus* as the most abundant bacterial species in the nutrient amendment plots (supplemented annually with $10g/m^2$ ammonium nitrate for 21 years) that was culturable on nematode growth media plates (data not shown). *Bacillus megaterium* and *Pseudomonas sp.* were isolated in association with soil nematodes from Konza prairie soils (*Oscheious sp.* and *Pellioditis sp.* respectively) (Jones et al. 2006a, Jones et al. 2006c). The 16S rDNA sequence of the *Pseudomonas sp.* isolated did not match any known taxon in the Ribosomal Database Project (RDP). The closest match was *Pseudomonas flourescens*, which was 98% identical across the entire 16S rDNA with our isolate (See Methods).

Wild-type *C. elegans* (*N2*) was grown on the three grassland soil bacterial species as well as *E. coli* (OP50) which served as a control, as it is the typical laboratory diet for *C. elegans* (Sulston and Hodgkin 1988) and allowed for comparisons with the wealth of *C. elegans* data that has been generated using *E. coli* (OP50) as the food source. The different bacteria served as *C. elegans* food source as well as their immediate environment during growth, as animals were grown on a bacterial lawn covering the entire surface of the nematode growth media agar plate. Ultimately, we were interested in the effects of different bacterial environments on nematode fitness. While brood size is often used as a surrogate for fitness, this can lead to an incomplete picture, as it does not take other important aspects of fitness into account. Therefore, we used life table analysis to estimate absolute fitness (λ), which accounts for age specific fecundity (m_x) and survival (l_x), as well as generation time (T) (Neal 2004) (and is subsequently much more comprehensive, See Methods). The absolute fitness of wild-type animals differed significantly in the different bacterial environments. Animals displayed the highest fitness when grown on *Pseudomonas sp.* (λ = 3.99), which was significantly greater (P = 0.021) than when grown on *E. coli* (λ = 3.60), *B. megaterium* (λ = 2.81, $P < 0.0001$) and *M. luteus* (λ = 2.63, $P < 0.0001$). Fitness of wild-type animals in the *E. coli* environment was also significantly higher than in either *B. megaterium* (P = 0.027) or *M. luteus* (P = 0.027) environments (Figure 3.1A, Table 3.1). It is interesting to note that the only previous study to use life tables to calculate fitness in *C. elegans* found highly similar values (λ = 3.85 on *E. coli*) demonstrating the reproducibility of these analyses in separate labs (Chen et al. 2007). In our experiments, *C. elegans* appeared to have higher fitness on Gram-negative, than on Gram-positive bacteria which could be due to

differences in the way these groups are recognized (Tenor and Aballay 2008) and lysed. However, this needs to be explored further with a greater number of bacterial species.

Another important aspect of nematode demography that has been well studied is lifespan, which is commonly measured as time to death for 50% of a population (TD_{50}) (Tan et al. 1999b, Tan and Ausubel 2000) using survivorship curves (Figures 3.1B, 3.2A-D) and is indicative of the quality and pathogenicity of *C. elegans* food sources. We observed differences in lifespan with wild-type animals having lower TD_{50} values when grown on *M. luteus* (TD_{50} = 4.1) than during growth on *E. coli* (TD_{50} = 5.6), while growth on both *Pseudomonas sp*. (TD_{50} = 8.7) and *B. megaterium* (TD_{50} = 12.3) increased lifespan with all pair-wise comparisons of the four bacterial environments significant ($P < 0.0001$) (Figure 3.1C, Table 3.1). The extended lifespan in the *B. megaterium* environment is not a consequence of starvation (Lakowski and Hekimi 1998, Kaeberlein et al. 2006, Lee et al. 2006b) as generation time is not altered as would be expected of worms under caloric restriction (Table 3.2).

To further characterize wild-type *C. elegans* response to the bacterial isolates we conducted food preference tests. Using a biased choice assay (Shtonda and Avery 2006) (Figure 3.3A) we determined food preference for all pair-wise combinations of bacterial isolates (Figure 3.3B). Comparisons of the pair-wise measures of preferences, revealed a hierarchy of food preferences (Figure 3.3C). Interestingly, this hierarchy mirrored the trend that was observed in fitness tests in the different bacterial environments (Figure 3.1A), with *C. elegans* preferring *Pseudomonas sp.* on which it was most fit, followed by *E. coli*, *B. megaterium*, and *M. luteus*, respectively. Thus *C. elegans* food preference

appears to correlate with fitness, with bacterial environments on which worms were most fit being preferred.

Genomic transcriptional response

Transcriptional responses of *C. elegans* adults were assayed after growth on each of the four bacteria: *E. coli*, *M. luteus*, *Pseudomonas sp.*, or *B. megaterium*. Adult animals were analyzed in order to reduce the possibility that age differences confound responses to the different environments. We hybridized cDNA to *C. elegans* whole genome spotted oligonucleotide microarray chips (Washington University St. Louis Genome Sequencing Center). Using six biological replicates, relative expression levels were obtained from all pair-wise comparisons of the different bacterial environments (Figures 3.4, 3.5). A total of 372 genes that were significantly differentially expressed were identified across all pair-wise comparisons using the MIXED procedure (Wolfinger et al. 2001) in SAS (SAS Institute Inc., Cary, North Carolina) and controlling for the increased experiment-wise error rate using the false discovery rate $q < 0.01$ (See Methods) (Storey and Tibshirani 2003). Due to overlap of identified gene sets, the 372 instances of differentially expressed genes across all comparisons correspond to a total of 204 unique genes identified (Tables 3.3, 8.1). Expression levels of ten genes were determined using qPCR and all matched levels observed by microarray (Tables 3.4, 3.5) validating the data quality.

Over-represented functional groups

Gene Ontology (GO) terms for the identified genes were used to group genes by similar function (See Methods). Genes annotated as metabolism were highly represented (9.3% of total) as expected. Interestingly, genes previously implicated in innate immunity were found in all six comparisons (9.8% of the total). Surprisingly, 8.8% of identified genes were involved in cuticle biosynthesis or were cuticular collagens. Finally, genes of unknown function made up the largest portion, 61% of the total (Figure 3.5B, Table 3.6), as one might expect, as *C. elegans* had not previously been exposed to these environments.

We also mapped the identified genes to the *C. elegans* co-expressed gene mountains and found similar groups over-represented, as quantified by the representation factor (RF) (Kim et al. 2001) (Table 3.7). 32/204 (RF = 3.4, P = 7.9e-05) of the identified genes mapped to mount 8, which is enriched with genes associated with mitosis as well as genes previously implicated in innate immunity. We also found genes (11/204) that were over-represented in mount 16 that contains retinoblastoma complex genes (RF = 4.1, P = 0.008). Mount 19, which is comprised predominately of genes involved in glycolysis, contained 14/204 (RF = 6.4, P = 4.2e-06) identified genes. 25/204 of our identified genes were found in the mount 22, which represents genes involved in carbohydrate metabolism (RF = 14.3, P = 4.7e-20). Mount 23 contained 10 of our differentially regulated genes and this mountain contains Ras pathway components (RF = 6.0, P = 6.2e-4) (Table 3.7). Thus our identified genes clustered similarly using two different methods, with *C. elegans* genomic response to different bacterial environments enriched for metabolic and defense mechanisms presumably for protection and nutrition.

Biological validation of identified genes

We obtained all available viable mutations for the 204 differentially expressed genes in our study, (21/204, ~10% of the total genes identified) from the *Caenorhabditis* Genetics Center (CGC) and used them for biological validation of our microarray results (Table 3.8). We performed functional tests measuring multiple aspects of life history including brood size, generation time (Table 3.2, Figure 3.6), absolute fitness and lifespan (TD_{50}) (Table 3.1, Figure 3.2) in all four bacterial environments. We first investigated mutational effects within each environment independently. We found that many of the mutations had effects on life history traits and differed significantly from wild type in a given bacterial environment. While most mutants had decreased fitness compared to wild type in each environment, surprisingly, a few had increased fitness when grown on *E. coli*, *M. luteus* and *Pseudomonas sp*. (Tables 3.1, 3.9). Interestingly, in the *B. megaterium* environment more mutants experienced increased fitness than decreased fitness.

A similar trend was found for brood size as most mutants had reduced numbers of progeny in response to growth on *E. coli*, *M. luteus*, and *Pseudomonas sp.*, while growth on *B. megaterium* resulted in equal numbers of mutant alleles that significantly increased and decreased brood size (Tables 3.2, 3.9). Similarly, generation times were slower for most mutant alleles on the same three bacterial environments and only in the *B. megaterium* environment were there more mutant alleles with faster generation times (Table 3.9). Surprisingly, lifespan showed a different trend. Growth of mutant strains on *Pseudomonas sp*., *M. luteus* and *B. megaterium,* primarily caused reductions in lifespan,

while growth on *E. coli* caused most mutant strains to significantly increase lifespan. Overall, many of the mutations affected life history as a result of the bacterial environment in which they were grown. These data demonstrate that transcriptional profiling identified genes of functional importance in each bacterial environment.

Specificity of functional response and genotype by environment interactions

In order to compare across bacterial environments we investigated genotype by environment interactions (GEI) and examined mutant norms of reaction across bacterial environments (Figures 3.6, 3.7). Reaction norms of brood size, generation time (Figure 3.6), fitness, and lifespan (Figure 3.7) across multiple environments, revealed differential effects of the bacterial environments on the different mutant genotypes. Not only were there differences across the different environments, but these differences were not always orderly with multiple reaction norms crossing, illustrating one type of GEI (Figure 3.7, Table 3.10). We found many instances of this phenomenon within our life history data demonstrating the specificity and complexity of mutational effects.

The simple assumption that any particular gene positively regulates a particular life history trait allows for predictions about the directionality of mutational effects. For example one might expect that loss of a gene function that is upregulated in a particular environment, would cause a reduction in fitness. One such example is *hsp-12.6* which encodes a heat-shock protein (Hsu et al. 2003) and was found to be up-regulated to a large extent when wild-type *C. elegans* was grown on *E. coil* compared to growth on *B. megaterium* (Table 3.10). Our simple assumption predicts that loss of *hsp-12.6* function would cause a specific reduction in fitness when grown on *E. coli*. This is exactly what

we found: *hsp-12.6* mutants have a 15% proportional reduction in fitness as compared to wild type when the mutant is grown on *E. coli* (Figure 3.8) which is significantly different ($p < 0.001$) from the proportional changes observed on *B. megaterium*. Not only is this difference significant, but fitness of *hsp-12.6* mutants was significantly increased when grown on *B. megaterium* relative to wild type (Table 3.1). This suggests that there was a cost associated with the expression of *hsp-12.6* in an environment in which it was not needed and a detriment to loss of function in an environment in which it was needed. Thus the *hsp-12.6* allele had an antagonistic pleiotrophic effect on fitness in these environments.

Another simple example of specificity was the effect of *rol-6* mutations on lifespan. Transcriptional profiling showed that *rol-6* was significantly upregulated in the *E. coli* versus *Pseudomonas sp.* environment. Thus, the simple prediction was that loss of *rol-6* function would cause a decreased lifespan, which was what we observed. When the *rol-6* mutant strain was grown on *E. coli* there was a 45% proportional reduction in TD_{50} compared to the 11% reduction observed when grown on *Pseudomonas sp.* and this difference was significant ($P < 0.0001$) (Table 3.10, Figure 3.9). Interestingly, *rol-6* has been well characterized and extensively studied (Kramer et al. 1990, Blaxter 1993, Kramer and Johnson 1993, Park and Kramer 1994), yet through use of alternate environments an additional function may have been uncovered.

Although we observed examples that met the expectations of the simple prediction that genes positively impact particular life history traits, in many cases the underlying gene regulation may be more complex involving positive and negative regulation. Thus in most cases we do not expect to be able to predict the directional

effect of a particular mutation on the trait. Instead we predicted that we would observe GEI between the environments in which differential expression was found. There were 37 instances of significant differential expression among the 21 genes tested (Table 3.10). We performed an ANOVA for each instance of differential expression and determined the significance of GEI for fitness and lifespan (See Methods). We found 49 percent (18/37) of the contrasts of mutant fitness in the six bacterial comparisons were significant. The significant fitness GEI is illustrated in graphs of reaction norms by bacterial comparison (Figure 3.7A). Using the same methodology for lifespan, we found that 35/37 (95%) of TD_{50} GEI tests were significant (Table 3.10, Figure 3.7B). Thus, it appears that the majority of differentially expressed genes are functionally important in the specific environments in which they were regulated; illustrating gene by environment interaction is likely a common feature to genes that are regulated in response to different bacterial environments.

Discussion

We characterized wild-type *C. elegans* responses to soil bacterial food sources/ environments to model naturally occurring interactions that may be driving bactivorous nematode community structural responses to land use change in grasslands. Although these bacterial environments might not be encountered by *C. elegans* in the wild, *C. elegans* is an excellent model for bactivorous nematodes. We used transcriptional profiling to identify 204 genes that are significantly differentially expressed when *C. elegans* is grown on different bacterial food sources/environments isolated from grassland soils. Of the identified genes, most were characterized to be involved in

metabolism, defense, cuticle biosynthesis, or were of unknown function. This suggests that the majority of the response to soil bacterial isolates by bactivorous nematodes could also be driven by metabolic and defense functions.

Using available mutants we conducted functional tests to biologically validate the list of identified genes and found that as many as 49% of the mutants had a significant effect on fitness. A unique aspect of this work is that we calculated fitness using classical life table analysis. To our knowledge this is the first use of such analyses to biologically validate candidate genes identified by transcriptional profiling. Furthermore, 95% of the mutants had a significant effect on innate immunity as measured by lifespan (TD_{50}). These results illustrate the power of transcriptional profiling to identify candidate genes important for responses to different environments. Interestingly, other studies that have used gene inactivation to biologically validate environmentally induced differential expression (Morozova et al. 2006, Cui et al. 2007), have also found that a large proportion of genes are functionally important suggesting this could be a common feature to transcriptional regulatory networks that are involved in response to external stimuli (Wittkopp 2007). Using new environments we have been able to assign new functions to genes of unknown function as well as to show alternate functions to previously well-characterized genes. We anticipate that using this methodology we will be able to continue characterizing the large proportion of *C. elegans* genome that is currently of unknown function. Furthermore, genotype-by-environment interactions were prevalent in our functional tests demonstrating the utility of environmental assays to elucidate gene function. For example, we identified *pab-2* in our microarrray experiments and functional tests demonstrated that it had a significantly higher fitness than wild type (λ =

4.14 vs. 3.60 respectively) when grown on *E. coli* (Table 3.1). This result is interesting as N2 was cultured on *E. coli* OP50 for decades (>1,000 generations) prior to being frozen (Schulenburg et al. 2004). It is likely that during this extended period of time that N2 became better adapted to life on *E. coli*. It is even more exciting that when we investigated the lifespan of *pab-2* mutant animals in the *E. coli* environment, we found that not only did this strain have greater fitness but it also had a longer lifespan (TD_{50} = 6.6 vs. 5.6) (Table 3.1), larger brood size (300.2 vs. 290.8) and a faster generation time (4.01 days) than wild type (4.4 days) (Table 3.2) when grown on *E. coli*. Interestingly this mutant does not follow the trend postulated by Hodgkin and Barns (Hodgkin and Barnes 1991) where there does not appear to be a trade-off between developmental rate and brood size. It will be interesting to see how this increased fitness is mediated, and further test the extent of this increase in fitness.

Interestingly, when we analyzed wild type food preference for the four bacterial isolates we found that the hierarchy of food preference mirrors the trend observed for fitness in the different bacterial environments. This suggests that *C. elegans* prefers the environment in which it will be most fit. It will be interesting to see how *C. elegans* makes this choice and ultimately maximizes fitness. This preference could also be a mechanism driving observed nematode community structure in grassland soils, as we have recently found that perturbations that mimic disturbances caused by land-use change not only alter nematode communities (Jones et al. 2006c) but also the bacterial community (See Chapter 2). We have also observed that native soil nematodes differ in their susceptibility to the different bacteria in terms of infection/colonization (Coolon and Herman unpublished data), thus pathogenicity may also contribute to soil nematode

community structure. Taken together we suggest that the expression of metabolism and defense functions may in part drive nematode community dynamics in grassland soil systems.

Acknowledgements

We thank the *Caenorhabditis* Genetics Center for providing strains. We thank Brett Sandercock for help with demographic methods, and Ted Morgan for statistical advice. Thanks to the members of the Ecological Genomics Institute at Kansas State University for discussions and helpful comments and the Gene Expression Facility at Kansas State University for use of the facility.

Figures and Tables

Figure 3.1 Effects of bacterial environment on wild type life history traits

(A) Absolute fitness (λ) values for wild type (*N2*) in each bacterial environment. (B) Survivorship curves showing the proportion of the starting population surviving at different times in each bacterial environment. (C) Time to death for 50 percent of the individuals in a population (TD_{50}) is shown for N2 grown in each bacterial environment. Standard error (s.e.m.) is indicated with error bars.

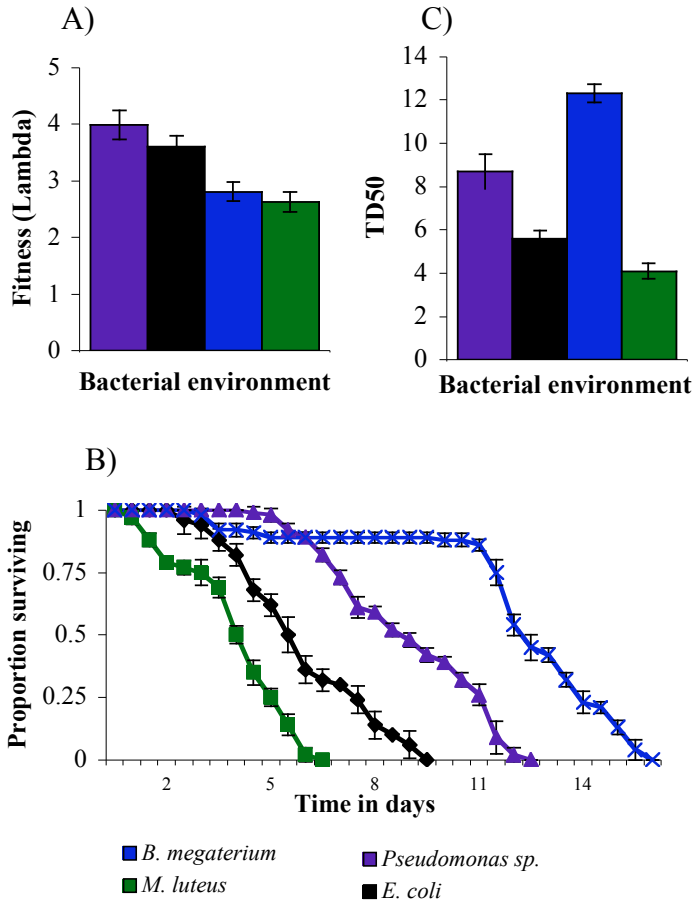


Figure 3.2 Survivorship curves

Survivorship curves are shown for N2 and mutant strains across time for bacterial environments: (A) *B. megaterium* (B) *E. coli OP50* (C) *M. luteus* (D) *Pseudomonas sp.*

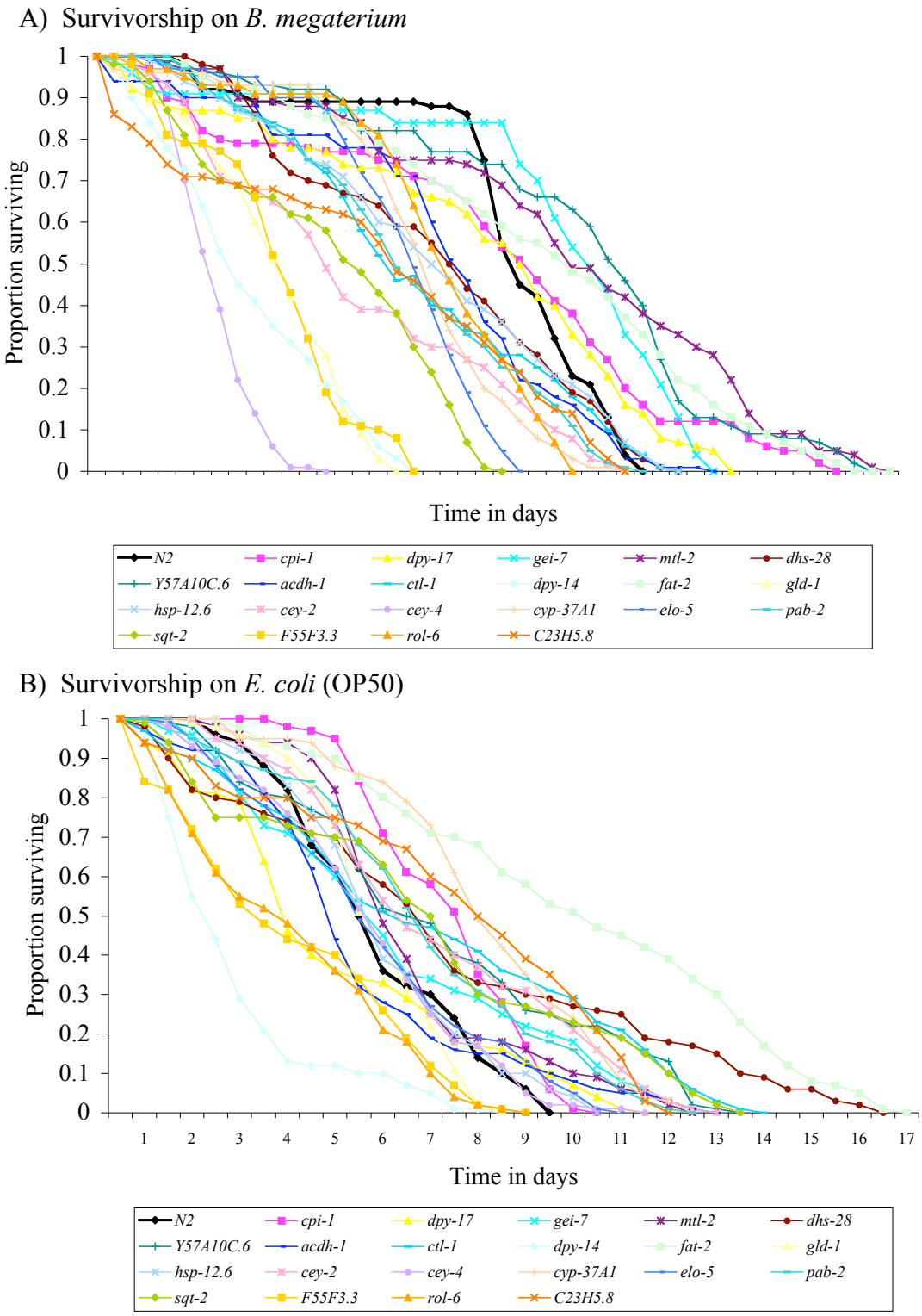
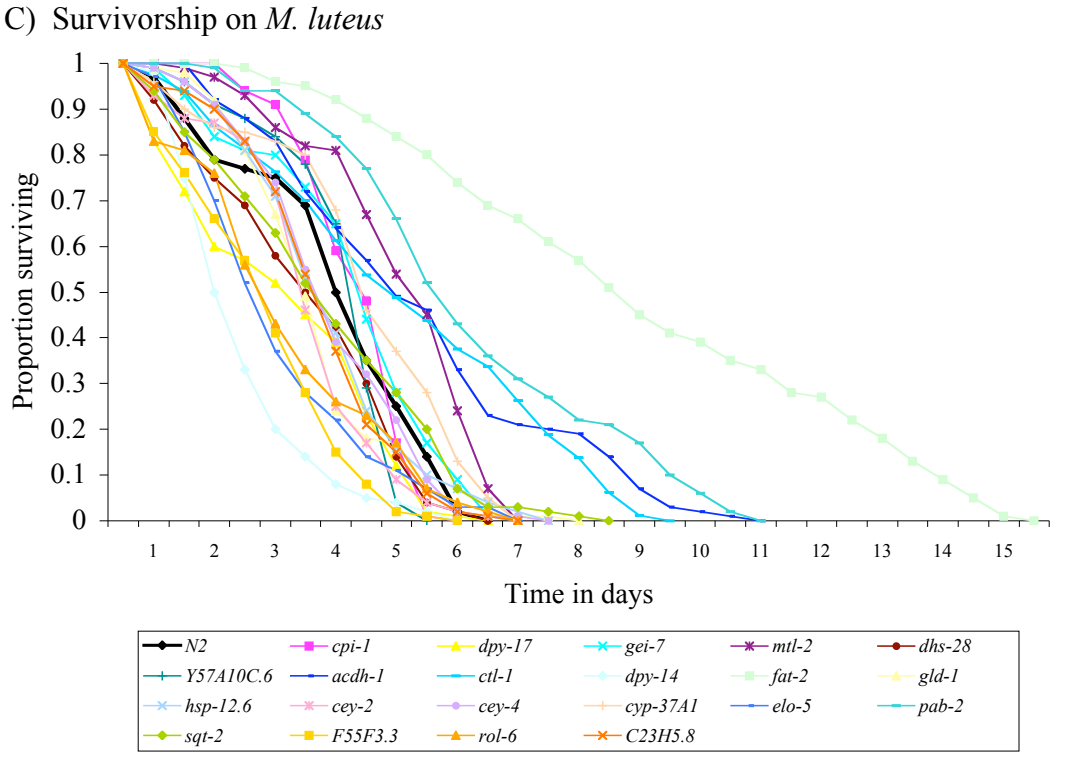


Figure 3.2 Survivorship curves

Survivorship curves are shown for N2 and mutant strains across time for bacterial environments: (A) *B. megaterium* (B) *E. coli OP50* (C) *M. luteus* (D) *Pseudomonas sp.*



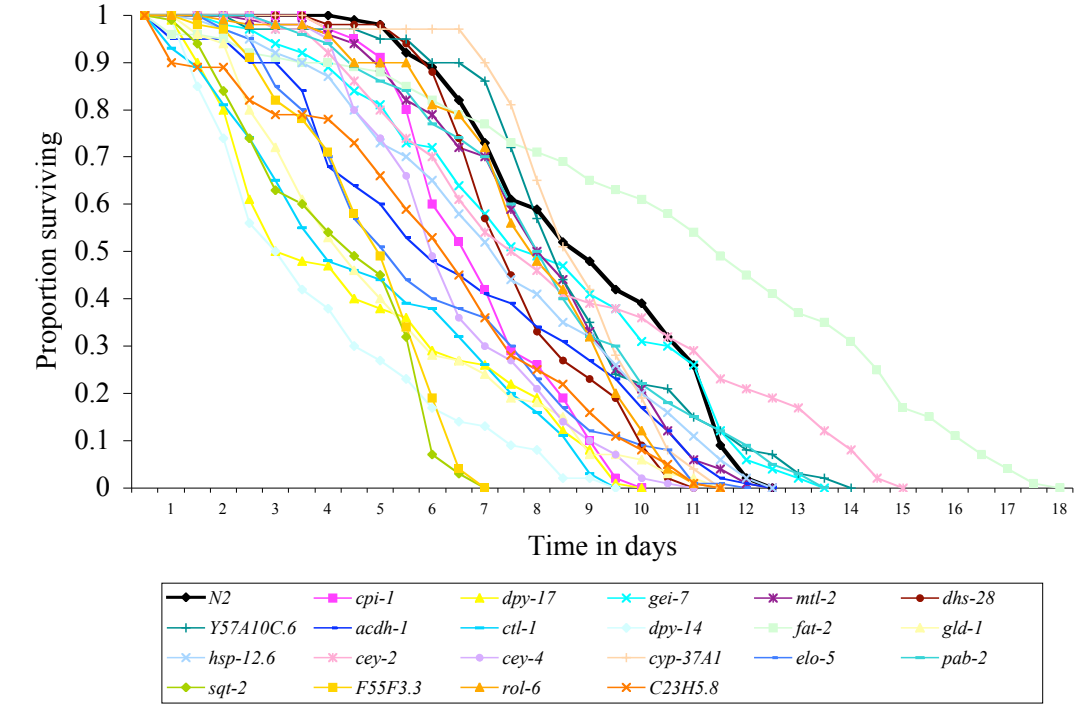
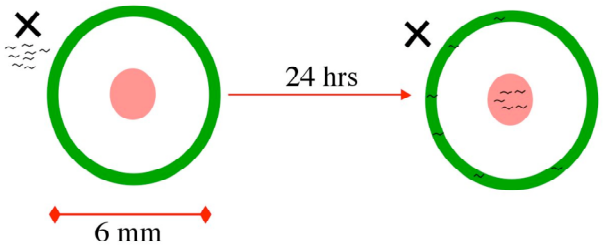


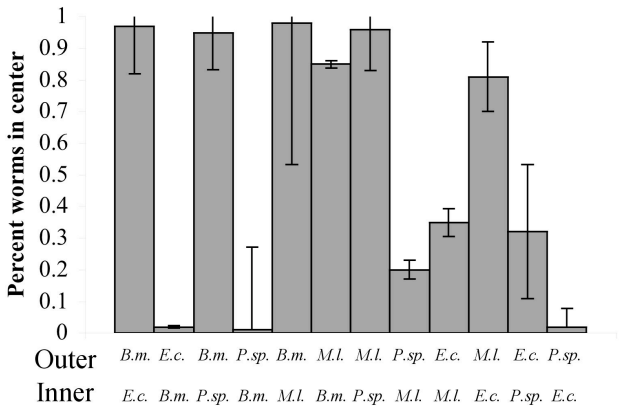
Figure 3.3 Food preference

(A) Food preferences of wild-type animals were measured in a biased choice assay modified from Shtonda and Avery (2006). Bacteria were arrayed on an agar plate with a center bacteria and surrounding this in a circle a second bacterial type. Synchronized L1 larvae were placed outside the outer circle (indicated by the X) and the fraction in the center bacterial type was determined after 24 hours. (B) Fraction of nematodes in the center bacterial type is shown for all pair-wise comparisons and reciprocal comparisons were used for *C. elegans* food preference. Standard error for each mean (s.e.m.) is indicated with error bars. The bacteria listed under each bar were compared and are either outer (outer ring) or inner (inner circle) and *B.m.* = *B. megaterium*, *M.l.* = *M. luteus*, *E.c.* = *E. coli*, *P.sp.* = *Pseudomonas sp.* (C) From the observed food preference in the biased-choice assay the hierarchy of food preferences for the four bacterial isolates was determined.

A)



B)



C)

Pseudomonas sp. ≥ *E. coli (OP50)* >> *B. megaterium* ≥ *M. luteus*

Figure 3.4 Microarray experimental design

The experimental design for the microarray comparisons made is shown. All pair-wise comparisons of adult *C. elegans* in the four bacterial environments were made in a factorial design. Six biological replicates were used and dye-swaps were preformed every other replicate.

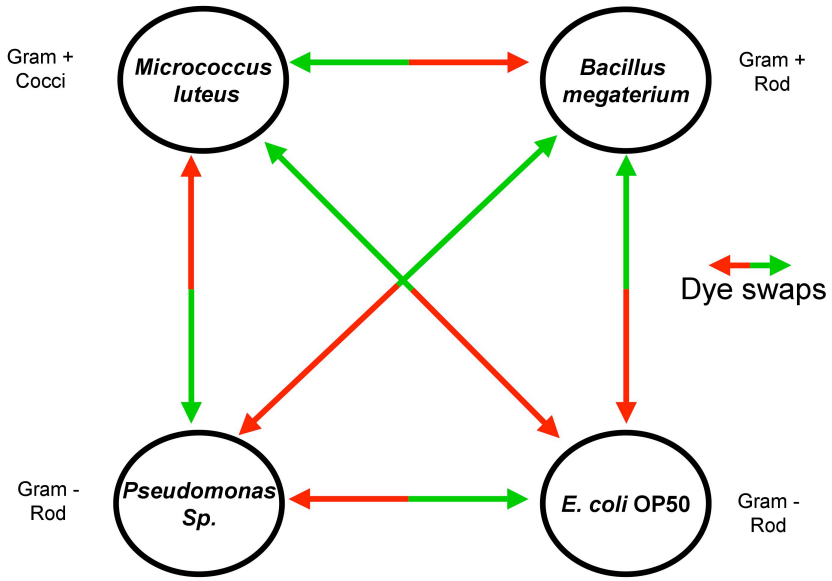
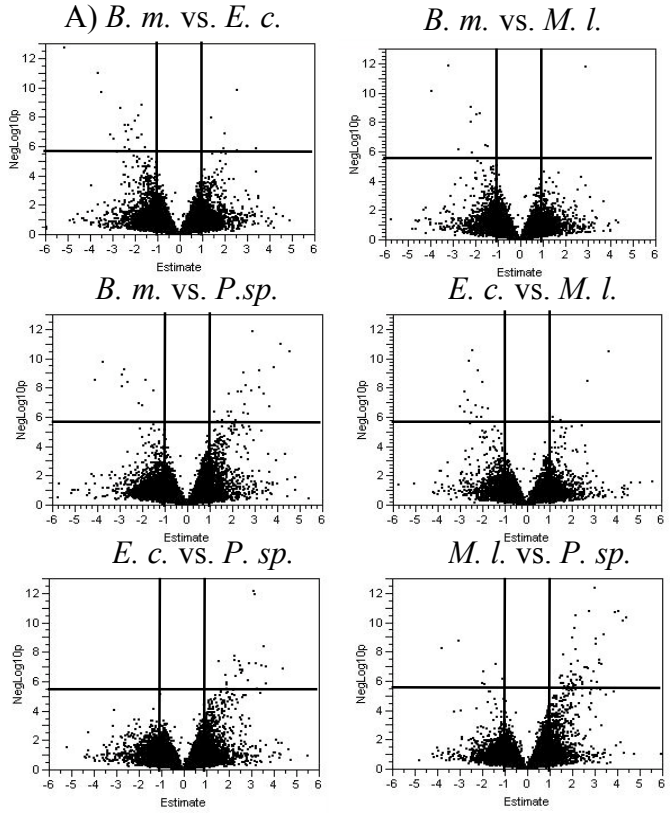


Figure 3.5 *C. elegans* differential gene expression in different bacterial environments
(A) Volcano plots are shown for each microarray comparison. For each, $-\log_{10}$(p-value) from statistical tests is plotted on the Y-axis and $\log_2$(fold change) is plotted on the X-axis. Data points represent the response of all the genes present on the microarrays used, with each point representing a single gene. Points above the horizontal line are significant at the false discovery rate $q < 0.01$. The two vertical lines show 2 fold up or down regulation in the treatment (relative to the second listed bacteria in graph titles). Multiple genes were significantly differentially expressed in response to the different bacterial environments. (B) Gene Ontology (GO) terms were amended with recently published information that was and not yet added to the annotation of particular genes and used to categorize the identified differentially expressed genes. Clustering was done manually by grouping GO terms of similar function.



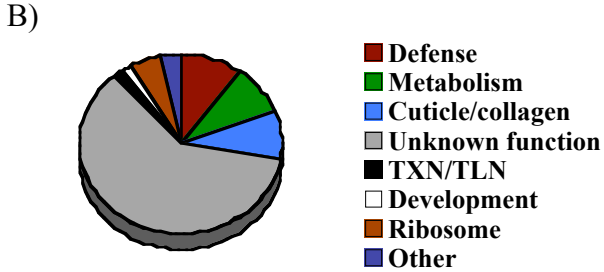


Figure 3.6 Reaction norms of the mutant strains to the different bacterial environments

(A) Brood size was measured for wild type (**bold**, black) and mutant strains in all four bacterial environments and reaction norms are shown for those with significant genotype by environment interactions. (B) Generation time was calculated as $T = (\Sigma x l_x m_x)/(\Sigma l_x m_x)$ using life tables. Generation time reaction norms are shown in days for those with significant genotype by environment interactions.

A) Brood size by bacterial environment

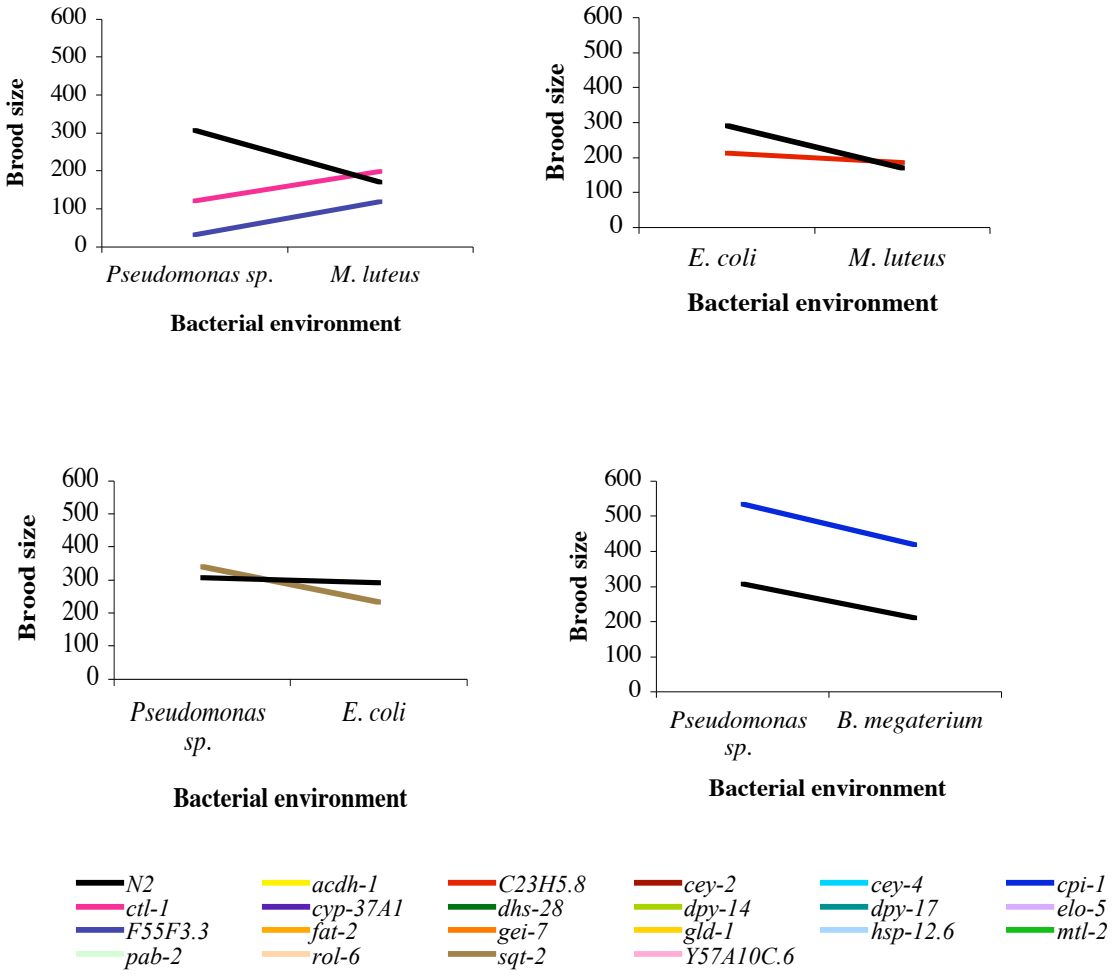


Figure 3.6 Reaction norms of the mutant strains to the different bacterial environments

(A) Brood size was measured for wild type (**bold**, black) and mutant strains in all four bacterial environments and reaction norms are shown for those with significant genotype by environment interactions. (B) Generation time was calculated as $T = (\Sigma x l_x m_x)/(\Sigma l_x m_x)$ using life tables. Generation time reaction norms are shown in days for those with significant genotype by environment interactions.

B) Generation time by bacterial environment

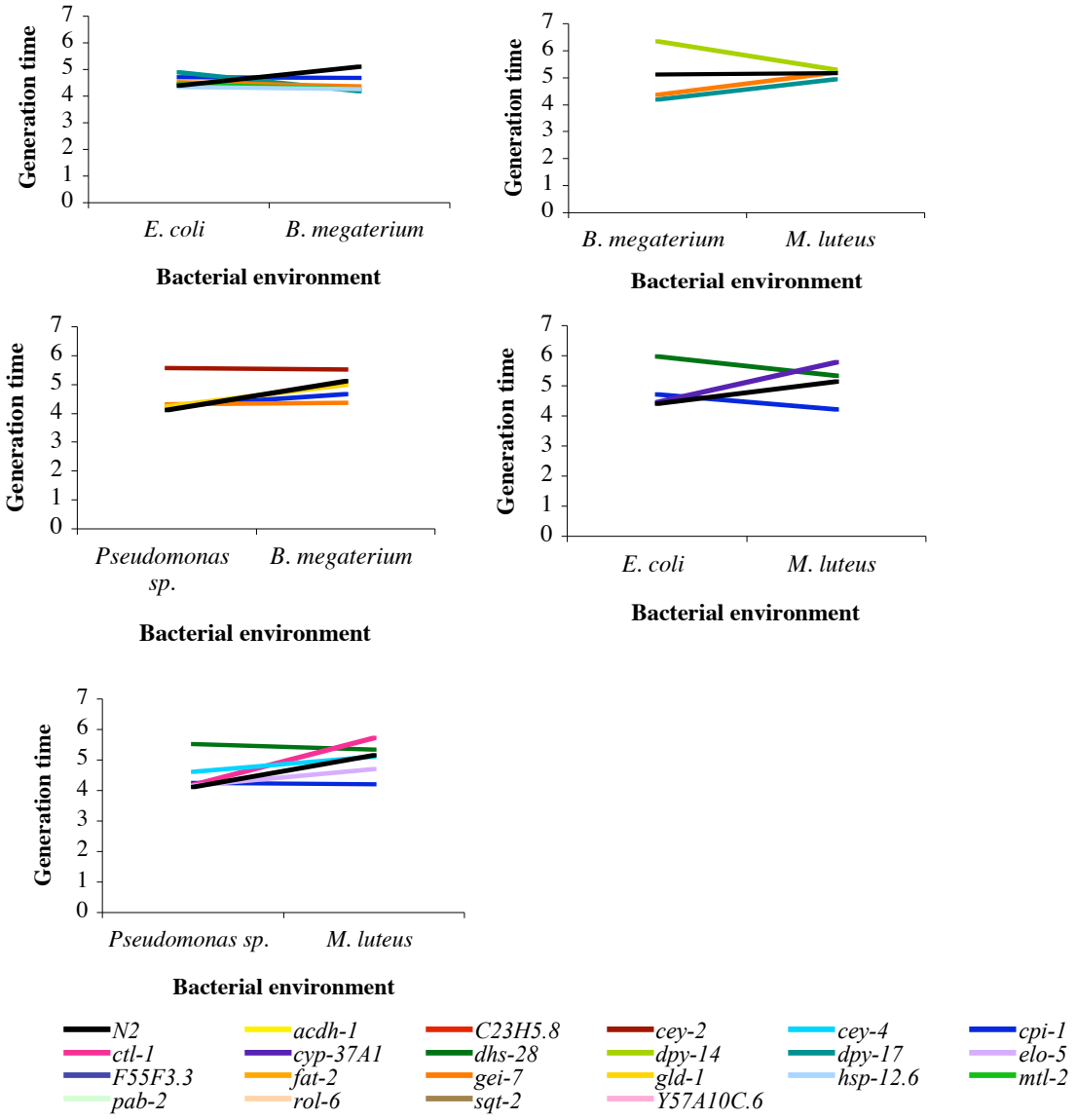


Figure 3.7 Life history reaction norms with significant gene by environment interaction

Significant gene by environment interactions of Lambda (A) and lifespan as measured by TD_{50} (B) are illustrated by reaction norms. All pairwise bacterial comparisons are shown (*B. megaterium* vs. *E. coli*, *B. megaterium* vs. *M. luteus, B. megaterium* vs. *Pseudomonas sp., E. coli* vs. *Pseudomonas sp., E. coli* vs. *M. luteus,* and *M. luteus* vs. *Pseudomonas sp*.).

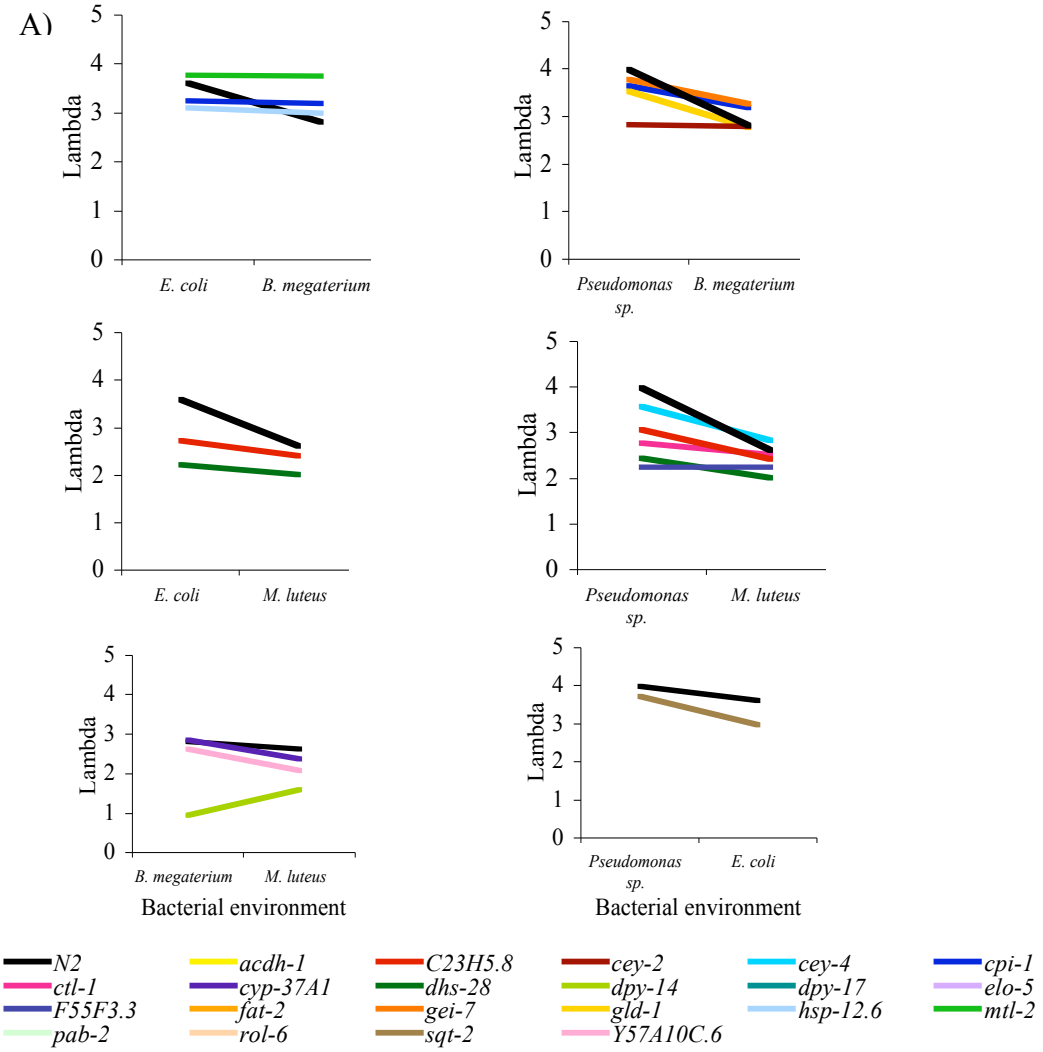


Figure 3.7 Reaction norms of the mutant strains to the different bacterial environments

Significant gene by environment interactions of Lambda (A) and lifespan as measured by TD_{50} (B) are illustrated by reaction norms. All pairwise bacterial comparisons are shown (*B. megaterium* vs. *E. coli*, *B. megaterium* vs. *M. luteus, B. megaterium* vs. *Pseudomonas sp., E. coli* vs. *Pseudomonas sp., E. coli* vs. *M. luteus,* and *M. luteus* vs. *Pseudomonas sp*.).

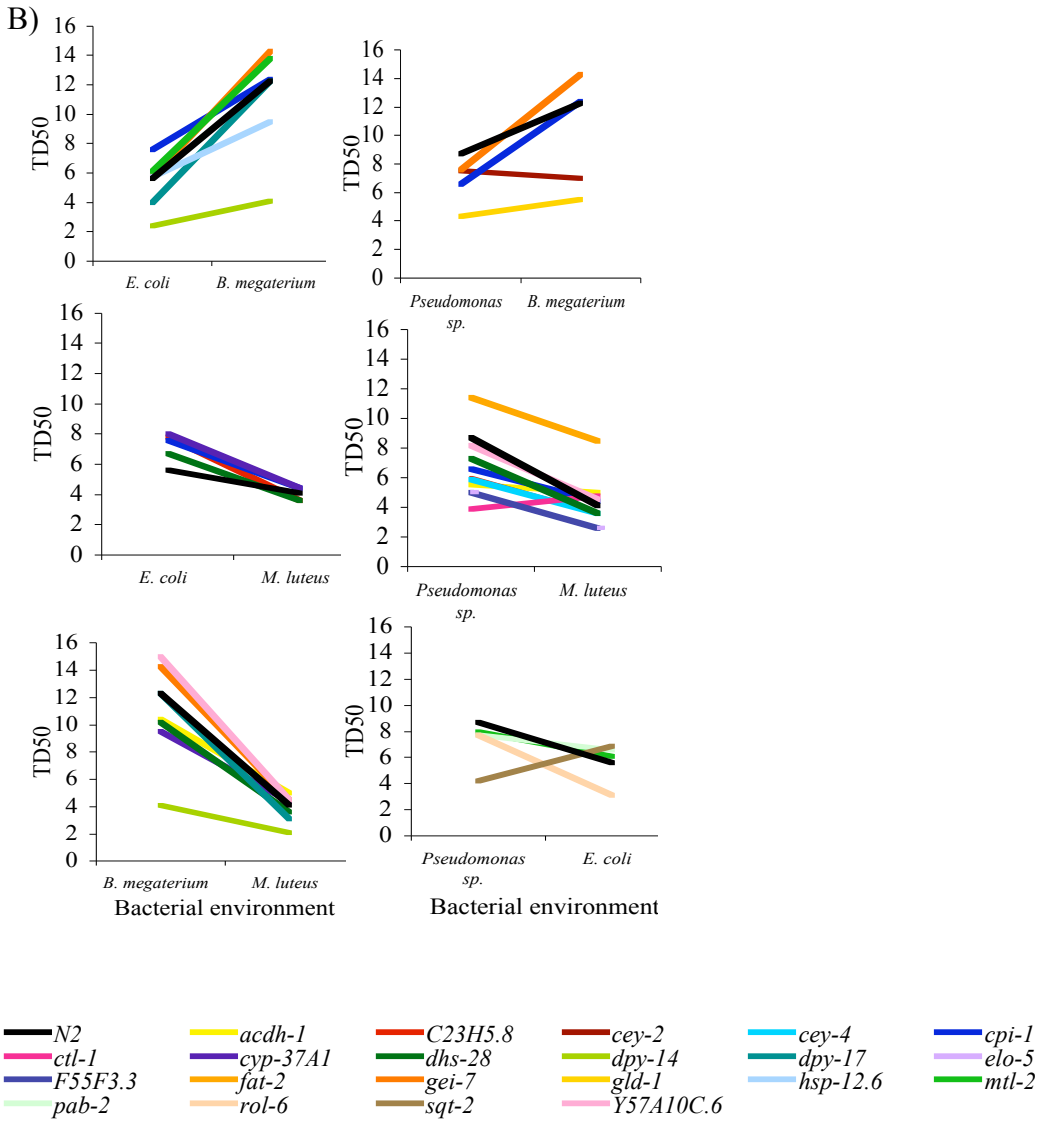


Figure 3.8 *hsp-12.6* proportional change in fitness

Proportional changes in *hsp-12.6* fitness were calculated as $(\mu_{N2}-\mu_{hsp\text{-}12.6})/\mu_{N2}$ by bacterial environment to make fitness relative to wild type. Letters indicate significantly different means ($P > 0.05$ from ANOVA).

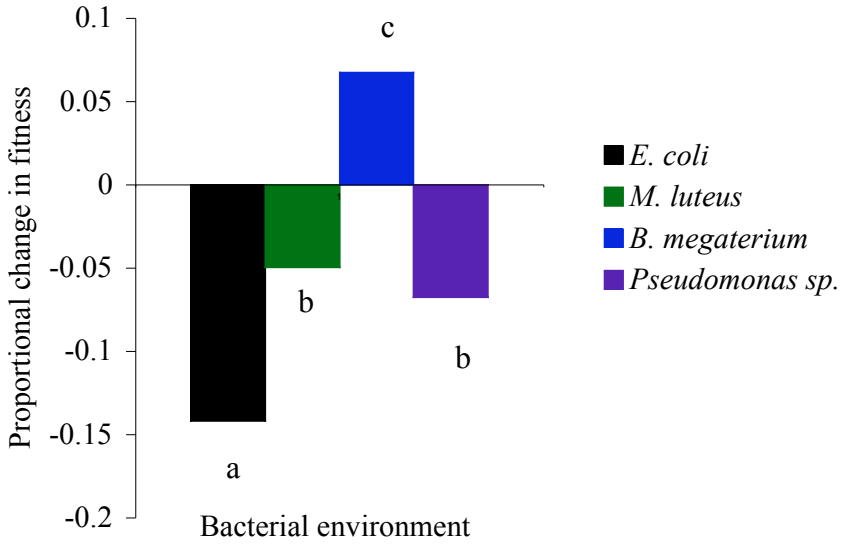


Figure 3.9 Proportional change in *rol-6* TD_{50}

Proportional changes in *rol-6* longevity (measured as TD_{50}) relative to wild type calculated as $(\mu_{N2}-\mu_{rol-6})/\mu_{N2}$ by bacterial environment. Letters indicate significantly different means ($P > 0.05$ from ANOVA).

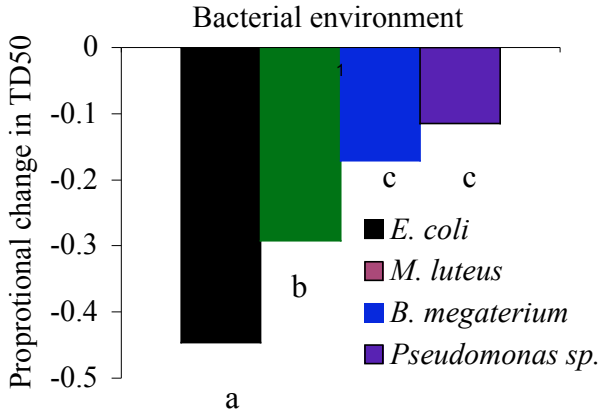


Table 3.1 Biological validation of identified genes

Wild-type *C. elegans* (N2) and mutant strains were grown on the four bacterial isolates and absolute fitness (λ) and time to death for 50% of the individuals in a population (TD_{50} in days) were measured. Standard error (s.e.m.) is given in parenthesis. Additionally, + indicates a significant ($P < 0.05$) increase relative to wild type and – indicates a significant ($P < 0.05$) decrease of the mutant (λ or TD_{50}) relative to wild type.

	E. coli (OP50)		*M. luteus*		*Pseudomonas sp.*		*B. megaterium*	
Gene	λ	TD_{50}	λ	TD_{50}	λ	TD_{50}	λ	TD_{50}
+	3.60(0.19)	5.6(0.22)	2.63(0.18)	4.1(0.22)	3.99(0.25)	8.7(0.27)	2.81(0.16)	12.3(0.27)
acdh-1	2.99(0.03)$^{-}$	5.0(0.35)$^{-}$	2.54(0.25)	5.0(0.35)$^{+}$	3.78(0.74)	5.5(0.79)$^{-}$	3.01(0.37)	10.4(0.42)$^{-}$
C23H5.8	2.72(0.03)$^{-}$	7.8(0.57)$^{+}$	2.42(0.04)$^{-}$	3.6(0.42)$^{-}$	3.07(0.02)$^{-}$	6.0(0.79)$^{-}$	3.30(0.04)$^{+}$	8.9(0.74)$^{-}$
cey-2	3.08(0.04)$^{-}$	6.1(0.42)	2.11(0.06)$^{-}$	3.5(0.35)$^{-}$	2.83(0.03)$^{-}$	7.5(0.61)$^{-}$	2.79(0.01)	7.0(0.35)$^{-}$
cey-4	3.51(0.13)	5.6(0.42)	2.84(0.06)$^{+}$	3.6(0.42)$^{-}$	3.57(0.07)$^{-}$	5.9(0.22)$^{-}$	2.95(0.02)	3.7(0.27)$^{-}$
cpi-1	3.25(0.15)$^{-}$	7.6(0.22)$^{+}$	3.01(1.17)	4.4(0.22)	3.65(0.43)	6.6(0.42)$^{-}$	3.19(0.41)	12.4(0.42)
ctl-1	2.91(0.07)$^{-}$	6.2(0.84)	2.53(0.07)	4.8(0.29)$^{+}$	2.77(0.18)$^{-}$	3.9(0.42)	2.29(0.07)$^{-}$	8.5(0.35)$^{-}$
cyp-37A1	3.59(0.08)	8.0(0.50)$^{+}$	2.37(0.06)$^{-}$	4.4(0.42)	3.64(0.03)$^{-}$	8.5(0.35)$^{-}$	2.85(0.04)	9.5(0.50)$^{-}$
dhs-28	2.23(0.18)$^{-}$	6.7(0.27)$^{+}$	2.01(0.21)$^{-}$	3.6(0.22)$^{-}$	2.43(0.14)$^{-}$	7.3(0.27)$^{-}$	1.86(0.27)$^{-}$	10.2(0.76)$^{-}$
dpy-14	1.89(0.44)$^{-}$	2.4(0.22)$^{-}$	1.60(0.07)$^{-}$	2.1(0.22)$^{-}$	1.85(0.17)$^{-}$	3.1(0.42)$^{-}$	0.96(0.02)$^{-}$	4.1(0.42)$^{-}$
dpy-17	2.84(0.52)$^{-}$	4.0(0.35)$^{-}$	2.70(0.34)	3.1(0.42)$^{-}$	3.20(0.45)$^{-}$	3.0(0.35)$^{-}$	2.69(0.80)	12.3(0.57)
elo-5	4.11(0.07)$^{+}$	5.5(0.35)	3.02(0.10)$^{+}$	2.6(0.42)$^{-}$	4.07(0.12)	5.0(0.50)$^{-}$	4.18(0.05)$^{+}$	9.5(0.35)$^{-}$
F55F3.3	3.53(0.15)	3.1(0.55)$^{-}$	2.25(0.14)$^{-}$	2.6(0.55)$^{-}$	2.24(0.07)$^{-}$	5.0(0.35)$^{-}$	2.06(0.07)$^{-}$	5.5(0.35)$^{-}$
fat-2	3.27(0.13)$^{-}$	9.9(0.82)$^{+}$	2.97(0.04)$^{+}$	8.5(0.35)$^{+}$	4.23(0.04)	11.4(0.74)$^{+}$	3.18(0.09)$^{+}$	13.7(1.15)$^{+}$
gei-7	3.52(0.25)	5.7(0.27)	2.73(0.12)	4.5(0.00)$^{+}$	3.77(0.26)	7.6(0.22)$^{-}$	3.27(0.48)	14.3(0.27)$^{+}$
gld-1	3.15(0.13)$^{-}$	5.6(0.22)	2.51(0.28)	3.5(0.35)$^{-}$	3.53(0.06)$^{-}$	4.3(0.57)$^{-}$	2.78(0.04)	5.5(0.35)$^{-}$
hsp-12.6	3.10(0.08)$^{-}$	5.7(0.45)	2.50(0.18)	3.7(0.27)$^{-}$	3.72(0.14)	6.6(1.29)$^{-}$	3.00(0.08)$^{+}$	9.5(1.00)$^{-}$
mtl-2	3.77(0.17)	6.1(0.22)$^{+}$	3.02(0.23)$^{+}$	5.2(0.27)$^{+}$	4.09(0.28)	8.0(0.35)$^{-}$	3.75(0.40)$^{+}$	13.8(0.27)$^{+}$
pab-2	4.14(0.06)$^{+}$	6.6(0.42)$^{+}$	2.72(0.47)	5.4(0.42)$^{+}$	4.29(0.24)	7.7(0.57)$^{-}$	3.20(0.11)$^{+}$	8.9(0.74)$^{-}$
rol-6	2.82(0.22)$^{-}$	3.1(0.82)$^{-}$	2.28(0.11)$^{-}$	2.9(0.22)$^{-}$	3.11(0.09)$^{-}$	7.7(0.45)$^{-}$	2.56(0.04)$^{-}$	10.2(0.76)$^{-}$
sqt-2	2.97(0.01)$^{-}$	6.9(0.42)$^{+}$	2.69(0.06)	3.7(0.57)	3.72(0.06)$^{-}$	4.2(0.57)$^{-}$	3.39(0.47)$^{+}$	7.2(1.35)$^{-}$
Y57A10C.6	3.37(0.18)	6.3(0.45)$^{+}$	2.09(0.09)$^{-}$	4.5(0.00)$^{+}$	3.41(0.33)$^{-}$	8.2(0.57)	2.62(0.23)	15.0(0.35)$^{+}$

Table 3.2 Functional tests of brood size and generation time

Wild-type *C. elegans* and mutant strains were grown in the four bacterial environments and brood size and generation time were determined. Generation time was determined as $T = (\Sigma x l_x m_x)/(\Sigma l_x m_x)$ (in days) using life tables. Standard error (s.e.m.) is given in parenthesis, and significant differences between mutant and wild type is denoted by + or - ($P < 0.05$) following values. Additionally + indicates an increase relative to wild type and a – indicates a decrease relative to wild type.

	E. coli (OP50)		*M. luteus*		*Pseudomonas sp.*		*B. megaterium*	
Gene	Brood	T	Brood	T	Brood	T	Brood	T
+	290.80(60.87)	4.40(0.19)	171.00(70.12)	5.15(0.36)	307.00(49.16)	4.11(0.25)	210.83(57.22)	5.10(0.20)
acdh-1	283.20(43.15)	5.14(0.16)	172.67(53.43)	5.46(0.55)	290.83(104.7)	4.32(0.39)	184.83(38.73)	4.80(0.59)
C23H5.8	212.60(9.92)$^-$	5.35(0.04)$^+$	186.00(7.42)	5.93(0.07)$^+$	313.40(21.03)	5.13(0.03)$^+$	239.20(17.61)	4.58(0.04)$^-$
cey-2	348.20(18.03)	5.20(0.07$^+$	164.80(31.98)	6.80(0.18)$^+$	328.00(10.03)	5.56(0.04)$^+$	289.00(10.37)$^+$	5.52(0.04)$^+$
cey-4	313.80(16.48)	4.58(0.14)	208.60(3.29)	5.12(0.09)	356.60(12.26)	4.62(0.08)$^+$	363.40(19.73)$^+$	5.45(0.05)$^+$
cpi-1	260.20(46.09)	4.71(0.13)$^+$	122.00(101.4)	4.21(0.49)$^-$	228.17(27.38)$^-$	4.24(0.42)	207.50(14.94)	4.68(0.54)
ctl-1	177.00(22.99)$^-$	4.84(0.06)$^+$	198.40(8.96)	5.71(0.15)$^+$	120.20(58.98)$^-$	4.17(0.18)	73.00(17.21)$^-$	4.77(0.15)$^-$
cyp-37A1	301.00(18.06)	4.46(0.03)	149.00(14.41)	5.78(0.19)$^+$	311.40(11.06)	4.44(0.05)$^+$	298.60(58.54)$^+$	5.42(0.22)
dhs-28	129.60(53.53)$^-$	5.99(0.16)$^+$	46.50(23.38)$^-$	5.34(0.46)	138.33(45.63)$^-$	5.51(0.22)$^+$	30.00(20.92)$^-$	5.80(1.08)$^+$
dpy-14	53.60(41.51)$^-$	5.44(0.36)$^+$	16.60(15.26)$^-$	5.29(0.21)	38.40(18.50)$^-$	4.93(0.13)$^+$	0.60(0.55)$^-$	6.33(0.58)$^+$
dpy-17	125.83(15.87)	4.90(1.12)	144.83(45.08)	4.93(0.39)	173.00(113.0)$^-$	4.19(0.42)	142.83(137.8)	4.19(0.46)$^-$
elo-5	276.00(13.42)	3.97(0.05)$^-$	179.00(8.86)	4.70(0.10)$^-$	342.60(13.58)	4.16(0.08)	247.00(7.58)	3.85(0.02)$^-$
F55F3.3	198.00(7.11)$^-$	4.19(0.11)	118.20(56.23)	5.27(0.46)	31.20(4.66)$^-$	4.26(0.10)	72.40(15.01)$^-$	5.91(0.26)$^+$
fat-2	238.20(44.79)	4.61(0.10)	225.60(13.18)	4.98(0.08)	328.40(9.50)	4.02(0.02)	283.40(18.64)$^+$	4.89(0.11)
gei-7	299.20(26.99)	4.54(0.20)	186.67(56.70)	5.18(0.34)	302.33(40.83)	4.31(0.22)	200.67(85.45)	4.36(0.29)$^-$
gld-1	222.40(19.19)$^-$	4.71(0.10)$^+$	81.40(27.65)$^-$	4.77(0.22)	215.00(50.91)$^-$	4.24(0.27)	165.80(14.94)	5.00(0.05)
hsp-12.6	136.40(12.99)$^-$	4.34(0.04)	97.40(21.13)	4.99(0.15)	135.40(27.93)$^-$	3.72(0.07)$^-$	109.80(14.72)$^-$	4.27(0.08)$^-$
mtl-2	361.20(42.44)	4.44(0.17)	244.67(32.66)$^+$	5.00(0.41)	346.00(28.64)	4.16(0.21)	269.50(42.79)	4.26(0.26)$^-$
pab-2	300.20(11.65)	4.01(0.03)$^-$	100.20(58.38)	4.04(0.30)$^-$	304.00(8.12)	3.94(0.14)	279.40(16.82)$^+$	4.84(0.10)$^-$
rol-6	162.00(22.36)$^-$	4.74(0.27)	200.60(27.44)	6.43(0.23)$^+$	211.60(23.57)$^-$	4.72(0.02)$^+$	257.60(18.39)	5.91(0.03)$^+$
sqt-2	233.80(8.11)	5.01(0.04)$^+$	180.60(12.14)	5.25(0.10)	339.00(12.57)	4.44(0.03)$^+$	218.00(107.1)	4.20(0.12)$^-$
Y57A10C.6	267.80(43.61)	4.60(0.13)	60.00(80.02)$^-$	4.98(0.87)	256.67(12.74)$^-$	4.41(0.29)	184.33(40.76)	5.30(0.51)

Table 3.3 Identification of differentially expressed genes

Statistical analysis was performed using a mixed model ANOVA. False discovery rate ($q < 0.01$) was used as a significance threshold for identification of differentially expressed genes. Because there is some overlap of differentially expressed genes identified between the six comparisons, the number of unique genes identified is also included.

Comparison	# of genes identified
B. megaterium vs. *E. coli* (OP50)	55
B. megaterium vs. *M. luteus*	25
B. megaterium vs. *Pseudomonas sp*.	81
E. coli (OP50) vs. *M. luteus*	41
E. coli (OP50) vs. *Pseudomonas sp.*	62
M. luteus vs. *Pseudomonas sp.*	108
TOTAL	372
UNIQUE	204

Table 3.4 qPCR primers used in validation of microarray results

Genes listed were identified in microarray comparisons as significantly differentially expressed in response to the bacterial environment. For each gene used for validation, primer pairs were made using BeaconDesigner3 (BioRad laboratories) for use in qPCR. Housekeeping genes were selected based on very low variance and lack of treatment effects from our microarray data.

Gene	Forward primer	Reverse primer
lys-4	TTCATTCAACCAGTGTCTAC	TCAAAGCATCAAGAGTTTCC
ilys-3	CCCTTTCTTGTGGATATTATCAG	GACTCTTGTAGCGGTTGTAG
cpi-1	TGTCTGATGTGAATGCCTCTG	TGACTCTCCAACAAGAACTTCC
dpy-17	CACAAGTCTATCAGCAAGTC	CTACCGTATCCACCATATCC
gei-7	TTCGTCACTTGAATGCGTCTC	TCGTTCTTCCAGCCTCGTAG
mtl-2	CTGCCAGTGAGAAGAAATGC	CGAACAATATCAATTAGTAGGAATTTG
dhs-28	TGCTTCTTCTGGTGCTTCTTC	GCTCGTTCTTTCCGTCAGTG
Y57A10C.6	ACATCGTTGGAGTCGGTATG	GCAATCATCAAGAGCAGTAGTG
act-4	AAGTGCGATTAGGATGAACGG	AGGACTGGATGCTCTTCTGG
acdh-1	TGCCTCGTCTCTGTTCTGATAGTCTTG	ACCTGTGCCTCTCCTGAATTAGTAATCC
sodh-1	ATTGGTTGGAGGACACGAAGGAG	CGCAGTTGAGGCAGTTGAAGTTC
F41F3.3	CCACAACTCCCACCACTTTC	TCCGTATCCTCCTCCGATTG
Housekeeping		
ubq-1	CACTTGGTTCTTCGTCTTAG	CCTCCTTGTCTTGAATCTTG
gdp-2	CAATGTTCGTCGTCGGAGTC	CAAGTGGAGCAAGGCAGTTAG
nhx-4	AACAACAACAACAACACCTCAG	AAGCATTAGGACTACCGATTCAG
F32A7.4	TCTGCTGCTGCCTATTCTCTG	GCTTCACTGGACTGTTCGTTC

Table 3.5 qPCR validation results

We used quantitative reverse transcriptase real time PCR to validate the results of the microarray experiments. Three new biological replicates that were not used in the microarray experiments were used for validation. cDNA was synthesized from RNA samples using a two-step iScript cDNASynthesis Kit (BioRad Laboratories) and these three replicate cDNA stocks for each bacterial environment were used for all subsequent tests. qPCR was performed with a Bio-Rad icycler (Bio-Rad Laboratories) using transcript specific primers shown in Table 3.4. PCR reaction parameters were optimized as needed and housekeeping genes were used to standardize and calculate ΔCT values. From this ΔΔCT values were calculated and are shown in in the columns labeled qPCR. Microarray expression differences are shown for comparison. Melt curve analyses were performed to test for specific amplification. In all cases amplification was specific.

	B. megaterium vs. *E. coli*		*B. megaterium* vs. *M. luteus*		*B. megaterium* vs. *Pseudomonas*		*E. coli* vs. *Pseudomonas*		*E. coli* vs. *M. luteus*		*M. luteus* vs. *Pseudomonas*	
Gene	Array	qPCR	Array	qPCR	Array	qPCR	Array	qPCR	Array	qPCR	Array	qPCR
lys-4	1.814	2.01	NS	NS	4.9792	4.43	3.1652	3.03	1.8028	2.12	4.9681	5.31
ilys-3	NS	NS	1.7127	1.98	2.5388	2.12	2.2666	2.32	1.9848	1.88	4.2514	4.11
cpi-1	3.6191	4.03	NS	NS	2.8718	2.77	NS	NS	2.7359	2.53	1.9885	2.61
dpy-17	1.6517	1.72	1.1972	1.45	NS	NS	NS	NS	NS	NS	NS	NS
gei-7	1.8005	1.67	1.1629	1.28	1.8091	2.21	NS	NS	NS	NS	NS	NS
mtl-2	3.0595	2.89	NS	NS	NS	NS	2.1706	2.22	NS	NS	NS	NS
dhs-28	NS	NS	1.5478	1.67	NS	NS	NS	NS	1.9573	1.92	2.0024	2.53
Y57A10C.6	NS	NS	2.1269	2.03	NS	NS	NS	NS	2.6368	2.33	3.0965	3.62
act-4	NS	NS	NS	NS	1.8426	1.55	2.5454	2.13	NS	NS	2.0024	2.16
acdh-1	NS	NS	2.7392	2.56	NS	NS	NS	NS	NS	NS	2.7293	2.48
sodh-1	3.4823	3.06	NS	NS	2.8544	3.98	NS	NS	3.6534	5.31	3.0254	6.17
F41F3.3	NS	NS	NS	NS	NS	NS	2.7586	1.98	NS	NS	3.1077	2.68

Table 3.6 Functional categories of identified genes

Functional categories of the 204 identified genes were clustered based on gene ontology and manual grouping of identified genes ($q < 0.01$). TXN = transcription and TLN = translation.

Category	# Identified	% total
Unknown function	124	61
Defense	20	9.8
Metabolism	19	9.3
Cuticle/collagen	18	8.8
Ribosome	10	4.9
Other	7	3.4
TXN/TLN	3	1.5
Development	3	1.5

Table 3.7 q<0.01 genes mapped to co-expression matrix mountains

Microarray identified genes were mapped to the *C. elegans* co-expression matrix for all *C. elegans* genes to group genes by similar function. Number of genes overlapping between a given mountain and identified genes is listed along with a representation factor calculated at (http://elegans.uky.edu/gl/cgi-bin/gene_list.cgi) and the associated p-value according to the likelihood of the overlap by chance given the size of the genome, and the number of genes in the lists compared.

Topomap cluster	Genes in common	Representation factor	Bonferroni corrected p-value	Annotation/Biogroup
mount00	1	0.0	5.555e-12	Other/collagen
mount01	4	0.2	4.922e-05	Amine oxidase
mount02	4	0.2	0.002	Cyclin
mount04	3	0.2	0.007	RNA Pol II transcription
mount05	3	0.3	0.071	Chromatin
mount07	5	0.5	1.000	Tc4-Tc5
mount08	32	3.4	7.942e-08	Mitosis/defense
mount11	2	0.3	0.781	DNA repair
mount12	2	0.4	1.000	Cell cycle
mount14	1	0.2	1.000	DNA synthesis
mount15	10	3.5	0.061	G-protein coupled receptor
mount16	11	4.1	0.008	Retinoblastoma complex
mount17	2	1.0	1.000	tRNA synthesis
mount18	1	0.5	1.000	Cell migration
mount19	14	6.4	4.159e-06	Glycolysis
mount20	2	1.0	1.000	Topoisomerase
mount21	3	1.7	1.000	Germ-line enriched
mount22	25	14.3	4.743e-20	Carbohydrate metabolism
mount23	10	6.0	6.209e-04	Ras pathway
mount24	4	2.6	1.000	Neuronal
mount25	1	1.0	1.000	Chemosensation
mount27	3	3.0	1.000	Oocyte enriched
mount29	1	2.2	1.000	Nucleotide synthesis
mount31	1	3.5	1.000	Proteases
mount34	1	5.1	1.000	Hermaphrodite enriched

Table 3.8 Mutants used for functional tests

List of 21 mutants used for functional tests, predicted molecular functions are indicated.

Gene	Allele	Predicted molecular function
acdh-1	*ok1489*	Acyl-CoA dehydrogenase
C23H5.8	*ok651*	Unknown function
cey-2	*ok902*	Cold-shock/Y-box domain containing
cey-4	*ok858*	Unknown function
cpi-1	*ok1213*	Homolog of cysteine protease inhibitors (cystatins)
ctl-1	*ok1242*	Cytosolic catalase
dhs-28	*ok450*	17-Beta-hydroxysteroid dehydrogenase 4
dpy-14	*e188*	Type III (alpha 1) collagen
dpy-17	*e1295*	Cuticle collagen
elo-5	*gk182*	PUFA elongase
cyp-37A1	*ok673*	Unknown function
F55F3.3	*ok1758*	Unknown function
fat-2	*ok873*	Delta-12 fatty acyl desaturase
gei-7	*ok531*	Predicted isocitrate lyase/malate synthase
gld-1	*op236*	Meiotic cell cycle/oogenesis
hsp-12.6	*gk156*	Predicted heat shock protein
mtl-2	*gk125*	Metallothionein
pab-2	*ok1851*	Polyadenylate-binding protein
rol-6	*e187*	Cuticle collagen
sqt-2	*sc108*	Cuticle collagen
Y57A10C.6	*ok693*	Predicted thiolase

Table 3.9 Directionality of mutant life history response by bacterial environment
Mutant life history traits: Fitness, Lifespan, Brood size and Generation time were determined in each bacterial environment (Tables 3.1, 3.2) and using analysis of variance difference from wild type was determined for each mutant in each environment. The number of the 21 mutants tested that were significantly different ($P < 0.05$) from wild type is shown and the directionality of mutational effects is indicated.

		E. coli	*M. luteus*	*Pseudomonas sp.*	*B. megaterium*
Fitness (λ)	# up	2	4	0	7
	# down	13	8	13	5
	total	15	12	13	12
Lifespan (TD_{50})	# up	9	7	1	4
	# down	5	11	18	15
	total	14	18	19	19
Brood size	# up	0	1	0	5
	# down	8	4	10	5
	total	8	5	10	10
Generation time (T)	# up	9	5	8	6
	# down	2	3	1	9
	total	11	8	9	15

Table 3.10 Significance of gene by environment interactions

Using all available strains with mutant alleles from the CGC we tested 37 hypotheses generated from the differential expression of genes in the different bacterial environments. Each gene is listed along with the direction and magnitude of differential expression (Log_2fold change) from microarray comparisons (E = *E. coli*, M = *M. luteus*, P = *Pseudomonas sp.*, B = *B. megaterium*). Significance of Genotype by Environment Interactions (GEI) for fitness and TD_{50} between bacterial environments (shown in column 2) is in columns 4 and 5.

Gene	Microarray contrast	Log_2(Fold change)	Fitness GEI *P*-value	TD_{50} GEI *P*-value
acdh-1	B < M	2.74	0.196907	4.81E-08
	M > P	2.73	0.706524	3.63E-08
C23H5.8	E < M	1.13	0.0000277	7.76E-07
	M > P	1.2	0.0000836	0.000104
cey-2	B > P	1.76	1.25E-07	4.47E-09
cey-4	M > P	1.65	0.000345	1.86E-07
cpi-1	B < E	3.62	0.003999	2.18E-06
	B < P	2.87	0.015478	3.23E-06
	E > M	2.74	0.168364	2.50E-07
	M < P	1.99	0.163894	1.05E-07
ctl-1	M > P	1.54	1.55E-06	6.56E-12
cyp-37A1	B < M	1.99	0.017136	6.97E-08
	E < M	2.06	0.064124	7.58E-06
	M > P	2.45	0.505982	0.105647
dhs-28	B < M	1.55	0.066303	0.000784
	E < M	1.96	0.00018	1.18E-06
	M > P	2	0.0000113	0.000979
dpy-14	B < E	1.39	0.572207	2.29E-12
	B < M	1.21	0.0000585	8.21E-14
dpy-17	B < E	1.65	0.195549	0.000232
	B < M	1.2	0.601654	0.012109
elo-5	M > P	1.75	0.052139	5.75E-06
F55F3.3	M > P	2.08	4.35E-08	5.75E-06
fat-2	M > P	1.81	0.485521	0.000611
gei-7	B < E	1.8	0.053643	4.71E-07
	B < M	1.16	0.131696	5.55E-07
	B < P	1.81	0.014723	5.01E-10
gld-1	B > P	1.26	0.00614	0.0000034
hsp-12.6	B < E	2.42	0.0000208	0.0000374
mtl-2	B < E	3.06	0.002662	0.000385
	E > P	2.17	0.725084	0.000161
pab-2	E > P	1.75	0.183789	0.0000357
rol-6	E > P	2.69	0.602943	0.004032
sqt-2	E > P	1.41	0.025087	1.98E-11
Y57A10C.6	B < M	2.13	0.042391	1.85E-08
	E < M	2.64	0.062286	0.238398
	M > P	3.1	0.824269	0.008479

CHAPTER 4 - DAF-2/Insulin and TOL-1 regulated longevity is environment specific

Abstract

Use of more natural environments for genetic and phenotypic analysis is crucial for understanding gene function. Here we found this is true of the well-studied *C. elegans* defense pathways. *daf-2* mutants have been found to be resistant to infection by all bacteria tested to date through de-repression of the transcription factor DAF-16, which leads to up-regulation of antimicrobial effectors. Here we show that the normally long-lived *daf-2* mutant has reduced lifespan on different bacterial types. We found that DAF-16 plays a role in pathogen resistance and we also show that TOL-1 is required for defense against the Gram-positive bacterium *Bacillus megaterium*.

Materials and Methods

C. elegans and bacteria strains and maintenance

The following mutant strains were used: Bristol (*N2*), *dbl-1(nk3), daf-2(e1368), daf-16(mgDf50), tol-1(nr2033), pmk-1(km25),* and *sek-1(km4)*. Growth and maintenance conditions were as described (Brenner 1974, Sulston and Hodgkin 1988).

Pathogenicity assays

Longevity assays were performed as previously described (Tan et al. 1999a, Tan and Ausubel 2000) and TD_{50} was calculated from survivorship curves as time to death for 50% of individuals in a population. Briefly, worms were synchronized by bleaching to collect eggs and hatched in M9 overnight. Worms were then grown to L4 on *E. coli* (OP50) to standardize test populations, and then transferred to the test bacteria as young adults (20 worms per plate) and were maintained at 25°C. Surviving worms were then re-plated daily and the fraction surviving was determined every 12 hours. Worms were determined to be dead when they no longer responded to touch with platinum wire. All pathogenicity tests were conducted in at least ten independent replicate experiments.

Statistical analysis

Analysis of variance using the GLM procedure in SAS (SAS Institute Inc., Cary, North Carolina, USA) was used for statistical tests. The model used for these tests is shown.

Model: $TD_{50} = \mu$ + genotype + bacteria + genotype X bacteria + error

Results and Discussion

Laboratory studies of genetic model organisms have provided a detailed understanding of the genetic architecture underlying organismal function. Yet, because these functions evolved over millions of generations outside the laboratory and gene function can be environment specific, it is essential to consider the effect of more natural environments on the genetic mechanisms that control these functions (Shimizu and Purugganan 2005, Ungerer et al. 2008). For example, flowering time in *Arabidopsis* is a well studied trait, but quantitative trait loci (QTL) that affected flowering time in laboratory experiments showed non-overlap with QTL identified in field experiments (Weinig et al. 2003), suggesting that additional gene functions are can be revealed in more natural settings. We wondered whether this might also be true of other well-studied genetic pathways, especially those that govern organismal responses and interactions with the environment. To address this question we focused on the multiple *C. elegans* defense pathways involved in resistance to microbial infections (Schulenburg et al. 2004) (See Chapter 6). The insulin signaling pathway *(daf-2/daf-16,* insulin-like receptor/FOXO transcription factor) (Garsin et al. 2003), the p38 MAPK pathway (*sek-1*/*pmk-1*) (Troemel et al. 2006), the TGF-β-like pathway (*dbl-1*) (Mallo et al. 2002) and the toll-like receptor pathway (*tol-1*) (Tenor and Aballay 2008) have all been shown to be involved in the response of *C. elegans* to microbial pathogens (See Chapter 1). These studies have established *C. elegans* as an experimental model for human innate immunity. However, it is unlikely that *C. elegans* evolved in association with these human pathogens raising the possibility that the natural functions of these defense pathways may not have been identified.

While *C. elegans* is an excellent model for human innate immunity, it might be an even better model for the responses of soil nematodes to bacterial environments. Recently we used *C. elegans* to model the responses of grassland soil nematodes to changes in the bacterial community. We isolated bacteria from grassland soils (*Micrococcus luteus, Bacillus magaterium* and a *Pseudomonas sp*.) and found that lifespan was dependent on the bacterial isolate on which it was grown (See Chapter 3). Lifespan on these bacterial isolates differed from that on *C. elegans* commonly used laboratory food source, *E. coli* (OP50) (See Chapter 3). To determine the contributions of each defense pathway in the different soil bacterial environments, we measured survivorship of representative defense pathway mutants on the soil isolates and *E. coli* (Figure 4.1).

daf-2/insulin-like receptor mutants have extended lifespan and increased pathogen resistance to all bacteria tested to date. These include both Gram-positive (*Enterococcus faecalis*, *Staphlococcus aureus*) and Gram-negative (*Pseudomonas aeruginosa*, *Salmonella enterica*) bacterial pathogens, as well as the relatively nonpathogenic *Bacillus subtilis* (Garsin et al. 2003). Similarly we found that *daf-2* mutants have extended lifespan when grown on both *E. coli* and *Pseudomonas sp* (Figure 4.2). Surprisingly, we found that *daf-2* animals are not more resistant to *M. luteus,* as measured by time to death for 50% of a population (TD_{50}). Furthermore, *daf-2* mutants also have reduced lifespan when grown on *B. megaterium*, which is surprising, as wild-type animals have increased lifespan in this environment compared to growth on *E. coli* (OP50) (Figures 4.1, 4.2).

Similar to previous studies, we found that lifespan of *daf-16* mutants grown on *E. coli* was not significantly different from wild type (Garsin et al. 2003, Troemel et al.

2006), and found this was also the case for growth on *M. luteus*. Interestingly, lifespan of *daf-16* animals was reduced when grown on *Pseudomonas sp.* or *B. megaterium* (Figures 4.1, 4.2). Thus, it seems that DAF-16 is important for resistance to these bacteria, plus a larger role in resistance to *B. megaterium*.

tol-1 mutants were previously reported to be susceptible to growth on Gram-negative bacteria (Tenor and Aballay 2008), which is similar to what we found for growth on *E. coli* and our *Pseudomonas sp.* isolate (Figure 4.2). However, whereas *tol-1* mutants were previously reported to be more resistant to certain Gram-positive bacteria, we found that they were not resistant to either of our Gram-positive isolates, *M. luteus* or *B. megatarium*. In fact, *tol-1* mutants were more susceptible than were wild-type animals to *B. megatarium* (Figure 4.2). This suggests that *tol-1* is required for lifespan regulation on both Gram-positive and -negative bacteria in a manner more complex that previously suggested (Tenor and Aballay 2008).

Our results demonstrate that examination of the genetic basis of organismal functions in alternative, and somewhat more natural, environments can provide a more complete picture of gene function, even for well-studied genetic pathways. For example, insulin signaling is a highly conserved pathway that is important in various processes in invertebrate and mammalian organisms. In *C. elegans daf-2*/insulin-like signaling, is important for resistance to bacterial infection, longevity, and dauer larvae formation (Garsin et al. 2003). It has been thought that *daf-2* mutant animals are always long-lived and we have shown that lifespan is environment specific, even for *daf-2* mutants. It will be interesting to see the extent to which other well studied genetic pathways also function

in environment specific manner and how these revelations will help us to better understand genome function.

Acknowledgements

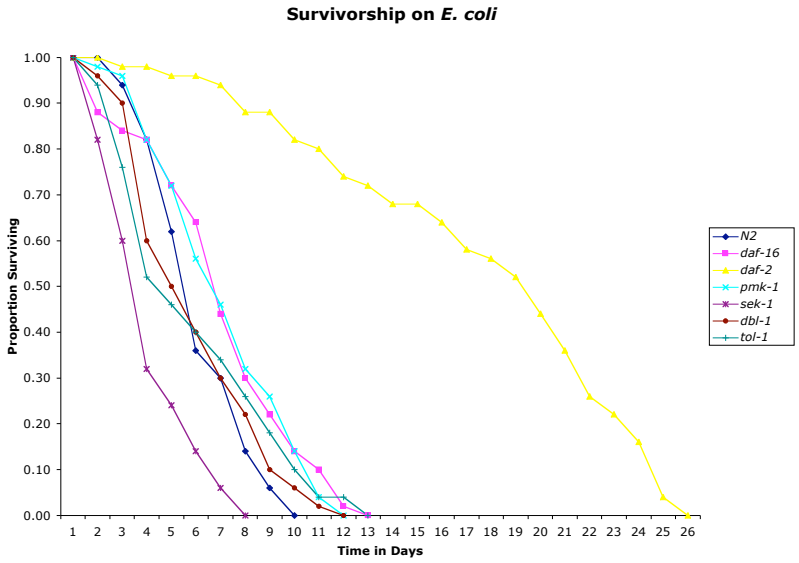
We thank the *Caenorhabditis* Genetics Center for providing strains. Thanks to the members of the Ecological Genomics Institute at Kansas State University for discussions and helpful comments.

Figures and Tables

Figure 4.1 Pathway mutant survivorship curves on soil bacteria

Survival of *C. elegans* strains wild type (*N2*), *daf-16(mgDf50)*, *daf-2(e1368)*, *pmk-1(km25)*, *sek-1(km4)*, *dbl-1(nk3)*, and *tol-1(nr2033)* feeding on A) *E. coli* (OP50), B) *M. luteus*, C) *Pseudomonas sp*., D) *B. megaterium.* Survivorship was determined daily until death. Each combination was repeated 5 times and means are shown.

A)



B)

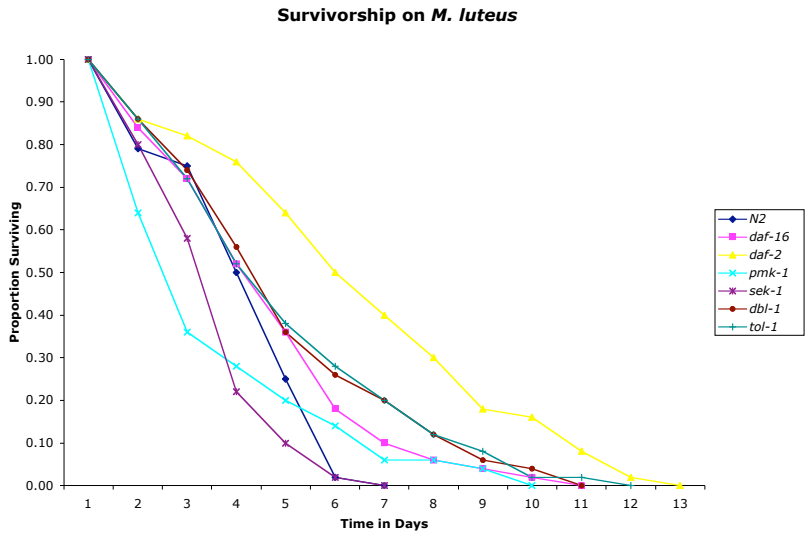
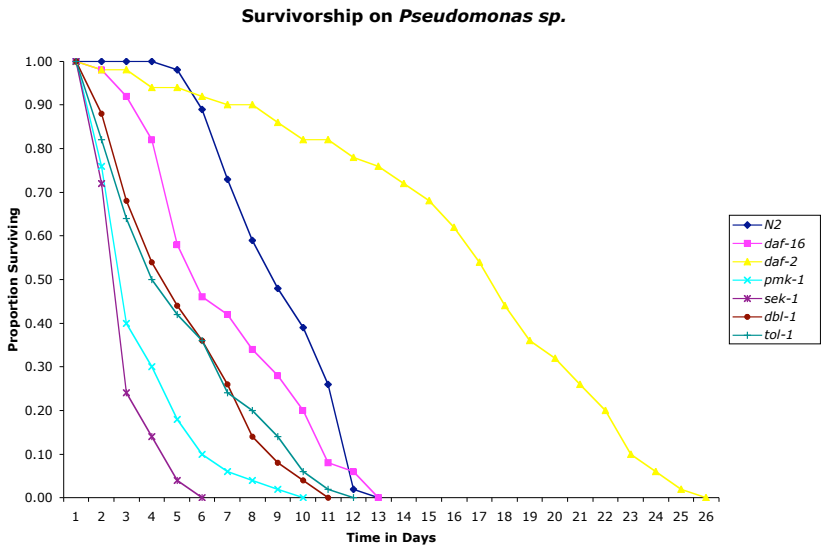


Figure 4.1 Pathway mutant survivorship curves on soil bacteria

Survival of *C. elegans* strains wild type (*N2*), *daf-16(mgDf50)*, *daf-2(e1368)*, *pmk-1(km25)*, *sek-1(km4)*, *dbl-1(nk3)*, and *tol-1(nr2033)* feeding on A) *E. coli* (OP50), B) *M. luteus*, C) *Pseudomonas sp*., D) *B. megaterium.* Survivorship was determined daily until death. Each combination was repeated 5 times and means are shown.

C)



D)

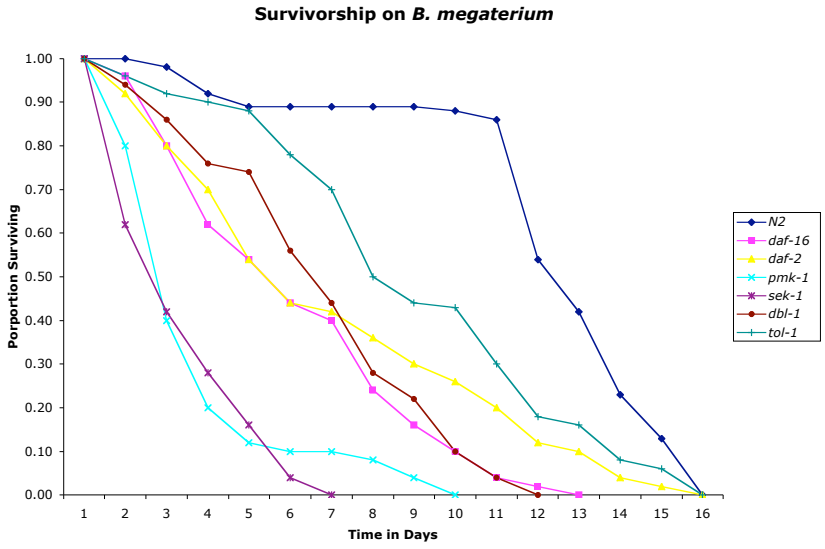
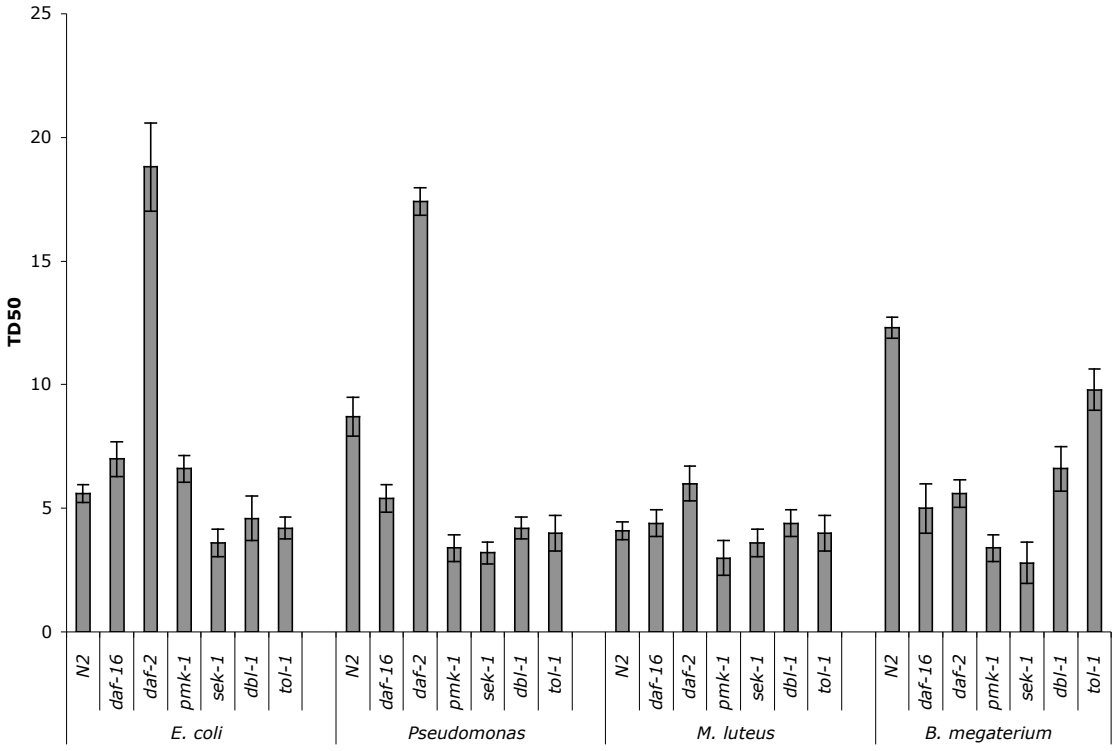


Figure 4.2 Analysis of *C. elegans* defense pathway mutant response to natural soil bacteria

Lifespan of *C. elegans* strains wild type (*N2*), *daf-16(mgDf50)*, *daf-2(e1368)*, *pmk-1(km25)*, *sek-1(km4)*, *dbl-1(nk3)*, and *tol-1(nr2033)* feeding on *E. coli* OP50, *M. luteus*, *Pseudomonas sp.*, and *B. megaterium*. TD_{50} was calculated as the time to death for 50% of individuals from survivorship curves (Figure 4.1) and means from 5 replicates are shown. Error bars are shown (s.e.m.).



CHAPTER 5 - Determining epistatic interactions of new bacteria response genes with known defense pathway components

Abstract

Innate immunity is a highly conserved process of non-adaptive defense response to microbial pathogens. Due to the high degree of conservation, *C. elegans* is a good model for human innate immunity. In *C. elegans*, the two major pathways involved in response to human pathogenic bacteria are the insulin-like receptor and the p38 MAPK pathways. Recently, we found that many *C. elegans* genes are differentially expressed in response to exposure to different grassland soil bacteria. Using functional tests we found that many of the identified genes affect lifespan, which indicates differential susceptibility to bacterial infection and death. Here we aim to determine whether these genes are components of either the insulin-like or p38 MAPK pathways, by investigating epistatic interactions. We found that some of the identified genes that respond to different grassland soil bacteria appear to lie downstream of a particular defense pathway. Surprisingly we also found that mutations in certain cuticular collagen genes can suppress *daf-2/Ins* extended longevity.

Introduction

Innate immunity is an ancient, highly conserved mechanism for defense against bacterial infection. This non-specific immunity is thought to be the first line of defense in higher organisms and has its origin in distantly related organisms (Schulenburg et al. 2004). Because of the high degree of evolutionary conservation, the innate immune system of *C. elegans* is a good model for human innate immunity (Tan et al. 1999a, Mahajan-Miklos et al. 2000, Aballay and Ausubel 2002, Tan and Ausubel 2002, Sifri et al. 2005). The development of *C. elegans* as a model for human innate immunity has resulted from the investigation of a variety of human pathogens that kill *C. elegans* (Tan et al. 1999b, Labrousse et al. 2000, Sifri et al. 2002, Aballay et al. 2003, Sifri et al. 2003, Moy et al. 2004, Tenor et al. 2004). Use of an invertebrate experimental model for host-pathogen interactions has greatly facilitated the dissection of innate immunity (Aballay and Ausubel 2002). From the use of this genetically tractable model system many new components of innate immunity have been identified (Ewbank 2006), which could potentially represent novel antimicrobials in an age of increasing antibiotic resistance for most human pathogens (Moy et al. 2006). Recently, we identified many genes that were differentially expressed in response to different bacterial environments in *C. elegans* (See Chapter 3). When we investigated the differentially expressed genes using available mutants and pathogenicity assays we found that 14/21 altered lifespan on *E. coli* (OP50). Of the mutants that altered lifespan nine significantly increased lifespan and five significantly reduced lifespan and none had been previously shown to be involved in *C. elegans* defense response.

The objective of this study was to determine through which known defense pathway(s) the observed mutant induced alterations in lifespan are mediated. Transcriptional profiling in pathway component mutants have been used to identify genes downstream of various defense pathways including the insulin-like signaling pathway (*daf-2*/*daf-16)* (McElwee et al. 2003, McElwee et al. 2004, Murphy 2006, Singh and Aballay 2006) and the p38 MAP kinase pathway (*pmk-1*/*sek-1*) (Kim et al. 2002, Huffman et al. 2004a, Kim et al. 2004, Troemel et al. 2006). However, none of the genes we identified that affected time to death for 50% of the individuals in a population (TD_{50}) were contained within these microarray datasets (See Chapter 1). RNAi screens to identify additional genes in the pathway have also been used to identify genes that influence aging in wild type (Hamilton et al. 2005, Lee 2006) and *daf-2* mutant genetic backgrounds (Samuelson et al. 2007), but none of the genes we identified were identified in these screens. Thus we decided to directly test within which pathways our genes functioned.

Our approach was to simultaneously inactivate the genes identified that influenced lifespan on *E. coli* OP50 from a previous study (See Chapter 3) and representative genes from the known defense pathways, *daf-16*, *daf-2*, *pmk-1* and *sek-1* to determine whether the identified genes fall into one of the two major defense pathways. We found one gene that appears to function in the *daf-2*/*daf-16* pathway, however we were unable to confidently determine through which pathway most of the other genes affect lifespan. Interestingly, we found that mutations in cuticular collagen genes suppressed *daf-2* extended longevity through an unknown mechanism.

Materials and Methods

C. elegans strains and maintenance

The following strains were used: Bristol (*N2*), *cpi-1* (*ok1213*), *dpy-17* (*e1295*), *mtl-2* (*gk125*), *dhs-28* (*ok450*), *Y57A10C.6* (*ok693*), *acdh-1* (*ok1489*), *rol-6* (*e187*), *dpy-14* (*e188*), *fat-2* (*ok873*), *cyp-37A1* (*ok673*), *pab-2* (*ok1851*), *sqt-2* (*sc108*), *F55F3.3* (*ok1758*), *C23H5.8* (*ok651*), *dpy-5* (*e63*), *dpy-20* (*e1282*), *dpy-21* (*e428*), *sqt-3* (*sc63*). Growth and maintenance conditions were as described (Brenner 1974, Sulston and Hodgkin 1988).

Bacterial strains

Bacterial strains used in this study include *E. coli* strain OP50 cultured in Luria-Bertani (LB) broth and spotted on Nematode Growth Media (NGM). *E. coli* strain HT115 carrying the RNAi vector L4440 or L4440 derived vectors. These vectors have been designed to express double stranded RNA (dsRNA) and *daf-16* and *sek-1* vectors were obtained courtesy of the Timmons laboratory (Kamath et al. 2001, Kamath et al. 2003). *daf-2* and *pmk-1* dsRNA expressing vectors were constructed using amplicons made with primers CENIX:66-G11_T3 and CENIX:66-G11_T7 that amplify secuence of the *daf-2* gene from wild-type *C. elegans.* These PCR fragments were then cloned into the L4440 vector (Kamath et al. 2001, Kamath et al. 2003). *pmk-1* dsRNA expressing vectors were constructed similarly using the primers CENIX:83-A1_T3, CENIX-A1_T7.

Feeding RNA interference

RNAi clones used in this study were from the Ahringer library (Kamath et al. 2001, Kamath et al. 2003) except for the clones that correspond to *daf-2* and *pmk-1* which we constructed as described above. All RNAi constructs were verified by sequencing. HT115 carrying RNAi vectors was grown overnight at 37°C in LB supplemented with 50μg/ml ampicillin. They were then seeded onto RNAi plates that contain 10μg/ml carbinicillin and 1mM isopropylthiolgalactosidase. RNAi clones were grown for two days to induce dsRNA expression before 20 L4/young adult *C. elegans* were transferred to plates and maintained at 25°C for longevity assays. L4440 empty vector was used as a negative control and N2 fed RNAi clones was used as a positive control.

Longevity assays

Longevity assays were performed as previously described (Tan et al. 1999a, Coolon et al. In Prep) and from survivorship curves TD_{50} was calculated as time to death for 50% of individuals in a population. Briefly, worms were synchronized by bleaching to collect eggs. Worms were then grown to L4 on *E. coli* (OP50) to standardize test populations, and then transferred to the RNAi plates as young adults (20 worms per plate) and were maintained at 25°C. Surviving worms were then re-plated daily and the fraction surviving was determined daily. Worms were determined to be dead when they no longer responded to touch with platinum wire. All pathogenicity tests were conducted in at least three independent replicate experiments.

Statistical analysis

Hypothesis testing was done using the GLM procedure in SAS (SAS Institute Inc., Cary, North Carolina, USA). The model used for these tests is shown.

Model: $Y = \mu + \text{genotype} + \text{RNAi} + \text{genotype x RNAi} + \text{error}$

Where Y is equal to TD_{50} and each hypothesis tested using test statements.

Results

Effects of growth on E. coli HT115 containing L4440 empty vector

We examined the effect of RNAi knockdown of *daf-2*, *daf-16*, *sek-1* and *pmk-1* in the background of mutations in each of the identified genes. Thus mutant animals for each gene were fed bacteria expressing dsRNA corresponding to *daf-2*, *daf-16*, *sek-1*, and *pmk-1*. This will be referred to as *daf-2*(RNAi), *daf-16*(RNAi), *sek-1*(RNAi) and *pmk-1*(RNAi). We examined the interaction of candidate defense mutants identified in functional tests of microarray identified genes (See Chapter 3) with known *C. elegans* defense/ageing pathways using feeding RNAi (Kamath et al. 2001, Kamath et al. 2003) directed against components of *C. elegans* defense/aging pathways.

In order to knock down the expression of defense pathway genes we must use the *E. coli* strain HT115, which has T7 RNA polymerase under the control of an isopropylthiolgalactosidase (IPTG) inducible promoter. When we transform *E. coli* HT115 with the L4440 vector containing T7 promoters but no sequence between them,

no dsRNA is expressed. This serves as a control for the presence of the vector in the bacteria. When a portion of a gene is cloned into L4440 between the T7 promoters, dsRNA is expressed corresponding to the gene of interest as the bacteria is grown on plates containing IPTG. This dsRNA serves to knockdown the level of mRNA for the gene targeted in *C. elegans* as it feeds on the *E. coli.* Feeding RNAi is a commonly used technique in *C. elegans* and leads to quite reliable gene knockdown through sequence specific mRNA degradation (Fire et al. 1998, Kamath et al. 2001, Kamath et al. 2003).

Here we investigated mutants that either increase or decrease lifespan and began to determine through which pathway(s) they act. It is known that genes of various functions contribute to defense and lifespan (Lee et al. 2003, Hamilton et al. 2005, Lee 2006), so it was not surprising that our list of candidate genes was also quite diverse. Molecular functions of the mutants we identified that alter *C. elegans* lifespan when grown on *E. coli* include metabolism, antimicrobial, collagen biosynthesis, stress response, and unknown function (See Chapter 3, Tables 3.1, 3.7).

When wild-type *C. elegans* (*N2*) was grown on *E. coli* HT115 containing L4440 empty vector as a control, we found that lifespan was much higher than when wild type was grown on *E. coli* OP50 (See Chapter 3). In fact wild type had a two-fold difference in lifespan between our previous work with OP50 and this study suggests there must be a difference between *E. coli* strains OP50 and HT115 to account for the differences observed in lifespan. To determine if this difference is attributable to the presence of the L4440 vector we also cultured wild type on HT115 without L4440 and found there was no difference from that containing empty vector L4440 (data not shown). We found that 9/14 mutant strains had significant changes in lifespan when grown on *E. coli* HT115 that

contained empty vector L4440, and of those two were significantly increased and seven were reduced (Table 5.1, Figure 5.1a). Of those who's lifespan was affected on both strains of *E. coli*, only four were affected in the same direction on both OP50 and HT115 and five were affected in opposite directions on the two *E. coli* strains. Those that were affected in the same direction on both *E. coli* strains were mutantions in either collagens (*dpy-17* and *dpy-14*) or genes of unknown function (*C23H5.8* and *F55F3.3*) (Table 5.1, Figure 5.1a). The genes that altered TD_{50} in opposite directions were of diverse molecular function (Table 5.1, Figure 5.1a). It will be interesting to determine the intrinsic differences between the two strains of *E. coli* that lead to the observed differences in mutant and wild type lifespan.

Interpretation of pathway analysis

Determining through which pathway a particular mutant affects a phenotype is done using genetic pathway analysis. This is performed by pairing null mutant alleles, each of which have effects on the same phenotype. Through comparisons of the individual mutational effects to the effect of the double mutant on a particular phenotype, one can infer whether the genes function in the same or parallel pathways. In addition, epistasis tests can be used to determine order within a pathway. This analysis can also be performed using RNAi to phenocopy the affects of a mutation in a gene. Interpretations based on RNAi are less conclusive, however it can shed some light on in which pathway a gene functions. Depending on the directionality of mutational effects on time to death for 50% of a population the different pairings of knockdowns in this study have different interpretations.

If two mutants affect TD_{50} in the same direction then additive or synergistic effects can be determined. Additive effects are indicated when the effect of the double knockdown on TD_{50} is comparable to the addition of the single mutational effects on TD_{50}. Synergistic effects are indicated when the double knockdown has phenotypic effects that exceed the addition of the two single mutational effects. Interpretation of these two types of effects (additive vs. synergistic), differ with synergistic effects indicating that the mutant tested functions in a parallel pathway. Additive effects are less definitive indicating that a gene might function in the same pathway or in parallel pathways. In this study, double knockdowns were made between test genes and genes that function in two different parallel innate immunity pathways. This aided our analysis because if a particular gene functions in a pathway we should observe additive effects of a mutant in this gene with one pathway and synergistic effects with the other pathway. If we observe additive effects of a particular mutant allele with mutations in both pathways then the interpretation is that the mutant tested is either downstream of both pathways or in a third parallel pathway.

If two mutants have effects in different directions on the same phenotype then epistasis can be determined. Epistasis is the masking of a phenotype of one mutant with the phenotype of another mutant when the two single mutations act in different directions on the same phenotype. In genetic pathways, the epistatic allele is the the allele who's phenotype is observed and is interpreted as downstream of the allele which it is epistatic to. However, when an allele is epistatic to a gene in one pathway and shows additive effects with a mutant in a parallel pathway ,it is difficult to determine to what pathway it belongs. For this set of analyses, due to the use of RNAi we will focus on only those

results that are conclusive. For the rest of the double knock down combinations, double mutants will need to be made to make stronger conclusions.

Insulin signaling pathway components

Knockdown of the downstream FOXO transcription factor of the insulin-like signaling pathway, *daf-16*(RNAi), resulted in a minor reduction in lifespan but this reduction was not significant (Table 5.1, Figure 5.1b). To determine whether the 14 soil bacterially induced defense genes interacted with *daf-16* we examined lifespan of each mutant in combination with *daf-16*(RNAi). Of the 14 genes tested only *dhs-28; daf-16*(RNAi) and *F55F3.3; daf-16*(RNAi) animals showed a significant decrease in lifespan when compared to *daf-16*(RNAi) alone (Table 5.1, Figure 5.1b). These effects are additive as both *dhs-28* and *F55F3.3* reduced lifespan with empty vector and *daf-16*(RNAi) alone reduced lifespan while the effects were non-significant (Table 5.1, Figure 5.1b).

daf-2(RNAi) caused about a two-fold lifespan extension (Table 5.1), similar to that seen by others in previous studies (Barsyte et al. 2001, Garsin et al. 2003). To identify interactions between our 14 bacterially induced defense genes and *daf-2* we made double knockdowns and found that many had significant effects on *daf-2*(RNAi) longevity. We found 11/14 genes interacted with *daf-2*(RNAi) and effected lifespan, and of those, mutations in four genes significantly increased lifespan and seven significantly reduced *daf-2*(RNAi) lifespan (Table 5.1, Figure 5.1c). Surprisingly, *mtl-2; daf-2*(RNAi), *acdh-1; daf-2*(RNAi), *cyp-37A1; daf-2*(RNAi), and *Y57A10C.6; daf-2*(RNAi) significantly increased lifespan as compared to *daf-2*(RNAi) (Table 5.1, Figure 5.1c).

Previous studies have not identified these four genes in large RNAi screens for genes that extend wild type (Hamilton et al. 2005) and *daf-2* longevity (Samuelson et al. 2007). However, this is not surprising as these two screens found sets of genes with very little overlap, which is probably a consequence of the use of feeding RNAi for the screen (Lee 2006).

We also found significant lifespan reductions for *fat-2; daf-2* (RNAi), *dhs-28; daf-2*(RNAi), *dpy-17; daf-2*(RNAi), *sqt-2; daf-2*(RNAi), *rol-6; daf-2*(RNAi), *F55F3.3; daf-2*(RNAi) and *dpy-14; daf-2*(RNAi) compared to *daf-2*(RNAi) (Table 5.1, Figure 5.1c). Interestingly, 57% of the genes that reduced lifespan encoded cuticular collagens (Table 5.1). Cuticular collagens in general have not been found to be transcriptionally regulated by *daf-2*/*daf-16* in microarray experiments (McElwee et al. 2003, Murphy et al. 2003, McElwee et al. 2004, Murphy 2006, Troemel et al. 2006) or found in RNAi screens for longevity effects (Hamilton et al. 2005, Samuelson et al. 2007). Surprisingly, *rol-6* lifespan reduction in the *daf-2*(RNAi) background is a synthetic phenotype, as *rol-6* has no lifespan phenotype alone. It is possible that the RNAi against the insulin signaling pathway could provide a sensitized background that allowed for uncovering the *rol-6* lifespan phenotype. We also found synthetic phenotypes for *cyp-37A1; daf-2*(RNAi) and *Y57A10C.6; daf-2*(RNAi) animals, interestingly both double knockdowns have significantly increased lifespan as compared to *daf-2*(RNAi). It will be interesting to determine the nature of this interaction.

p38 MAPK pathway components

We next focused on the p38 MAPK pathway by examining lifespan of each of the 14 mutants in combination with *sek-1*(RNAi). *sek-1*(RNAi) animals had a minor reduction in lifespan compared to empty vector but this reduction was not significant (Table 5.1, Figure 5.1d). Only two mutants significantly affected lifespan in the *sek-1*(RNAi) background with *C23H5.8; sek-1*(RNAi) having significantly increased lifespan and *dpy-14; sek-1*(RNAi) having significantly reduced lifespan compared to *sek-1*(RNAi) (Table 5.1, Figure 5.1d). Interestingly, very few of the mutant alleles interacted significantly with *sek-1*. Further analysis of *C23H5.8* and *dpy-14* will be necessary to conclusively determine the nature of their interactions with the defense pathways to mediate longevity.

We further investigated the p38 pathway and investigated interaction between the 14 bacterially induced defense genes and *pmk-1*. We found a slight increase of *pmk-1*(RNAi) lifespan as compared to empty vector and this effect was not significant (Table 5.1, Figure 5.1e). Surprisingly, in contrast with the data for *sek-1* (RNAi), seven mutants interacted with *pmk-1*(RNAi) (Table 5.1, Figure 5.1e). *mtl-2; pmk-1*(RNAi), *dpy-17; pmk-1*(RNAi), *dhs-28; pmk-1*(RNAi), *sqt-2; pmk-1*(RNAi), *fat-2; pmk-1*(RNAi), *F55F3.3 ; pmk-1*(RNAi) and *dpy-14; pmk-1*(RNAi) all had reduced lifespan compared to *pmk-1*(RNAi). Further investigation with double mutants will be necessary to determine the nature of these interactions.

Interaction of defense pathways and cuticular collagens

We found that all of the tested collagen genes suppressed *daf-2* extended longevity; therefore, we chose three others (*dpy-5*, *sqt-3* and *dpy-20*) that were not identified in our previous studies as differentially regulated in response to bacterial environment to serve as controls. As these all lead to morphological abnormalities (i.e. dumpy phenotype), we also included *dpy-21*, which leads to a similar dumpy phenotype, but is caused by a defect in dosage compensation (Yonker and Meyer 2003, Hartman and Ishii 2007). We found that both *dpy-5* and *dpy-21* had significant reductions in lifespan compared to wild type (Table 5.1, Figure 5.2a). Similarly, *dpy-5; sek-1*(RNAi), and *dpy-21; sek-1*(RNAi) had significantly reduced lifespan compared to *sek-1 (RNAi)* alone (Table 5.1, Figure 5.2b). In combination with *daf-16*(RNAi) we found that all four control genes resulted in significant reductions compared to *daf-16*(RNAi) (Table 5.1, Figure 5.2c). We found the same trend for double knockdown of *pmk-1* and the control genes with all four causing reductions compared to *pmk-1* (RNAi) (Table 5.1, Figure 5.2d). Surprisingly, all four control genes suppressed *daf-2* extended longevity (Table 5.1, Figure 5.2e). When this is extended to all the cuticular collagens that were tested we found that all significantly suppressed *daf-2* extended lifespan (Table 5.1, Figures 5.1c, 5.2e). This trend is surprising as cuticular collagens have not been found with RNAi screens (Hamilton et al. 2005, Samuelson et al. 2007) or transcriptional profiling in *daf-16*, *daf-2*, *pmk-1*, or *sek-1* mutants (McElwee et al. 2003, Murphy et al. 2003, McElwee et al. 2004, Troemel et al. 2006). We are not sure how this suppression occurs but it will be interesting to further dissect the molecular basis of this genetic interaction.

Discussion

Of the genes tested it appears that many interact with components of known defense/longevity pathways in *C. elegans*. We found differences in the behavior of wild type and several of the mutants when grown on *E. coli* HT115 containing the L4440 empty vector as compared to *E. coli*, OP50 (Table 5.1) (See Chapter 3). Strangely, this has not been discussed before and could have huge implications for other studies of lifespan using feeding RNAi. It will be interesting to determine the intrinsic differences between these two *E. coli* strains that lead to the observed differences in *C. elegans* lifespan. The behavior of wild-type *C. elegans* (*N2*) is odd because it was grown for decades on *E. coli* OP50 prior to freezing and therefore should have become better adapted to life in this environment. Surprisingly, to our knowledge this phenomenon had not been previously documented.

We tested 14 mutants found in our previous study (See Chapter 3) to have lifespan defects on *E. coli* OP50 for interaction with known defense pathways. We found that 9/14 mutant strains had significant changes in lifespan when grown on *E. coli* HT115 that contained the empty vector L4440 and of those, two were significantly increased and seven were reduced (Table 5.1, Figure 5.1a). Of those that were significant in both strains of *E. coli*, only four were in the same direction on both OP50 and HT115 and five were in opposite directions on the two *E. coli* strains. As was observed with wild type there are differences between *E. coli* OP50 and HT115 in mutational effects on lifespan. For those with effects on *E. coli* HT115 containing empty vector we could perform pathway analysis and attempt to determine if they follow expectations of a gene

downstream of a particular pathway. Only two genes had mutational effects consistent with being in a particular defense pathway.

F55F3.3 mutant animals had significantly reduced lifespan alone as well as *F55F3.3; daf-16*(RNAi) animals. Because the magnitude of effects in the double knockdown was comparable to the addition of the two single knockdown reductions the effects were additive. *F55F3.3* and *daf-2*(RNAi) had effects on lifespan in different directions and through the double knockdown analysis we found that *F55F3.3* is epistatic to *daf-2* as the *F55F3.3* phenotype was the one that was visible. Typically this would indicate that *F55F3.3* is downstream of *daf-2*, however, recent studies found that *pmk-1* and *sek-1* are epistatic to *daf-2* but reside in a parallel pathway. This evidence is inconclusive to determine a pathway for *F55F3.3*. When we investigated the components of the p38 MAPK pathway, we found no effect of *F55F3.3; sek-1*(RNAi), however we found a synergy between *pmk-1*(RNAi) and *F55F3.3* with effects greater in the double knockdown than the addition of the two knockdowns individually. This strongly suggests that *F55F3.3* is downstream of *daf-2* in the insulin-signaling pathway. Further analysis with double mutants will be necessary to make concrete conclusions.

dhs-28 also had reduced lifespan on *E. coli* HT115. Similar to *F55F3.3* there were additive effects with *daf-16*(RNAi) and *dhs-28* was epistatic to *daf-2*. Again we only found interaction with *pmk-1*(RNAi), however the effects of *dhs-28* and *pmk-1* were additive. As a result, interpretation of the pathway analysis is inconclusive. *dhs-28* could be downstream of both pathways or in a third parallel defense pathway. Epistatic analysis and further pathway analysis with other known pathways will be necessary to determine through which defense pathway the *dhs-28* lifespan reduction is mediated.

When we investigated the role of the 14 genes in *daf-2* extended longevity we found that most (11/14) significantly altered lifespan in combination with *daf-2*(RNAi) compared to *daf-2*(RNAi) alone (Table 5.1, Figure 5.1c). Interestingly, we found that some of the mutant alleles actually increased the already long-lived *daf-2*. We also found that many tested genes could suppress *daf-2* extended longevity. Of the mutant alleles that could suppress *daf-2* longevity, there was enrichment for genes that encode cuticular collagens. It is possible that the cuticle is under the control of the innate immunity system and contributes to defense. It is reasonable that this outer barrier to infection could be somehow modified in response to interaction with the bacterial environment. To see if this effect was consistent for other cuticular collagens we identified three other genes in this group and found that there were similar reductions in *daf-2* longevity when *daf-2*(RNAi) was paired with other cuticular collagen knockout mutant alleles (Table 5.1, Figure 5.2e). To determine if this effect was solely due to abnormal morphology we also tested *dpy-21*, which is involved in dosage compensation and found that it also reduced *daf-2* longevity. Upon further investigation we found that a previous study found that *dpy-21* had effects on lifespan. This was because of the negative effects of overexpression of the genes on the X-chromosome resulting from defective dosage compensation (Yonker and Meyer 2003, Hartman and Ishii 2007). Therefore the data collected on the cuticular collagens is inconclusive. It is possible that cuticular collagens are downstream of known defense pathways or that non-specific effects or morphology are contributing to lifespan. Only through further examination of other dumpy control genes that are not cuticular collagens will this be determined.

Further studies are needed to see if this trend is widespread across other cuticular collagens or if all strains with abnormal morphology also suppress *daf-2* longevity. Whether these genes also function in the dauer pathway and suppress the formation of dauer larvae will also be interesting to determine. This is the first instance that this trend has been reported and if it is widespread it should have been observed in lifespan RNAi screens. This possibly illustrates one downside to the use of RNAi screens as the do not reach saturation because some genes are more difficult to knockdown by RNAi, either through where they are expressed or via some other factor. Further double mutant analysis will be needed to verify whether the genes tested fall into a particular defense/longevity pathway. This could shed light on how *C. elegans* responds to bacterial pathogens and will be necessary in order to investigate how defense response contributes to soil nematode community dynamics in the wild, and to identify important factors that are conserved in human innate immune response to pathogens.

Acknowledgements

We would like to thank Lisa Timmons for providing RNAi clones and the *Caenorhabditis* Genetics Center for providing *C. elegans* strains.

Figures and Tables

Table 5.1 Effects of insulin signaling and p38 MAPK RNAi on mutant lifespan

Mutants used in pathway analysis are shown, with control genes indicated. Strains of *C. elegans* were fed *E. coli* HT115 expressing either L4440 empty vector or dsRNA targeted against *daf-16*, *daf-2*, *sek-1*, or *pmk-1*. Values given are average time to death for 50% of a population of 20 worms or TD_{50} and standard error is given in parentheses. Statistical significance in comparisons between mutant and wild type (+) within a given environment are indicated by (+) for significantly increased and (-) for significant decrease ($p < 0.05$). Statistical significance for effect of RNAi on wild type compared to empty vector is indicated by an *.

Gene	***L4440 (empty vector)***	***daf-16 RNAi***	***daf-2 RNAi***	***sek-1 RNAi***	***pmk-1 RNAi***
wt	12.0 (2.2)	10.0 (1.8)	20.5 (1.3)*	11.8 (3.3)	13.5 (1.9)
dpy-17	9.3 (1.5)-	8.0 (1.0)	11.3 (.58)-	10.3 (1.5)	9.3 (1.5)-
dhs-28	8.7 (1.5)-	6.3 (1.5)-	12.3 (2.5)-	8.0 (2.0)	9.3 (1.5)-
acdh-1	14.7 (1.5)+	9.0 (1.0)	23.0 (2.0)+	14.3 (.58)	14.7 (2.1)
Y57A10C.6	13.0 (1.0)	9.3 (.58)	22.3 (1.5)+	13.0 (3.0)	13.0 (1.0)
mtl-2	9.0 (1.0)-	8.3 (.58)	24.3 (2.5)+	8.3 (1.5)	9.7 (1.2)-
cpi-1	11.3 (2.1)	10.7 (2.1)	22.7 (3.5)	11.7 (1.5)	10.7 (1.5)
pab-2	12.3 (1.5)	8.7 (.58)	19.3 (1.5)	13.0 (3.0)	13.3 (2.3)
F55F3.3	7.0 (1.0)-	4.3 (.58)-	8.3 (.58)-	7.3 (1.5)	4.7 (1.2)-
cyp-37A1	13.0 (2.0)	8.7 (.58)	22.7 (2.1)+	13.3 (1.5)	10.3 (2.3)
sqt-2	9.0 (1.0)-	8.3 (2.1)	11.3 (1.2)-	10.3 (1.2)	9.0 (1.0)-
fat-2	7.3 (.58)-	9.7 (1.2)	17.0 (2.0)-	8.0 (1.7)	8.7 (.58)-
dpy-14	5.3 (1.5)-	8.7 (.58)	4.0 (1.0)-	4.3 (1.2)-	4.0 (1.0)-
C23H5.8	14.3 (2.1)+	9.7 (.58)	22.7 (3.1)	15.0 (2.0)+	12.0 (1.0)
rol-6	10.0 (1.0)	8.0 (1.0)	9.7 (1.2)-	8.0 (2.6)	12.7 (1.5)
Controls					
dpy-5	6.7 (1.5)-	5.7 (1.5)-	7.0 (1.7)-	6.3 (.58)-	6.3 (1.2)-
dpy-20	10.0 (1.0)	7.0 (1.0)-	12.7 (1.5)-	10.7 (1.5)	7.7 (1.5)-
sqt-3	11.7 (1.2)	7.0 (1.0)-	15.3 (2.1)-	12.0 (1.0)	9.0 (2.0)-
dpy-21	6.7 (.58)-	6.0 (1.0)-	12.0 (3.0)-	6.7 (1.5)-	5.7 (1.5)-

Figure 5.1 Lifespan of mutants grown on *E. coli* HT115

C. elegans mutants were grown on *E. coli* HT115 carrying either a) L4440 empty vector or dsRNA targeted knock-down of b) *daf-16* c) *daf-2* d) *sek-1* or e) *pmk-1*. Time to death for 50% of the population (TD_{50}) and standard error is indicated with error bars. Mutants are shown in grey and wild type (*N2*) is shown in black for comparison.

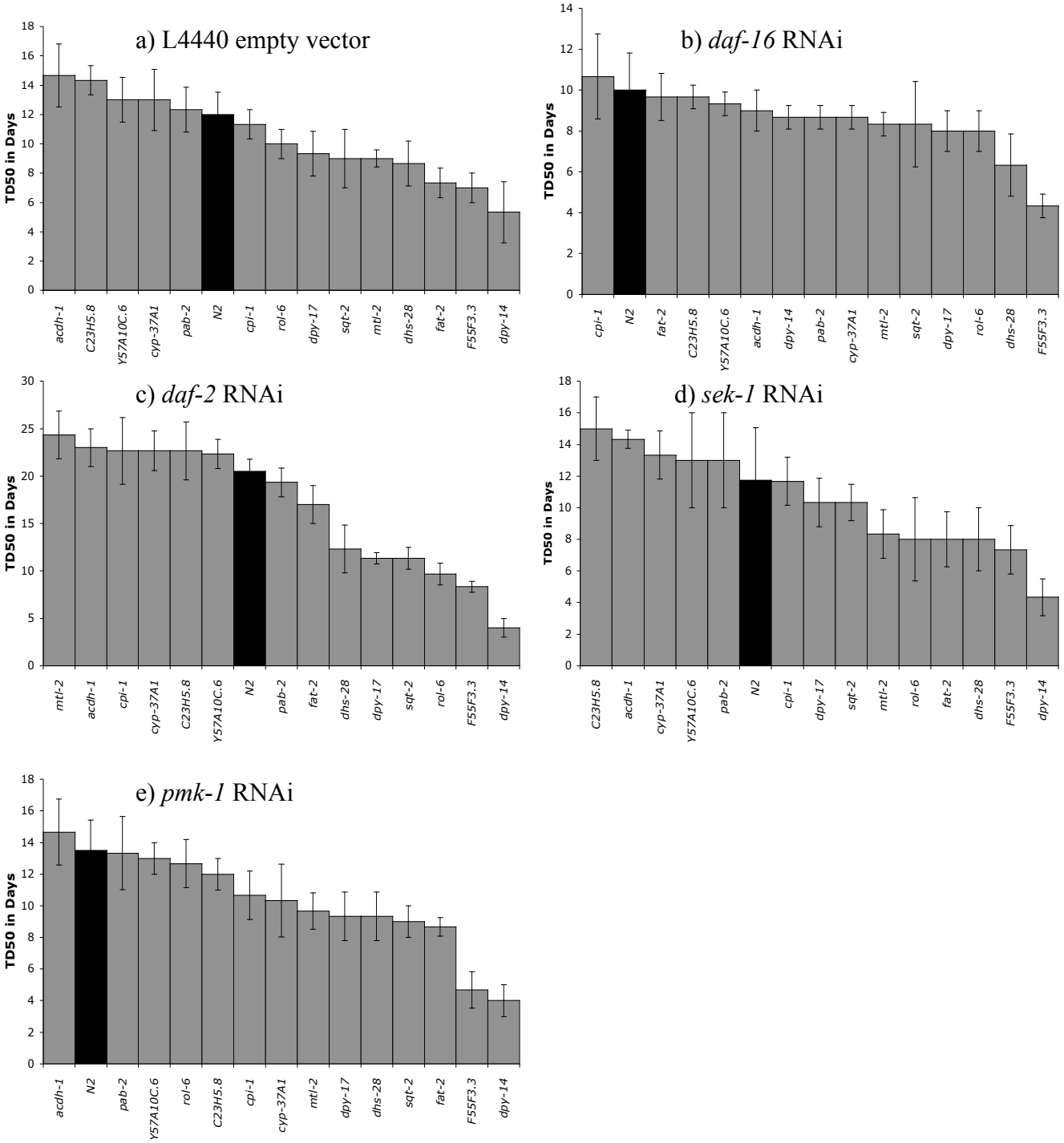
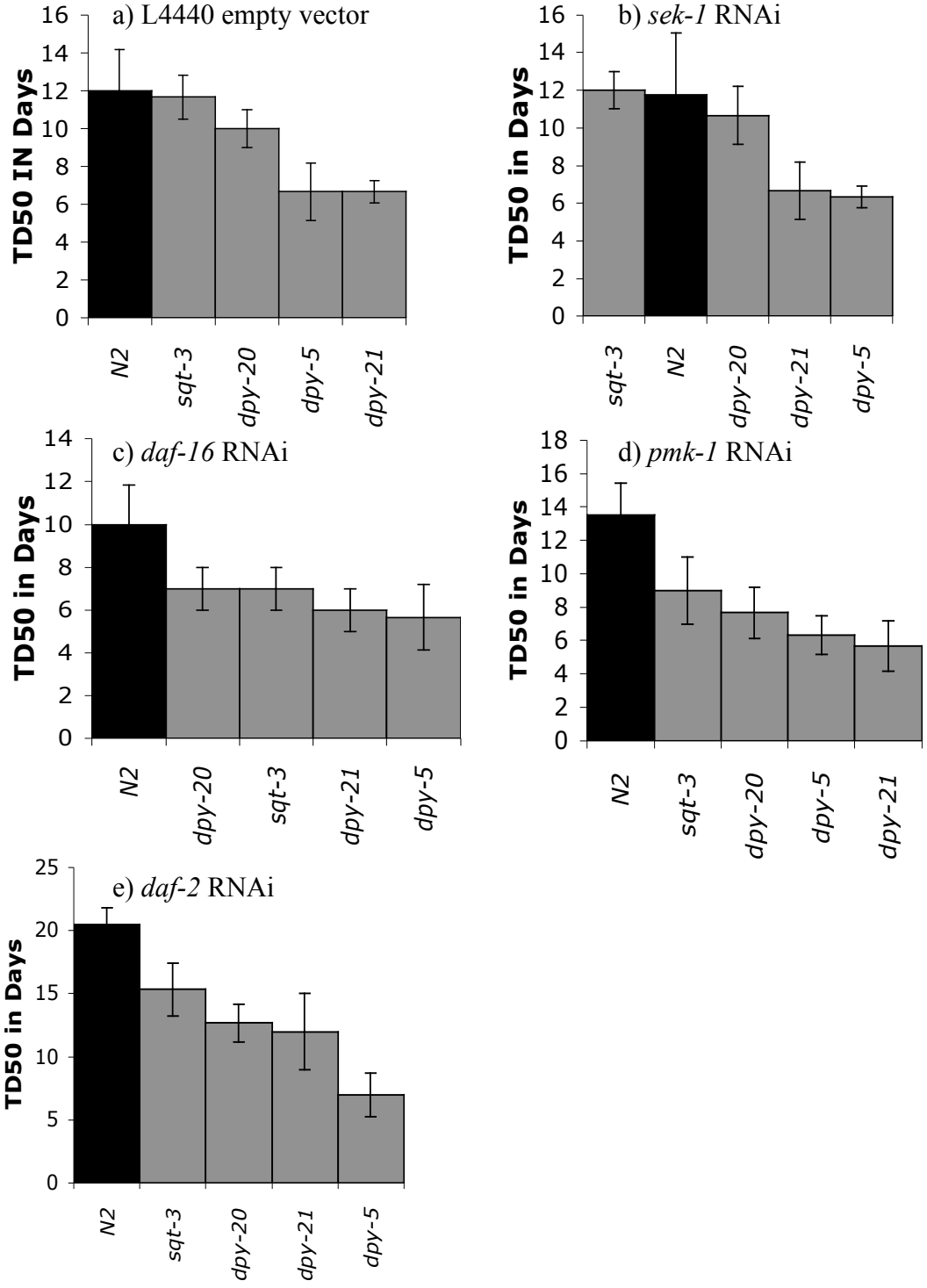


Figure 5.2 Lifespan of control mutants grown on *E. coli* HT115

C. elegans control mutants were grown on *E. coli* HT115 carrying either a) L4440 empty vector or dsRNA targeted knock-down of b) *daf-16* c) *daf-2* d) *sek-1* or e) *pmk-1*. Time to death for 50% of the population (TD_{50}) and standard error is indicated with error bars. Mutants are shown in grey and wild type (*N2*) is shown in black for comparison.



CHAPTER 6 - Profiling Innate Immunity: transcripts from the worm

Abstract

In the absence of an adaptive immune system, invertebrates need to mount other defenses against microbial pathogens. The genetic model organism *Caenorhabditis elegans* has become an emerging model for the study of non-specific or innate immunity. To identify the genes involved in innate immunity, many studies have focused on genes that are differentially regulated in response to stress or pathogen infection and in mutant genetic backgrounds. Transcriptional profiling is a powerful tool for identifying genes that are differentially regulated in response to different bacterial environments, and has been used to identify large numbers of genes implicated in *C. elegans* defense against bacterial pathogens. Other investigators have disabled genes in known defense pathways, typically the most upstream components, and compared mutant gene expression to that of the wild type to identify genes downstream of a particular pathway. These two complementary approaches have identified many candidate genes that play a role in defense. Here we present the first comprehensive review of this literature. We have compared these lists to determine overlap, and the degree to which the different experiments identify genes in common. This allowed us to make predictions, which in some cases are confirmed by available data. Because differential gene expression in *C. elegans* defense has been investigated for more than six years now, we should take a step

back to look at the data as a whole and make inferences, which will help us to better understand innate immunity.

C. elegans as a system to study innate immunity

The natural environment of *C. elegans* likely consists of nutrient-rich substrates, especially decaying organic matter as it is typically isolated in nature from compost, soil and decaying fruit (Barriere and Felix 2007). These nutrient-rich habitats are also typically highly populated with numerous species of bacteria, which serve as the food source for *C. elegans*. As a result *C. elegans* is faced with the possibility of infection by a multitude of bacterial pathogens and must defend itself from this onslaught. *C. elegans* lacks the adaptive immune system seen in higher organisms and therefore must utilize an innate immune system to combat pathogens (Mahajan-Miklos et al. 2000, Mallo et al. 2002, Millet and Ewbank 2004, Schulenburg et al. 2004, Ausubel 2005). Until recently the immune system of *C. elegans* was relatively uncharacterized, but since 1999 a variety of human and plant pathogens have been shown to colonize and kill *C. elegans*. In the last decade as many as five different *C. elegans* defense pathways have been identified and characterized (Schulenburg et al. 2004). The pathway most well-studied is the insulin-like signaling pathway that consists of the insulin-like receptor DAF-2 (for abnormal dauer formation), which, through a series of steps activates the FOXO transcription factor DAF-16 that is translocated into the nucleus where it regulates the expression of multiple downstream antimicrobial effectors (Murphy et al. 2003, Murphy 2006) (Figure 6.1). This pathway has been extensively studied for its role in lifespan, as *daf-2* mutants are extremely long-lived. The second pathway involved in *C. elegans* innate immunity is the TGF-β-like pathway controlled by DBL-1 (for Dpp, BMP-Like)

that is a TGF-β-like ligand. The *dbl-1* pathway also has roles in male tail development and body size, as well as a role in defense. The third pathway is a MAP kinase pathway made up of the MAPKKK NSY-1, MAPKK SEK-1 and MAPK PMK-1 (p38 family MAPK, See Figure 6.1). The p38 MAPK pathway is highly conserved and is a component of human defense against bacterial pathogens (Pisegna et al. 2004, Dai et al. 2008). The fourth pathway is the toll-like receptor pathway (TOL-1), which is conserved in flies and mammals. In *C. elegans* this pathway was originally thought to only be involved in avoidance behavior towards *Serratia marcescens* (Pujol et al. 2001). However, recently it was found to play a role in defense against Gram-negative bacterial pathogens, but had no effect on sensitivity to the Gram-positive pathogens tested (Tenor and Aballay 2008). The final defense pathway found in *C. elegans* is the ERK pathway, which is also a MAP kinase pathway, comprised of LIN-45, MEK-2 and MPK-1 proteins (Nicholas and Hodgkin 2004a). This pathway is predominantly involved in response to the Gram-positive natural nematode pathogen *Microbacterium nematophilum*, which adheres to the post-anal region resulting in a swelling response characteristic of this particular infection. Finally, it is thought that bacterial pathogens elicit a general stress response, in addition to any specific responses (Figure 6.1). It is possible that this stress response is mediated through a separate pathway(s) or that stress feeds into one or more of the aforementioned pathways in a manner that is not yet well understood.

Caenorhabditis elegans has been co-cultured with a variety of microbial pathogens and different food sources. These different microbes elicit distinct responses in *C. elegans* and are defended against by the previously discussed pathways. Because of the specificity and complexity of these responses, as well as the large number of genes

involved in defense, multiple approaches have been taken to elucidate the underlying genetic mechanisms. One major avenue of research to determine the genes involved in defense is the investigation of *C. elegans* transcriptional responses to various bacterial environments. Here we used published datasets of genes that are differentially expressed in response to different bacterial environments, stress or genes that are regulated downstream of known defense pathways. We first determined the overlap between lists of identified genes, and using statistical tests determined whether there was a greater number of genes in common between two lists than would be expected by chance alone. From significant overlap we can make predictions and in some cases the necessary data is available to confirm the predictions made. Our aim is for this analysis to become a resource that the *C. elegans* community could use to better understand the process of innate immunity.

Analysis

The experiments

Transcriptional response to pathogens

We collected datasets from multiple transcriptional profiling studies where the goal was to identify genes that were involved in innate immunity. The various transcriptional profiling experiments included in this synthesis different greatly in the details of the expression analysis, including differences in the microarray platform utilized, the number of biological replications, and ultimately the number of genes identified (Table 6.1). We acquired datasets from studies that identified genes differentially expressed in response to altered bacterial environments (Mallo et al. 2002,

Couillault et al. 2004, Troemel et al. 2006, O'Rourke et al. 2006, Wong et al. 2007, Coolon et al. In Prep).

To identify components of *C. elegans* innate immune system Troemel *et al.* (2006) investigated transcriptional responses to exposure to the Gram-negative bacterial pathogen *Pseudomonas aeruginosa.)*. Sterile *fer-15(b26)/fem-1(hc17* mutant hermaphrodites were used because in times of stress *C. elegans* will retain embryos until they hatch and kill the mother, which leads to various changes in gene expression that were not of interest. Transcriptional profiling of *fer-15(b26)/fem-1(hc17)* animals grown on *E. coli* OP50 was compared to growth on *P. aeruginosa* PA14 for 4hrs and for 8hrs (Troemel et al. 2006).

In a similar experiment O'Rourke *et al.* (2006) investigated the transcriptional response of *C. elegans* to the Gram-positive nematode pathogen *Microbacterium nematophilum*, which is one of the only nematode-specific pathogens investigated with the *C. elegans* innate immunity model system to date. This nematode pathogen adheres to the anal and post-anal region of the worm and causes a characteristic dar (deformed anal region) or swelling response, as well as constipation (Nicholas and Hodgkin 2004a, O'Rourke et al. 2006). A total of seventy-nine genes were identified to be differentially expressed in response to a 6hr incubation of L2/L3 larvae on virulent *M. nematophilum* CBX102 compared to comparably aged worms grown on an avirulent strain of *M. nematophilum* UV336 (O'Rourke et al. 2006). Many lectins were differentially expressed in response to *M. nematophilum*, and it was suggested that this could be a means by which pathogens are recognized by *C. elegans*. Interestingly, the genes that were identified are clustered in the genome and this suggests that groups of pathogen response

genes could be transcribed as an operon. It will be interesting to see if other innate immunity genes also share the same genomic clustering and are regulated similarly.

In an early study of *C. elegans* transcriptional response to pathogens, Mallo *et al*. (2002) used nylon filter macroarrays containing ~7,500 spotted cDNAs to identify genes responsive to infection by *Serritia marcescens*. Only ten genes were identified as differentially regulated in synchronized sterile *fer-15* mutant animals that had been infected for either 24 or 48hrs with *S. marcescens* Db11. This study found that the genes identified were under the control of the *dbl-1* MAPK pathway through transcriptional analysis of mutants. They showed that the expression of identified genes was no longer regulated in the *dbl-1* mutant background (Mallo et al. 2002). Surprisingly, they identified only 10 differentially regulated genes. This could be due to use of the nylon filter macroarrays (Table 6.1), which the authors suggested were as reliable as spotted glass microarrays. However it was later found that nylon filter macroarrays were less reproducible and required a very large change in expression level to be statistically significant.

Wong *et al*. (2007) investigated transcriptional response of wild-type *C. elegans* to four bacterial pathogens: the Gram-negative *S. marcescens* Db11, Gram-negative *Erwinia carotovora* CFBP2141, Gram-negative *Photorhabdus luminescens* Hb, and Gram-positive *Enterococcus faecalis* OG1RF, and all were compared to wild type grown on *E. coil* OP50. Late L4/young adult worms were exposed to the bacterial pathogens for a duration of 24hrs. The genes identified were enriched in pathogen recognition and antimicrobial functions as well as genes involved in proteolysis, cell death and general stress response. Through use of reporter constructs they were able to visualize response

to the various pathogens in a variety of cells and tissues of living worms (Wong et al. 2007). They found that necrosis is a common feature to *C. elegans* response to bacterial pathogens while it conferred no resistance to bacterial infection. Their attempts to identify regulatory motifs common to infection responsive genes were unsuccessful, suggesting that transcriptional regulation in response to pathogens is more complex than expected. Overall they suggested *C. elegans* has both pathogen-specific and pathogen shared transcriptional responses to infection by bacterial pathogens.

Our lab has studied *C. elegans* transcriptional response to bacteria isolated from grassland soils, some of which were isolated in association with soil nematodes from the Konza Prairie (See Chapter 3). The bacteria used were the Gram-positive *Micrococcus luteus*, which was the most abundant bacteria that was culturable on nematode growth media, from prairie soil samples. We also used *Bacillus megaterium*, which was isolated in association with a nematode of the *Oscheious* genus, a species of *Pseudomonas* that was isolated in association with a nematode of the *Pellioditis* genus and was most closely related to *Pseudomonas flourescens* (See Chapter 3). *E. coli* OP50 was also used as a control and all pair-wise comparisons of the four bacterial environments were made using adult *C. elegans*. We found that there was enrichment for genes involved in defense, metabolism and collagens as well as genes of unknown function. We further characterized this list of genes and biologically validated their importance in tests of mutant life history in the different bacterial environments (See Chapter 3). We found that most significantly affected lifespan or fitness and there was abundant genotype by environment interaction (See Chapter 3).

We have also included the transcriptional response to the fungal pathogen *Drechmeria coniospora* (Couillault et al. 2004). This is the only study, to our knowledge, that has examined *C. elegans* transcriptional response to a fungus. Couillault *et al*. (2004) found that only six genes were differentially expressed in response to infection by *D. coniospora,* which attaches to the cuticle of the worm, pierces through and digests tissues as it grows. Many of the genes that were identified had antimicrobial properties and conferred some resistance to infection and animals carrying mutant alleles of some of the genes were tested and found to be hypersensitive to *D. coniospora* infection. Additionally, the genes identified were dependant on functional TIR-1 (Toll-interleukin-1 receptor) (Couillault et al. 2004) for regulated expression.

Finally, Huffman *et al*. (2004a) investigated *C. elegans* response to the *Bacillus thuringiensis* pore-forming toxin Cry5B. This toxin is secreted by bacteria, forms pores in target cells that will liberate cytosolic nutrients, and eventually leads to cell lysis. Cry5B is a member of the crystal family of pore-forming toxins (PFTs) that makes holes of 25-30nm in diameter in the plasma membrane of target cells (Huffman et al. 2004a). PFTs are important virulence factors for many pathogenic bacteria in humans and invertebrates (Huffman et al. 2004b). To investigate cellular response to this type of virulence factor Affymetrix microarrays were used in a comparison of *C. elegans* growth on *E. coli* JM103 expressing Cry5B to *E. coli* JM103 with empty vector (Table 6.1). Some of the genes identified as differentially regulated contributed to resistance to the toxin as mutant alleles in these genes caused hypersensitivity to the toxin (Huffman et al. 2004a). The response to Cry5B was mediated by upregulation of two pathways, the p38 MAPK and the c-Jun N-terminal Kinase or JNK-like pathways. Functional tests showed

that both pathways provide a significant cellular defense against the Cry5B toxin and the response is conserved in mammals (Huffman et al. 2004b).

Identification of genes regulated by defense pathways

The final group of studies is transcriptional profiling experiments performed on defense pathway component mutants compared to wild type. This type of study aims to determine genes downstream of particular defense pathways, and typically identifies effector genes, some of which encode products with antimicrobial properties. The first of these was transcriptional profiling of the insulin-like signaling pathway where *daf-16* and *daf-2* loss of function mutants as well as *daf-16* (RNAi) and/or *daf-2* (RNAi) animals were compared to wild type (Murphy et al. 2003). The major aim of this study was to identify gene products that were involved in aging downstream of the FOXO transcription factor DAF-16. Two classes of genes were identified, those that were induced in *daf-2* mutant and *daf-2* (RNAi) animals and that were reduced in the double knockdown of both *daf-16* and *daf-2* and were proposed to extend lifespan (Murphy et al. 2003) and those that exhibited the opposite regulation (down in *daf-2* and up in *daf-16:daf-2* double RNAi, proposed to shorten lifespan). Not surprisingly, many of the genes identified were either involved in metabolism or anitmicrobial defense (Murphy et al. 2003). Another study identified the genes that are differentially expressed in dauer larvae. The insulin-like signaling pathway is known to control the formation of the dauer stage and McElwee *et al*. (2004) aimed to identify a gene expression signature of the dauer stage to compare to mutants in the insulin-like signaling pathway. They found that many genes were differentially regulated in comparisons of dauer larvae to comparably staged non-dauers (Table 6.1).

Another study aimed at identifying the downstream components of defense pathways was performed by Troemel *et al*. (2006). They investigated the effect that p38 MAPK *pmk-1* and the upstream MAPKK *sek-1* mutants had on gene expression. First, *pmk-1(km25)* mutants were investigated in a *daf-2(1368)* mutant background. This study found that the mutant allele of *pmk-1(km25)* resulted in altered expression of 36 downstream genes. When *sek-1(km4)* mutants were assayed for expression differences from wild type with microarrays, 110 genes were found to be differentially regulated. From this analysis, comparison with genes downstream of *daf-16* and lifespan assays where *daf-2*-extended lifespan required both *pmk-1* and *sek-1* but that neither *pmk-1* or *sek-1* had reductions in lifespan by themselves, Troemel *et al*. (2006) determined that the p38 MAPK pathway is a separate or parallel pathway that contributes to lifespan and resistance to pathogens.

Methodology

We examined the number of overlapping genes in every pair-wise combination of the above anlyses. From this we calculated a representation factor ($\log_{(\text{number observed in overlap/number expected in overlap})}$) to illustrate the degree of overlap, and determined statistical significance of the overlap, based on the number of genes in the genome (20,000, which is conservative as there is evidence to suggest that not all genes are represented on the commercially available microarrays), and also took into account the sizes of the compared lists and the number of overlapping genes to determine if the overlap was greater than that expected by chance alone (Chi Square test as in Kim et al. 2001). The results of this analysis are found in Table 6.2 where the upper triangular matrix is shown.

Each cell of the column and row labels contains the environmental change inducing expression differences from the aforementioned transcriptional profiling experiments as well as the number of genes identified in that study. The rest of the cells of the table contain a representation factor (RF) of the number of genes in common between the row and column datasets and a significance calculated for the degree of overlap, and cells are color coded based on significance.

Expected trends

Several comparisons were expected to have a large proportion of overlapping genes. For example, responses to either 4hr or 8hr exposure to the pathogen *P. aeruginosa* should be largely overlapping. This was confirmed as 291 genes ($P < 1e-100$) were found to be in common between the two gene lists (Table 6.2). Another obvious prediction was that there should be a large overlap between genes expressed during the dauer stage and genes downstream of *daf-16* because it is known that the insulin-like signaling pathway controls the development of dauer larvae (Gottlieb and Ruvkun 1994, Matyash et al. 2004, McElwee et al. 2004, Masse et al. 2005). We found 102 genes in common between these two groups of regulated genes and the overlap was highly significant ($P < 8.44e-29$, Table 6.2). Interestingly, there was only significant overlap between genes predicted to extend lifespan (class 1) from Murphy *et al*. (2003) and very little overlap with genes predicted to shorten lifespan (class 2) (Table 6.2). This makes sense as dauer larvae are extremely long-lived and the overlapping genes could be those involved in dauer lifespan extension. This prediction could be tested by

monitoring lifespan of mutants of the common genes that are forced into the dauer stage by caloric restriction (Lakowski and Hekimi 1998, Lee et al. 2006b).

New observations

There was a larger than expected overlap between *P. aeruginosa* (both 4 and 8 hrs) and *M. nematophilum* exposure (Table 6.2). This is surprising given the differences between the two bacteria (i.e. cell wall structure) and the different locations of infection, with *P. aeruginosa* colonizing the anterior portion of the gut and *M. nematophilum* adhering to the anal region (Nicholas and Hodgkin 2004a, 2004b). We also noted a significant overlap between *P. aeruginosa* exposure and Cry5B exposure (Table 6.2). This indicates that a substantial component of the *P. aeruginosa* response could be due to stress; perhaps demonstrating that exposure to *P. aeruginosa* causes cell lysis in *C. elegans*. Interestingly, there was significant overlap between *P. aeruginosa* infection and both classes of *daf-16* as well as *pmk-1* and *sek-1* regulated genes. This suggests that response to *P. aeruginosa* likely elicits both the insulin-like pathway as well as the p38 MAPK pathway in a defense process. Thus, one might predict that mutations in either of these two pathways could result in hypersensitivity to *P. aeruginosa* infection and earlier mortality in the *P. aerugionosa* environment. Interestingly, this is exactly what was observed, when either *pmk-1* or *sek-1* mutant animals were grown on *P. aeruginosa* (Troemel et al. 2006) illustrating the benefit of an analysis across experiments. However, *daf-16* mutants did not display an increased sensitivity to *P. aeruginosa*, thus not all predictions based on this analysis will be confirmed (Troemel et al. 2006). This could, however, indicate that the p38 MAPK pathway may be more important in *C. elegans*

defense against *P. aeruginosa* and further experiments will be necessary to confirm this hypothesis. Surprisingly, very little overlap in transcriptional response was found between exposure to *Pseudomonas aeruginosa* and the *Pseudomonas sp*. isolated from grassland soils used in our study (See Chapter 3) (Table 6.2). This is likely due to the differences in pathogenicity of these *Pseudomonas* species: ours was appearantly not pathogenic whereas *P. aeruginosa* is highly virulent (See Chapter 3). It will be interesting to determine the genetic differences between these two species of *Pseudomonas* species that lead to the differences in virulence.

Surprisingly, there was a significant overlap between dauer expressed genes and thegenes downstream of *pmk-1* and *sek-1* (Table 6.2). To our knowledge, a role for the p38 MAPK pathway in dauer formation has not been documented. However, it is possible that crosstalk between the p38 MAPK and insulin-like pathways contributes to dauer formation, as has been shown for lifespan and pathogen resistance (Troemel et al. 2006) or there could be a direct role of the p38 MAPK pathway in dauer formation (Figure 6.2). Similarly, there was significant overlap between genes downstream of *daf-16* and the p38 pathway. Interestingly, there was more overlap of *daf-16* genes that reduce lifespan than genes that increase lifespan with both *pmk-1* and *sek-1* regulated genes (Table 6.2). Also there was greater overlap between *sek-1* regulated genes and both classes of *daf-16* regulated genes, which could suggest that pathway crosstalk could be primarily through interactions of *sek-1*, and *daf-16* that is independent of *pmk-1* (Figure 6.2). However, not surprisingly, there was a great degree of overlap between genes regulated downstream of *sek-1* and *pmk-1* (Table 6.2) as *sek-1* directly interacts

with *pmk-1* to activate it through phosphorylation (Figure 6.2) (Kim et al. 2002, Troemel et al. 2006).

Genes that are responsive to *M. nematophilum* infection had significant overlap with genes downstream of both classes of *daf-16* and *pmk-1* but not *sek-1* regulated genes (Table 6.2). Therefore, we would predict that *pmk-1* and *daf-16* mutants would be more sensitive than *sek-1* to infection by *M. nematophilum*. It will be interesting to see if this prediction can be confirmed. We also found significant overlap between Cry5 and Cd regulated genes and those regulated in response to *M. nematophilum*. It is possible that *M. nematophilum* causes some cytolysis in a similar manner as do the two stressors and this overlap could represent the genes that are responsive to this aspect of infection by *M. nematophilum*. It has not been shown that *M. nematophilum* infection causes a great deal of cytolysis, however this could be specifically investigated with relatively simple experiments.

We also found significant overlap between *sek-1* regulated genes and genes regulated in response to Cry5 toxin but there was not between Cry5 responsive genes and *pmk-1* regulated genes (Table 6.2). We would therefore expect that *sek-1* mutant animals would be more sensitive to both stressors than would *pmk-1* mutants. Notably, Huffman *et al*. (2004a) found *sek-1* to be more sensitive to both toxins whereas *pmk-1* was not, confirming this prediction. Interestingly this aspect of the gene expression results was not discussed specifically in Huffman *et al.* (2004a).

There was only a very minor overlap between the two lists of genes that were identified by different groups to be regulated in response to exposure to *S. marcescens*. This could be due to the different platforms used in the expression analysis and because

so few genes were identified by Mallo *et al.* (2002). We also found that *C. elegans* transcriptional response to the fungus *D. coniospora* shared almost no genes in common with any of the other lists of genes suggesting that response and defense against fungal pathogens may be entirely different than response to stress and bacterial pathogens. It also seems possible that this response could be regulated through a yet to be discovered pathway in *C. elegans* as there is little overlap between fungal infection and *daf-16/daf-2, sek-1* or *pmk-1* regulated genes.

While the collection of datasets presented here is the most comprehensive to date, we cannot rule out the existence of other transcriptional profiling studies that were perhaps overlooked. However, we were able to observe new patterns observed, and thus the analysis contributes to the study of *C. elegans* innate immunity. We do not intend for this to be a rigorous meta-analysis and held to the same statistical standards, however it is a review of literature that could potentially be used to generate hypotheses and potentially shed some light on lists of genes that have been reported. If there was a lack of significance for the overlap of a particular comparison it could be due to the differences inherent in the studies compared as different platforms were used to generate the lists of genes discussed. However, as a proof of concept, predictions made herein were confirmed with available data suggesting the method could be robust to some experimental differences.

Conclusions

Overall, it appears that we have only just begun to characterize *C. elegans* innate immunity. The type of analysis shown here may prove to be a useful way to synthesize

data contained in analyese of differential expression, which will likely only get better as more datasets are added. We have shown that predictions based on significant overlap between microarray-identified genes can be confirmed through review of the literature, and further hypothesis testing will be needed to confirm other predictions. It seems that innate immunity is complex; with each new bacterial environment illustrating the surprising specificity of *C. elegans* defense response.

Figures and Tables

Figure 6.1 *C. elegans* defense pathways

C. elegans defense pathways are illustrated with known connections from the literature between genes and crosstalk between pathways. Inputs into the various pathways are shown as well as the documented outputs. The dotted line between the p38 MAPK pathway (SEK-1, PMK-1) is a known interaction between the two pathways, however the particular level where the crosstalk occurs is not well understood. Other dotted lines are predicted interactions of general stress and defense pathways. Diagram is modified from Troemel *et al*. (2006)

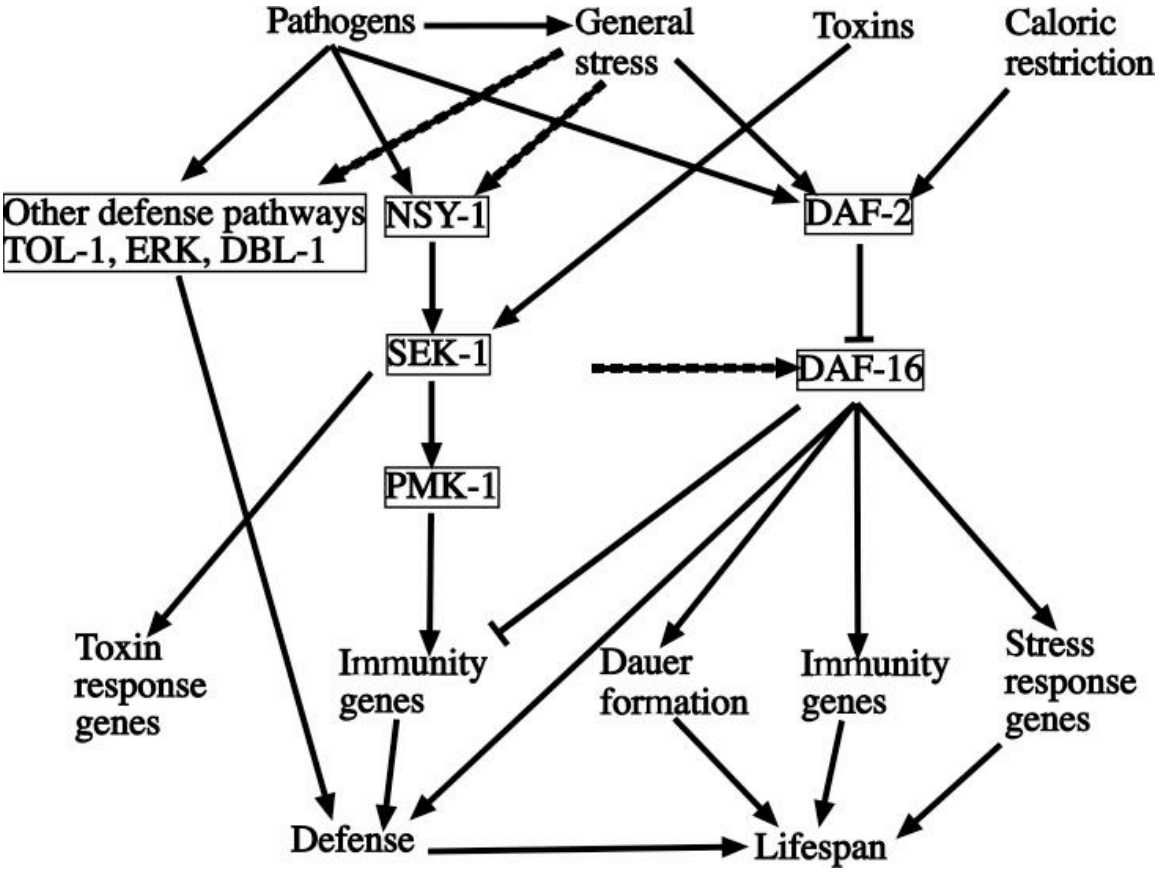


Figure 6.2 *C. elegans* defense pathways updated

Defense pathways are as described in Figure 1.1 with the addition of predicted interactions from the comparison of transcriptional profiling experiments shown in blue.

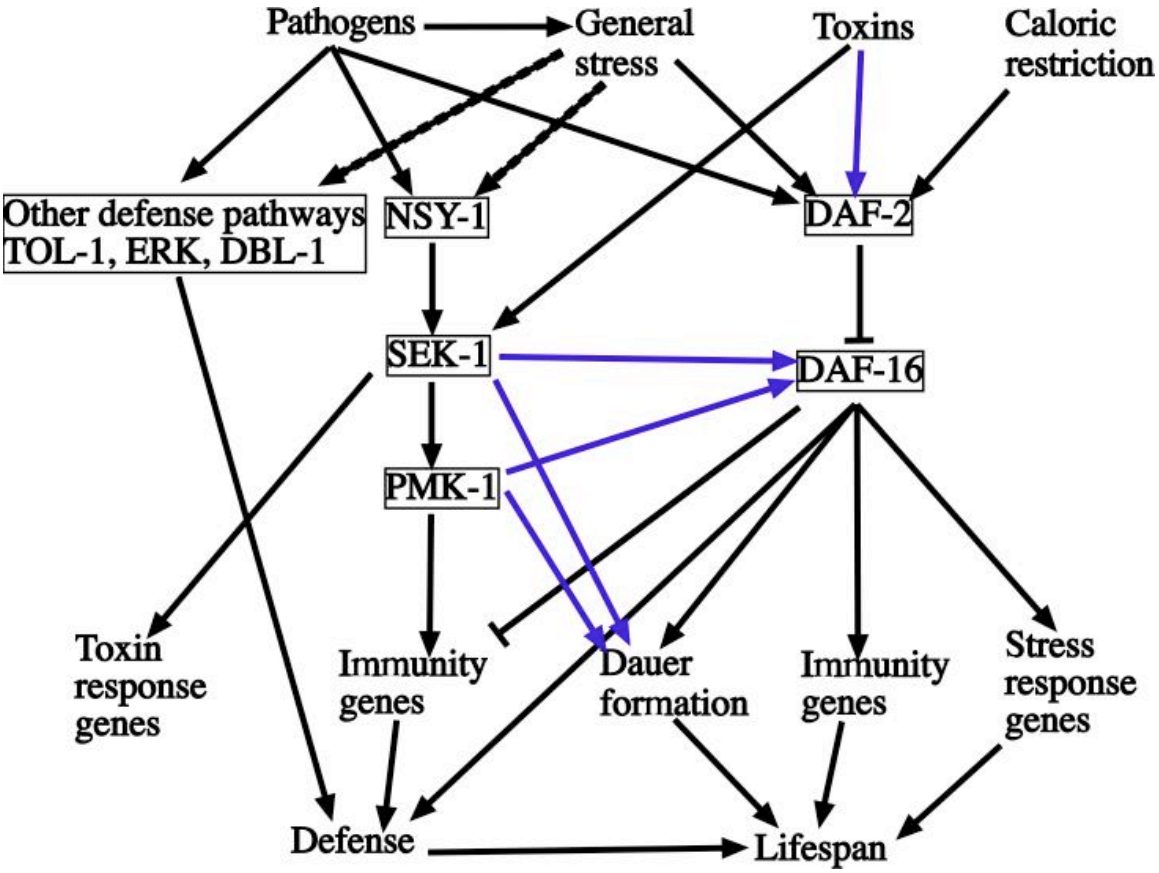


Table 6.1 Summary of innate immunity microarray experiment details

Summary of microarray details including the platform used, number of genes identified and number of biological replications (Reps) performed. In some cases the number of replications was not reported (NR).

Experiment treatment	Citation	Platform	# Genes identified	# Reps
P. aeruginosa 4hrs	Troemel *et al.* 2006	Affymetrix GeneChip	432	3
P. aeruginosa 8hrs	Troemel *et al.* 2006	Affymetrix GeneChip	506	3
M. nematophilum	O'Rourke *et al.* 2006	Affymetrix GeneChip	79	3
S. marcescens	Mallo *et al.* 2002	Nylon filter macroarray	10	2
S. marcescens	Wong *et al.* 2007	Spotted Wash U	1239	6
E. faecalis	Wong *et al.* 2007	Spotted Wash U	1258	6
E. carotovora	Wong *et al.* 2007	Spotted Wash U	1029	6
P. luminescens	Wong *et al.* 2007	Spotted Wash U	1265	6
B. megaterium vs. *E. coli*	Coolon *et al.* In Prep	Spotted Wash U	55	6
B. megaterium vs. *M. luteus*	Coolon *et al.* In Prep	Spotted Wash U	25	6
B. megaterium vs. *Pseudomonas sp.*	Coolon *et al.* In Prep	Spotted Wash U	81	6
E. coli vs. *M. luteus*	Coolon *et al.* In Prep	Spotted Wash U	41	6
E. coli vs. *Pseudomonas sp.*	Coolon *et al.* In Prep	Spotted Wash U	62	6
M. luteus vs. *Pseudomonas sp.*	Coolon *et al.* In Prep	Spotted Wash U	108	6
D. coniospora	Couillault *et al.* 2004	Nylon filter macroarray	6	NR
Cry5B toxin	Huffman *et al.* 2004	Affymetrix GeneChip	815	3
daf-16 class 1	Murphy *et al.* 2003	Spotted UCSF	259	NR
daf-16 class 2	Murphy *et al.* 2003	Spotted UCSF	250	NR
Dauer	McElwee *et al.* 2004	Affymetrix GeneChip	2466	5
pmk-1	Troemel *et al.* 2006	Affymetrix GeneChip	36	3
sek-1	Troemel *et al.* 2006	Spotted Stanford	110	3

Table 6.2 Comparison of defense regulated genes

All pair-wise comparisons between lists of genes were made. Representation factors and statistical significance of the overlap are given in each cell (RF, *p*-value). Cells are color coded by significance with blue: *P* < 1e-20, green: *P* = 1e-20 to 1e-2, white: NS. Column and row labels: *Pa* = *P. aeruginosa*, *Mn* = *M. nematophilum*, c1 = class 1, c2 = class 2, *Sm* = *S. marcescens*, *Sm* M = *S. marcescens* Mallo *et al*. 2003, *Ef* = *E. faecalis*, *Ec* = *E. carotovora*, *Pl* = *P. luminescens*, *Dc* = *D. coniospora*, B = *B. megaterium*, M = *M. luteus*, P = *Pseudomonas sp*. and E = *E. coli* OP50.

	Pa 4Hr	*Pa* 8Hr	*Mn*	Cry5	*daf-16* c1	*daf-16* c2	*pmk-1*	*sek-1*	*Sm*	*Ef*	*Ec*	*Pl*	*Sm M*	Dauer	*Dc*	Bvs.E	Bvs.M	Bvs.P	Evs.M	Evs.P
Pa 8Hr	26.6, 1e-100																			
Mn	17.6, 8.06e-30	15.5, 3.47e-29																		
Cry5	8.3, 4.28e-95	7.0, 4.21e-82	6.2, 3.46e-11																	
daf-16 c1	5.5, 1.08e-14	6.3, 3.52e-21	9.8, 7.33e-8	4.1, 3.36e-15																
daf-16 c2	5.2, 1.13e-12	4.4, 4.82e-11	7.1, 5.91e-5	3.1, 1.02e-8	0.0, 0.038															
pmk-1	14.1, 1.56e-10	9.9, 2.03e-7	28.1, 1.21e-5	3.4, 0.015	2.1, 0.375	22.2, 1.49e-11														
sek-1	16.0, 1.17e-35	12.2, 7.76e-28	18.4, 1.22e-8	5.8, 2.27e-13	5.6, 9.32e-5	28.4, 1.18e-46	85.9, 8.3e-30													
Sm	0.7, 0.066	0.7, 0.044	0.2, 0.039	1.3, 0.03	1.5, 0.032	1.3, 0.150	0.0, 0.100	0.7, 0.317												
Ef	1.7, 6.22e-4	1.9, 3.59e-7	2.2, 0.010	1.5, 2.25e-4	2.0, 8.94e-5	1.6, 0.015	1.9, 0.153	3.0, 4,31e-6	1.9, 1.23e-15											
Ec	1.3, 0.059	1.3, 0.068	0.2, 0.081	1.0, 0.455	1.1, 0.464	1.2, 0.218	2.7, 0.036	2.8, 1.58e-4	2.1, 3.76e-17	2.6, 1.58e-30										
Pl	2.3, 4.38e-10	2.2, 7.70e-11	2.0, 0.027	1.8, 7.78e-8	1.7, 0.004	2.3, 1.21e-6	2.6, 0.024	3.6, 1.67e-8	1.6, 1.01e-8	2.6, 1.4e-37	2.2, 2.3e-18									
Sm M	4.6, 0.196	4.0, 0.226	25.3, 0.039	4.9, 0.060	15.4, 0.007	0.0, 0.118	111.1, 1.41e-4	36.4, 0.001	3.2, 0.124	3.2, 0.127	3.9, 0.090	7.9, 1.93e-4								
Dauer	2.0, 2.40e-13	2.1, 8.16e-17	2.2, 4.45e-4	2.5, 4.24e-48	3.2, 8.44e-29	1.6, 3.51e-4	3.4, 9.68e-6	3.5, 1.39e-15	0.9, 0.117	1.2, 0.017	0.9, 0.270	1.2, 0.003	1.6, 0.355							
Dc	15.4, 0.007	6.6, 0.143	0.0, 0.023	0.0, 0.221	0.0, 0.075	0.0, 0.073	185.2, 4.71e-5	30.3, 0.033	5.4, 0.049	5.3, 0.050	3.2, 0.272	2.6, 0.324	333.3, 0.003	1.4, 0.454						
Bvs.E	3.4, 0.031	5.7, 6.80e-5	13.8, 0.001	3.1, 0.007	12.6, 3.37e-8	2.9, 0.151	10.1, 0.094	13.2, 2.38e-4	2.6, 0.006	5.5, 4.17e-10	2.5, 0.022	3.2, 5.64e-4	0.0, 0.027	2.5, 2.22e-4	0.0, 0.016					
Bvs.M	14.8, 3.48e-8	11.1, 2.06e-6	0.0, 0.094	5.9, 4.09e-4	12.4, 2.81e-4	0.0, 0.270	0.0P < 0.044	0.0, 0.129	2.6, 0.066	5.1, 9.88e-5	0.8, 0.371	4.4, 6.97e-4	0.0, 0.012	3.6, 8.31e-5	0.0, 0.007	101.8, 3.70e-13				
Bvs.P	2.9, 0.031	4.4, 2.05e-4	9.4, 0.004	3.3, 4.32e-4	8.6, 1.03e-6	4.0, 0.019	6.9, 0.136	9.0, 0.001	2.4, 0.004	4.7, 8.51e-11	3.1, 2.37e-4	3.3, 1.01e-5	24.7, 0.040	2.5, 8.42e-6	0.0, 0.024	103.3, 5.06e-42	49.4, 4.79e-8			
Evs.M	10.2, 1.79e-7	3.9, 0.020	18.5, 5.68e-4	4.2, 0.001	17, 2.18e-9	2.0, 0.403	13.6, 0.071	17.7, 7.5e-5	1.6, 0.248	5.0, 7.61e-7	1.4, 0.354	4.6, 5.52e-6	48.8, 0.020	3.2, 1.34e-5	0.0 P 0.012	141.9, 9.38e-32	195.1, 1.27e-21	96.4, 1.03e-28		
Evs.P	0.7, 0.388	3.2, 0.020	0.0, 0.218	1.2, 0.466	2.5, 0.192	1.3, 0.458	17.9, 0.006	2.9, 0.290	3.6, 2.00e-5	3.6, 2.37e-5	1.6, 0.213	4.3, 1.82e-7	32.3, 0.031	2.7, 8.78e-6	0.0, 0.018	41.1, 3.51e-10	25.8, 0.003	79.6, 8.90e-34	31.5, 7.78e-6	
Mvs.P	3, 0.009	3.7, 4.15e-4	4.7, 0.068	4.1, 3.52e-7	7.2, 1.42e-6	4.4, 0.002	5.1, 0.177	5.1, 0.022	3.0, 9.33e-6	4.1, 7.86e-11	3.1, 3.58e-5	5.1, 2.21e-16	37.0, 0.001	3.1, 9.32e-12	0.0, 0.032	50.5, 3.47e-22	111.1, 1.10e-28	102.9, 2.32e-84	135.5, 3.37e-61	107.5, 6.20e-68

CHAPTER 7 - Conclusions

Understanding the genetic mechanisms that govern organismal responses to human-induced environmental change and disturbance is of prime importance for predicting the long-term effects of anthropogenic change. This is especially true in the grasslands of North America, where the tallgrass prairie is one of the most endangered ecosystem types. Previous data from the Konza Prairie LTER site has shown that many components of the above- and below-ground food web are affected by various disturbance regimes. Our group has previously shown that one of the most sensitive components of the soil food web, bacterial-feeding nematodes, responds to many disturbances including the addition of nitrogen fertilizer as well as alterations in burn frequency. Here we have shown that the soil microbial community also responds to the same perturbations. In addition, we believe that the changes in the microbial community could potentially have huge effects on the next trophic level, the bacterial-feeding nematodes.

To investigate the genes involved in responses to different bacterial environments as a driver for nematode community responses, we moved into the laboratory and used the genetic model organism *Caenorhabditis elegans* to model the interaction of grassland nematodes and soil bacteria. We identified many bacterially induced genes that in turn affected fitness and lifespan in the lab. Through use of the genetic model we have identified many candidate genes to investigate further in native nematode species. After the identification of the genes important for life in different bacterial environments, we

further began to elucidate the genetic and signaling pathways through which they may function to determine resistance to bacterial pathogens. Our data suggests that through metabolic and defense functions, the genes in native nematode genomes are likely controlling pathogenic and trophic interactions with soil bacteria. These interactions could be in part, those that dictate soil bacterial-feeding nematode responses to disturbances in the tallgrass prairie. The next step will be to determine which genes are conserved in native nematode species. Once the genes are identified in the native nematodes, their function should be determined through loss of function analysis by RNAi. Further work with *C. elegans* will help us to better understand the bacterially induced genes to facilitate genetic analysis in the native species. Finally, through evolutionary studies, we feel that it will be possible to determine the long-term importance of genes through signatures of selection in the native nematodes genomes.

References

Aballay, A., and F. M. Ausubel. 2002. *Caenorhabditis elegans* as a host for the study of host-pathogen interactions. Current Opinion In Microbiology **5**:97-101.

Aballay, A., E. Drenkard, L. R. Hilbun, and F. M. Ausubel. 2003. *Caenorhabditis elegans* innate immune response triggered by *Salmonella enterica* requires intact LPS and is mediated by a MAPK signaling pathway. Current Biology **13**:47-52.

Ajwa, H. A., C. J. Dell, and C. W. Rice. 1999. Changes in enzyme activities and microbial biomass of tallgrass prairie soil as related to burning and nitrogen fertilization. Soil Biology & Biochemistry **31**:769-777.

Allison, V. J., R. M. Miller, J. D. Jastrow, R. Matamala, and D. R. Zak. 2005. Changes in soil microbial community structure in a tallgrass prairie chronosequence. Soil Science Society of America Journal **69**:1412-1421.

Ausubel, F. M. 2005. Are innate immune signaling pathways in plants and animals conserved? Nature Immunology **6**:973-979.

Baath, E., A. Frostegard, T. Pennanen, and H. Fritze. 1995. Microbial community structure and pH response in relation to soil organic-matter quality in wood-ash fertilized, clear-cut or burned coniferous forest soils. Soil Biology & Biochemistry **27**:229-240.

Baer, S. G., D. J. Kitchen, J. M. Blair, and C. W. Rice. 2002. Changes in ecosystem structure and function along a chronosequence of restored grasslands. Ecological Applications **12**:1688-1701.

Barriere, A., and M. Felix. 2007. Temporal Dynamics and linkage disequilibrium in Natural *Caenorhabditis elegans* populations. Genetics **176**:999-1011.

Barsyte, D., D. A. Lovejoy, and G. J. Lithgow. 2001. Longevity and heavy metal resistance in daf-2 and age-1 long-lived mutants of *Caenorhabditis elegans*. FASEB Journal **15**:627-634.

Bittman, S., T. A. Forge, and C. G. Kowalenko. 2005. Responses of the bacterial and fungal biomass in a grassland soil to multi-year applications of dairy manure slurry and fertilizer. Soil Biology & Biochemistry **37**:613-623.

Blair, J. M. 1997. Fire, N availability, and plant response in grasslands: A test of the transient maxima hypothesis. Ecology **78**:2359-2368.

Blair, J. M., P. J. Bohlen, and D. W. Freckman. 1996. Soil invertebrates as indicators of soil quality. Pages 283-301 *in* A. J. Jones, editor. Methods for assessing soil quality. SSSA Special Publication No. 49, Soil Science Society of America, Madison, WI.

Blair, J. M., T. C. Todd, and J. Callaham, M.A. 2000. Responses of grassland soil invertebrates to natural and anthropogenic disturbances. Pages 43-71 *in* P. F. Hendrix, editor. Invertebrates as Webmasters in Ecosystems. CAB International Press, Wallingford, UK.

Blaxter, M. L. 1993. Cuticle Surface-Proteins Of Wild-Type And Mutant *Caenorhabditis elegans*. Journal Of Biological Chemistry **268**:6600-6609.

Blumenthal, T., R. D. Evans, C. Link, A. Guffanti, D. Lawson, D. Terry-mieg, W. L. Chiu, K. Duke, M. Kiraly, and S. K. Kim. 2002. A Global Analysis of *Caenorhabditis elegans* Operons. Nature **417**:851-854.

Booijink, C., E. G. Zoetendal, M. Kleerebezem, and W. M. de Vos. 2007. Microbial communities in the human small intestine: coupling diversity to metagenomics. Future Microbiology **2**:285-295.

Brazma, A., P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton, T. Gaasterland, P. Glenisson, F. C. P. Holstege, I. F. Kim, V. Markowitz, J. C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, and M. Vingron. 2001. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. Nature Genetics **29**:365-371.

Brenner, S. 1974. The genetics of *Caenorhabditis elegans*. Genetics **77**:71-94.

Callaham, M. A., J. M. Blair, T. C. Todd, D. J. Kitchen, and M. R. Whiles. 2003. Macroinvertebrates in North American tallgrass prairie soils: effects of fire, mowing, and fertilization on density and biomass. Soil Biology & Biochemistry **35**:1079-1093.

Chen, J., D. Senturk, J. Wang, H. Muller, J. R. Carey, H. Caswell, and E. Caswell-Chen. 2007. A Demographic Analysis of the Fitness Cost of Extended Longevity in *Caenorhabditis elegans*. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences **62**:126-135.

Churchill, G. A., and B. Oliver. 2001. Sex, flies and microarrays. Nature Genetics **29**:355-356.

Coleman, J. S., L. Rochefort, F. A. Bazzaz, and F. I. Woodward. 1991. Atmospheric CO2, Plant Nitrogen Status And The Susceptibility Of Plants To An Acute Increase In Temperature. Plant Cell And Environment **14**:667-674.

Collins, S. L., A. K. Knapp, J. M. Briggs, J. M. Blair, and E. M. Steinauer. 1998. Modulation of diversity by grazing and mowing in native tallgrass prairie. Science **280**:745-747.

Consortium, T. C. e. S. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282**:2012-2018.

Coolon, J. D., K. L. Jones, T. C. Todd, B. C. Carr, and M. A. Herman. In Prep. Genomic basis of nematode-bacteria interactions.

Couillault, C., N. Pujol, J. Reboul, L. Sabatier, J. F. Guichou, Y. Kohara, and J. J. Ewbank. 2004. TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. Nature Immunology **5**:488-494.

Cui, Y. X., S. J. McBride, W. A. Boyd, S. Alper, and J. H. Freedman. 2007. Toxicogenomic analysis of *Caenorhabditis elegans* reveals novel genes and pathways involved in the resistance to cadmium toxicity. Genome Biology **8**.

Curry, J. P. 1994. Grassland Invertebrates. Ecology, Influence on Soil Fertility and Effects on Plant Growth. Chapman and Hall., New York.

Dai, X., K. Sayama, M. Tohyama, Y. Shirakata, L. Yang, S. Hirakawa, S. Tokumaru, and K. Hashimoto. 2008. The NF-KappaB, p38 MAPK and STAT1 pathways differentially regulate the dsRNA-mediated innate immunity responses of epidermal keratinocytes. International Immunology **20**:901-909.

Daily, G. C. 1997. The potential impacts of global warming on managed and natural ecosystem: Implications for human well-being. Abstracts Of Papers Of The American Chemical Society **213**:12-ENVR.

Dell, C. J., and C. W. Rice. 2005. Short-term competition for ammonium and nitrate in tallgrass prairie. Soil Science Society of America Journal **69**:371-377.

Dell, C. J., M. A. Williams, and C. W. Rice. 2005. Partitioning of nitrogen over five growing seasons in tallgrass prairie. Ecology **86**:1280-1287.

Dionisi, H. M., A. C. Layton, G. Harms, I. R. Gregory, K. G. Robinson, and G. S. Sayler. 2002. Quantification of *Nitrosomonas oligotropha*-like ammonia-oxidizing bacteria and *Nitrospira spp*. from full-scale wastewater treatment plants by competitive PCR. Applied and Environmental Microbiology **68**:245-253.

Dobson, A. P. 1997. Hopes for the future: Restoration ecology and conservation biology. Science **277**:515-522.

Ewbank, J. J. 2006. Signaling in the immune response.WormBook Online.

Fire, A., S. Xu, M. Montgomery, S. Kostas, S. Driver, and C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature **391**:806-811.

Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The relationship between the number of species and the number of individuals in a random sample of an animal population. Journal of Animal Ecology **12**:42-58.

Foley, J. A. 2005. Global consequences of land use. Science **309**:570-574.

Freckman, D. W. 1988. Bacterivorous Nematodes and Organic-Matter Decomposition. Agriculture Ecosystems & Environment **24**:195-217.

Freckman, D. W., and C. H. Ettema. 1993. Assessing Nematode Communities in Agroecosystems of Varying Human Intervention. Agriculture Ecosystems & Environment **45**:239-261.

Frey, S. D., E. T. Elliott, and K. Paustian. 1999. Bacterial and fungal abundance and biomass in conventional and no-tillage agroecosystems along two climatic gradients. Soil Biology & Biochemistry **31**:573-585.

Garsin, D. A., J. M. Villanueva, J. Begun, D. H. Kim, C. D. Sifri, S. B. Calderwood, G. Ruvkun, and F. M. Ausubel. 2003. Long-lived *C. elegans daf-2* mutants are resistant to bacterial pathogens. Science **300**:1921-1921.

Gottlieb, S., and G. Ruvkun. 1994. *daf-2*, *daf-16* and *daf-23* - Genetically Interacting Genes-Controlling Dauer Formation In *Caenorhabditis elegans*. Genetics **137**:107-120.

Gravato-Nobre, M. J., and J. Hodgkin. 2005. *Caenorhabditis elegans* as a model for innate immunity to pathogens. Cellular Microbiology **7**:741-751.

Groffman, P. M., C. W. Rice, and J. M. Tiedje. 1993. Denitrification in a tallgrass prairie landscape. Ecology **74**:855-862.

Hamilton, B., Y. Q. Doug, M. Shindo, W. Y. Liu, I. Odell, G. Ruvkun, and S. S. Lee. 2005. A systematic RNAi screen for longevity genes in *C. elegans*. Genes & Development **19**:1544-1555.

Hannah, L. 1995. Human Disturbance And Natural Habitat - A Biome Level Analysis Of A Global Data Set. Biodiversity and Conservation **4**:128-155.

Hansen, A. A., R. A. Herbert, K. Mikkelsen, L. L. Jensen, T. Kristoffersen, J. M. Tiedje, B. A. Lomstein, and K. W. Finster. 2007. Viability, diversity and composition of the bacterial community in a high Arctic permafrost soil from Spitsbergen, Northern Norway. Environmental Microbiology **9**:2870-2884.

Hartman, P. S., and N. Ishii. 2007. Chromosome dosage as a life span determinant in *Caenorhabditis elegans*. Mechanisms Of Ageing And Development **128**:437-443.

Hiraishi, A., and Y. Ueda. 1994. *Rhodoplanes* gen. nov., a new genus of phototrophic bacteria including *Rhodopseudomonas rosea* as *Rhodoplanes roseus* comb. nov. and *Rhodoplanes elegans sp.* nov. International Journal of Systematic Bacteriology **44**:665–673.

Hodgkin, J., and T. M. Barnes. 1991. More is Not Better: Brood Size and Population Growth in a Self-Fertilizing Nematode. Proceedings of the Royal Society: Biological Sciences **246**:19-24.

Holmes, W. E., and D. R. Zak. 1999. Soil microbial control of nitrogen loss following clearcut harvest in northern hardwood ecosystems. Ecological Applications **9**:202-215.

Hooper, A. B., T. Vannelli, D. J. Bergmann, and D. M. Arciero. 1997. Enzymology of the oxidation of ammonia to nitrite by bacteria. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology **71**:59-67.

Hsu, A. L., C. T. Murphy, and C. Kenyon. 2003. Regulation of aging and age-related disease by DAF-16 and heat-shock factor. Science **300**:1142-1145.

Huang, X. Q., and A. Madan. 1999. CAP3: A DNA sequence assembly program. Genome Research **9**:868-877.

Huffman, D. L., L. Abrami, R. Sasik, J. Corbeil, F. G. van der Goot, and R. V. Aroian. 2004a. Mitogen-activated protein kinase pathways defends against bacterial pore-forming toxins. Proceedings Of The National Academy Of Sciences Of The United States Of America **101**:10995-11000.

Huffman, D. L., L. J. Bischof, J. S. Griffitts, and R. V. Aroian. 2004b. Pore worms: Using *Caenorhabditis elegans* to study how bacterial toxins interact with their target host. International Journal Of Medical Microbiology **293**:599-607.

Jones, K. L., T. C. Todd, and M. A. Herman. 2006a. Development of taxon-specific markers for high-throughput screening of microbial-feeding nematodes. Molecular Ecology Notes **6**:712-714.

Jones, K. L., T. C. Todd, J. L. Wall-Beam, J. D. Coolon, J. M. Blair, and M. A. Herman. 2006b. Molecular approach for assessing responses of microbial-feeding nematodes to burning and chronic nitrogen enrichment in a native grassland. Molecular Ecology **15**:2601 –2609.

Jones, K. L., T. C. Todd, J. L. Wall-Beam, J. D. Coolon, J. M. Blair, and M. A. Herman. 2006c. Molecular approach for assessing responses of microbial-feeding nematodes to burning and chronic nitrogen enrichment in a native grassland. Molecular Ecology **15**:2601-2609.

Juretschko, S., G. Timmermann, M. Schmid, K. H. Schleifer, A. Pommerening-Roser, H. P. Koops, and M. Wagner. 1998. Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and

Nitrospira-like bacteria as dominant populations. Applied and Environmental Microbiology **64**:3042-3051.

Kaeberlein, T. L., E. D. Smith, M. Tsuchiya, K. L. Welton, J. H. Thomas, S. Fields, B. K. Kennedy, and M. Kaeberlein. 2006. Lifespan extension in *Caenorhabditis elegans* by complete removal of food. Aging Cell **5**:487-494.

Kamath, R. K., M. Martinez-Campos, P. Zipperlen, A. Fraser, and J. Ahringer. 2001. Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. Genome Biology **2**:1-10.

Kamath, R. S., A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D. P. Welchman, P. Zipperlen, and J. Ahringer. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. **421**:231-237.

Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, A. Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima, M. Kohara, M. Matsumoto, S. Shimpo, H. Tsuruoka, T. Wada, M. Yamada, and S. Tabata. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. DNA Research **9**:189-197.

Kim, D. H., R. Feinbaum, G. Alloing, F. E. Emerson, D. A. Garsin, H. Inoue, M. Tanaka-Hino, N. Hisamoto, K. Matsumoto, M. W. Tan, and F. M. Ausubel. 2002. A conserved p38 MAP kinase pathway in *Caenorhabditis elegans* innate immunity. Science **297**:623-626.

Kim, D. H., N. T. Liberati, T. Mizuno, H. Inoue, N. Hisamoto, K. Matsumoto, and F. M. Ausubel. 2004. Integration of Caenorhabditis elegans MAPK pathways mediating immunity and stress resistance by MEK-1 MAPK kinase and VHP-1 MAPK phosphatase. Proceedings Of The National Academy Of Sciences Of The United States Of America **101**:10990-10994.

Kim, S. K., J. Lund, M. Kiraly, K. Duke, M. Jiang, J. M. Stuart, A. Eizinger, B. N. Wylie, and G. S. Davidson. 2001. A gene expression map for *Caenorhabditis elegans*. Science **293**:2087-2092.

Knapp, A. K., J. M. Briggs, D. C. Hartnett, and S. L. Collins, editors. 1998. Grassland Dynamics. Long-Term Ecological Research in Tallgrass Prairie. Oxford University Press, New York.

Knapp, A. K., P. A. Fay, J. M. Blair, S. L. Collins, M. D. Smith, J. D. Carlisle, C. W. Harper, B. T. Danner, M. S. Lett, and J. K. McCarron. 2002. Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. Science **298**:2202-2205.

Konstantinidis, K. T., A. Ramette, and J. M. Tiedje. 2006. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Applied and Environmental Microbiology **72**:7286-7293.

Konstantinidis, K. T., and J. M. Tiedje. 2007. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Current Opinion in Microbiology **10**:504-509.

Kramer, J. M., R. P. French, E. C. Park, and J. J. Johnson. 1990. The *Caenorhabditis elegans rol-6* Gene, Which Interacts With The *sqt-1* Collagen Gene To Determine Organismal Morphology, Encodes A Collagen. Molecular And Cellular Biology **10**:2081-2089.

Kramer, J. M., and J. J. Johnson. 1993. Analysis Of Mutations In The *sqt-1* And *rol-6* Collagen Genes Of *Caenorhabditis elegans*. Genetics **135**:1035-1045.

Labrousse, A., S. Chauvet, C. Couillault, C. L. Kurz, and J. J. Ewbank. 2000. *Caenorhabditis elegans* is a model host for *Salmonella typhimurium*. Current Biology **10**:1543-1545.

Lakowski, B., and S. Hekimi. 1998. The genetics of caloric restriction in *Caenorhabditis elegans*. Proceedings Of The National Academy Of Sciences Of The United States Of America **95**:13091-13096.

Lauro, F. M., and D. H. Bartlett. 2008. Prokaryotic lifestyles in deep-sea habitats. Extremophiles **12**:15-25.

Lee, D. G., J. M. Urbach, G. Wu, N. T. Liberati, R. L. Feinbaum, S. Miyata, L. T. Diggins, J. X. He, M. Saucier, E. Deziel, L. Friedman, L. Li, G. Grills, K. Montgomery, R. Kucherlapati, L. G. Rahme, and F. M. Ausubel. 2006a. Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. Genome Biology **7**.

Lee, G. D., M. A. Wilson, M. Zhu, C. A. Wolkow, R. de Cabo, D. K. Ingram, and S. G. Zou. 2006b. Dietary deprivation extends lifespan in *Caenorhabditis elegans*. Aging Cell **5**:515-524.

Lee, S. S. 2006. Whole genome RNAi screens for increased longevity: Important new insights but not the whole story. Experimental Gerontology **41**:968-973.

Lee, S. S., S. Kennedy, A. C. Tolonen, and G. Ruvkun. 2003. DAF-16 target genes that control *C. elegans* life-span and metabolism. Science **300**:644-647.

Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G. W. Nicol, J. I. Prosser, S. C. Schuster, and C. Schleper. 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature **442**:806-809.

Mahajan-Miklos, S., L. G. Rahme, and F. M. Ausubel. 2000. Elucidating the molecular mechanisms of bacterial virulence using non-mammalian hosts. Molecular Microbiology **37**:981-988.

Mallo, G. V., C. L. Kurz, C. Couillault, N. Pujol, S. Granjeaud, Y. Kohara, and J. J. Ewbank. 2002. Inducible antibacterial defense system in *C. elegans*. Current Biology **12**:1209-1214.

Manduchi, E., L. M. Scearce, J. E. Brestelli, G. R. Grant, K. H. Kaestner, and C. J. Stoeckert. 2002. Comparison of different labeling methods for two-channel high-density microarray experiments. Physiological Genomics **10**:169-179.

Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376-380.

Marschner, P., E. Kandeler, and B. Marschner. 2003. Structure and function of the soil microbial community in a long-term fertilizer experiment. Soil Biology & Biochemistry **35**:453-461.

Masse, I., L. Molin, M. Billaud, and F. Solari. 2005. Lifespan and dauer regulation by tissue-specific activities of *Caenorhabditis elegans* DAF-18. Developmental Biology **286**:91-101.

Matyash, V., E. V. Entchev, F. Mende, M. Wilsch-Brauninger, C. Thiele, A. W. Schmidt, H. J. Knolker, S. Ward, and T. V. Kurzchalia. 2004. Sterol-derived hormone(s) controls entry into diapause in *Caenorhabditis elegans* by consecutive activation of DAF-12 and DAF-16. PloS Biology **2**:1561-1571.

McElwee, J., K. Bubb, and J. H. Thomas. 2003. Transcriptional outputs of the *Caenorhabditis elegans* forkhead protein DAF-16. Aging Cell **2**:111-121.

McElwee, J. J., E. Schuster, E. Blanc, J. H. Thomas, and D. Gems. 2004. Shared transcriptional signature in *Caenorhabditis elegans* dauer larvae and long-lived *daf-2* mutants implicates detoxification system in longevity assurance. Journal Of Biological Chemistry **279**:44533-44543.

Millet, A. C. M., and J. J. Ewbank. 2004. Immunity in *Caenorhabditis elegans*. Current Opinion In Immunology **16**:4-9.

Morozova, T. V., R. R. H. Anholt, and T. F. C. Mackay. 2006. Transcriptional response to alcohol exposure in *Drosophila melanogaster*. Genome Biology **7**.

Moy, T. I., A. R. Ball, Z. Anklesaria, G. Casadei, K. Lewis, and F. M. Ausubel. 2006. Identification of novel antimicrobials using a live-animal infection model. Proceedings Of The National Academy Of Sciences Of The United States Of America **103**:10414-10419.

Moy, T. I., E. Mylonakis, S. B. Calderwood, and F. M. Ausubel. 2004. Cytotoxicity of hydrogen peroxide produced by *Enterococcus faecium*. Infection And Immunity **72**:4512-4520.

Murphy, C. T. 2006. The search for DAF-16/FOXO transcriptional targets: Approaches and discoveries. Experimental Gerontology **41**:910-921.

Murphy, C. T., S. A. McCarroll, C. I. Bargmann, A. Fraser, R. S. Kamath, J. Ahringer, H. Li, and C. Kenyon. 2003. Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. Nature **424**:277-284.

Neal, D. 2004. Introduction to Population Biology. Cambridge University Press, Cambridge.

Nicholas, H. R., and J. Hodgkin. 2004a. The ERK MAP kinase cascade mediates tail swelling and a to rectal infection protective response in *C. elegans*. Current Biology **14**:1256-1261.

Nicholas, H. R., and J. Hodgkin. 2004b. Responses to infection and possible recognition strategies in the innate immune system of *Caenorhabditis elegans*. Molecular Immunology **41**:479-493.

O'Rourke, D., D. Baban, M. Demidova, R. Mott, and J. Hodgkin. 2006. Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. Genome Research **16**:1005-1016.

Park, Y. S., and J. M. Kramer. 1994. The *C. elegans sqt-1* And *rol-6* Collagen Genes Are Coordinately Expressed During Development, But Not At All Stages That Display Mutant Phenotypes. Developmental Biology **163**:112-124.

Patten, M. A., E. Shochat, D. L. Reinking, D. H. Wolfe, and S. K. Sherrod. 2006. Habitat edge, land management, and rates of brood parasitism in tallgrass prairie. Ecological Applications **16**:687-695.

Pietikainen, J., and H. Fritze. 1995. Clear-cutting and prescribed burning in coniferous forest - Comparison of effects on soil fungal and total microbial biomass, respiration activity and nitrification. Soil Biology & Biochemistry **27**:101-109.

Pisegna, S., G. Pirozzi, M. Piccoli, L. Frati, A. Santoni, and G. Palmieri. 2004. p38 MAPK activation controls the TLR3-mediated up-regulation of cytotoxicity and cytokine production in human NK cells. Blood **104**:4157-4164.

Pujol, N., E. M. Link, L. X. Liu, C. L. Kurz, G. Alloing, M. W. Tan, K. P. Ray, R. Solari, C. D. Johnson, and J. J. Ewbank. 2001. Reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. Current Biology **11**:809-821.

Purkhold, U., A. Pommerening-Roser, S. Juretschko, M. C. Schmid, H. P. Koops, and M. Wagner. 2000. Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and amoA sequence analysis: Implications for molecular diversity surveys. Applied and Environmental Microbiology **66**:5368-5382.

Regan, J. M., G. W. Harrington, and D. R. Noguera. 2002. Ammonia- and nitrite-oxidizing bacterial communities in a pilot-scale chloraminated drinking water distribution system. Applied and Environmental Microbiology **68**:73-81.

Roesch, L. F., R. R. Fulthorpe, A. Riva, G. Casella, A. K. M. Hadwin, A. D. Kent, S. H. Daroub, F. A. O. Camargo, W. G. Farmerie, and E. W. Triplett. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME Journal **1**:283-290.

Samson, F., and F. Knopf. 1994. Prairie conservation in North America. BioScience **44**:418-421.

Samuelson, A. V., C. E. Carr, and G. Ruvkun. 2007. Gene activities that mediate increased life span of *C. elegans* insulin-like signaling mutants. Genes & Development **21**:2976-2994.

Sarathchandra, S. U., A. Ghani, G. W. Yeates, G. Burch, and N. R. Cox. 2001. Effect of nitrogen and phosphate fertilisers on microbial and nematode diversity in pasture soils. Soil Biology & Biochemistry **33**:953-964.

Schulenburg, H., C. L. Kurz, and J. J. Ewbank. 2004. Evolution of the innate immune system: the worm perspective. Immunological Reviews **198**:36-58.

Schuster, S. C. 2008. Next-generation sequencing transforms today's biology. Nature Methods **5**:16-18.

Shimizu, K. K., and M. D. Purugganan. 2005. Evolutionary and ecological genomics of *arabidopsis*. Plant Physiology **138**:578-584.

Shochat, E., D. H. Wolfe, M. A. Patten, D. L. Reinking, and S. K. Sherrod. 2005. Tallgrass prairie management and bird nest success along roadsides. Biological Conservation **121**:399-407.

Shtonda, B. B., and L. Avery. 2006. Dietary choice behavior in *Caenorhabditis elegans*. Journal Of Experimental Biology **209**:89-102.

Sifri, C. D., J. Begun, and F. M. Ausubel. 2005. The worm has turned - microbial virulence modeled in *Caenorhabditis elegans*. Trends In Microbiology **13**:119-127.

Sifri, C. D., J. Begun, F. M. Ausubel, and S. B. Calderwood. 2003. *Caenorhabditis elegans* as a model host for *Staphylococcus aureus* pathogenesis. Infection And Immunity **71**:2208-2217.

Sifri, C. D., E. Mylonakis, K. V. Singh, X. Qin, D. A. Garsin, B. E. Murray, F. M. Ausubel, and S. B. Calderwood. 2002. Virulence effect of *Enterococcus faecalis* protease genes and the quorum-sensing locus fsr in *Caenorhabditis elegans* and mice. Infection And Immunity **70**:5647-5650.

Singh, V., and A. Aballay. 2006. Heat-shock transcription factor (HSF)-1 pathway required for *Caenorhabditis elegans* immunity. Proceedings Of The National Academy Of Sciences Of The United States Of America **103**:13092-13097.

Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences of the United States of America **103**:12115-12120.

Sotomayor, D., and C. W. Rice. 1996. Denitrification in soil profiles beneath grassland and cultivated soils. Soil Science Society of America Journal **60**:1822-1828.

Storey, J. D., and R. Tibshirani. 2003. Statistical significance for genomewide studies. Proceedings Of The National Academy Of Sciences Of The United States Of America **100**:9440-9445.

Sulston, J., and J. Hodgkin. 1988. The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sulston, J., E. Schierenberg, J. G. White, and J. N. Thomson. 1983. The Embryonic cell lineage of the nematode *Caenorhabditis elegans*. Developmental Biology **100**:64-119.

Szewczyk, N. J., E. Kozak, and C. A. Conley. 2003. Chemically defined medium and *Caenorhabditis elegans*. BMC Biotechnology **3**.

Tan, M. W., and F. M. Ausubel. 2000. *Caenorhabditis elegans*: a model genetic host to study *Pseudomonas aeruginosa* pathogenesis. Current Opinion In Microbiology **3**:29-34.

Tan, M. W., and F. M. Ausubel. 2002. Alternative models in microbial pathogens. Pages 461-475 *in* Molecular Cellular Microbiology. Academic Press Inc, San Diego.

Tan, M. W., S. Mahajan-Miklos, and F. M. Ausubel. 1999a. Killing of *Caenorhabditis elegans* by *Pseudomonas aeruginosa* used to model mammalian bacterial pathogenesis. Proceedings Of The National Academy Of Sciences Of The United States Of America **96**:715-720.

Tan, M. W., L. G. Rahme, J. A. Sternberg, R. G. Tompkins, and F. M. Ausubel. 1999b. *Pseudomonas aeruginosa* killing of *Caenorhabditis elegans* used to identify *P. aeruginosa* virulence factors. Proceedings Of The National Academy Of Sciences Of The United States Of America **96**:2408-2413.

Tenor, J. L., and A. Aballay. 2008. A conserved Toll-like receptor is required for *Caenorhabditis elegans* innate immunity. EMBO Reports **9**:103-109.

Tenor, J. L., B. A. McCormick, F. M. Ausubel, and A. Aballay. 2004. *Caenorhabditis elegans*-based screen identifies *Salmonella* virulence factors required for conserved host-pathogen interactions. Current Biology **14**:1018-1024.

Todd, T. C. 1996. Effects of management practices on nematode community structure in tallgrass prairie. Applied Soil Ecology **3**:235-246.

Todd, T. C., J. M. Blair, and G. A. Milliken. 1999. Effects of altered soil-water availability on a tallgrass prairie nematode community. Applied Soil Ecology **13**:45-55.

Troemel, E. R., S. W. Chu, V. Reinke, S. S. Lee, F. M. Ausubel, and D. H. Kim. 2006. p38 MAPK regulates expression of immune response genes and contributes to longevity in *C. elegans*. Plos Genetics **2**:1725-1739.

Trosvik, P., B. Skanseng, K. S. Jakobsen, N. C. Stenseth, T. Naes, and K. Rudi. 2007. Multivariate analysis of complex DNA sequence electropherograms for high-throughput quantitative analysis of mixed microbial populations. Applied and Environmental Microbiology **73**:4975-4983.

Turner, C. L., J. M. Blair, R. J. Schartz, and J. C. Neel. 1997. Soil N and plant responses to fire, topography, and supplemental N in tallgrass prairie. Ecology **78**:1832-1843.

Ungerer, M. C., L. C. Johnson, and M. A. Herman. 2008. Ecological genomics: understanding gene functions in the natural environment. Heredity **100**:178-183.

Vazquez, F. J., M. J. Acea, and T. Carballas. 1993. Soil microbial populations after wildfire. Fems Microbiology Ecology **13**:93-103.

Watanabe, K., Y. Kodama, and S. Harayama. 2001. Design and evaluation of PCR primers to amplify bacterial 16S ribosomal DNA fragments used for community fingerprinting. Journal of Microbiological Methods **44**:253-262.

Weinig, C., L. A. Dorn, N. C. Kane, Z. M. German, S. S. Halldorsdottir, M. C. Ungerer, T. Y., T. F. Mackay, M. D. Purugganan, and S. J. 2003. Heterogeneous selection at specific loci in natural environments in *Arabidopsis thaliana*. Genetics **165**:321-329.

Williams, M. A., and C. W. Rice. 2007. Seven years of enhanced water availability influences the physiological, structural, and functional attributes of a soil microbial community. Applied Soil Ecology **35**:535-545.

Wittkopp, P. J. 2007. Variable gene expression in eukaryotes: a network perspective. Journal Of Experimental Biology **210**:1567-1575.

Wolfinger, R. D., G. Gibson, E. D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R. S. Paules. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. Journal of Computational Biology **8**:625-637.

Wong, D., D. Bazopoulou, N. Pujol, N. Tavernarakis, and J. J. Ewbank. 2007. Genome-wide investigation reveals pathogen-specific and shared signatures in the response of *Caenorhabditis elegans* to infection. Genome Biology **8**.

Wood, W. B. 1988. The Nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory Press, New York.

Xie, C.-H., and A. Yokota. 2006. *Sphingomonas azotifigens* sp. nov., a nitrogen-fixing bacterium isolated from the roots of *Oryza sativa*. International Journal of Systematic and Evolutionary Microbiology **56**:889-893.

Yeates, G. W. 2003. Nematodes as soil indicators: functional and biodiversity aspects. Biology and Fertility of Soils **37**:199-210.
Yonker, S. A., and B. J. Meyer. 2003. Recruitment of *C. elegans* dosage compensation proteins for gene-specific versus chromosome-wide repression. Development **130**:6519-6532.

Appendix A - Differentially expressed genes

Table 8.1 Differentially expressed genes (q < 0.01)
Significantly differentially expressed genes are listed (by cosmid name as they appear on the microarray .gal file) along with the directionality of the differential expression, the significance for statistical tests (see methods) and the observed fold change. E = *E. coli*, M = *M. luteus*, P = *Pseudomonas sp*., B = *B. megaterium*.

Gene Name	p-value	TREAT1	Direction	TREAT2	Fold Change
6R55.1a	3.08038E-05	E	<	M	1.2248
B0035.13	1.4915E-07	B	>	E	2.0202
B0035.13	4.7722E-07	B	>	P	1.8395
B0213.15	0.000032	B	<	E	1.4622
B0213.3	8.87301E-06	E	>	M	1.2145
B0213.3	2.88538E-05	E	>	P	1.1204
B0222.7	5.27293E-06	B	<	E	2.039
B0222.7	4.82184E-08	E	>	P	2.4527
B0546.1	2.24887E-05	E	>	P	1.1331
C01B10.5a	1.22042E-05	B	<	E	1.8983
C01B10.5a	6.33232E-06	E	>	P	1.931
C01B12.1	1.66192E-05	E	>	P	1.4134
C01G6.7	0.000012927	B	<	M	1.2716
C02E7.6	0.000027024	B	<	E	2.3747
C02E7.7	6.17277E-06	B	<	E	1.8661
C05E4.9a	8.84188E-09	B	<	E	1.8005
C05E4.9a	1.11267E-05	B	<	M	1.1629
C05E4.9a	3.12848E-09	B	<	P	1.8091
C05E4.9b	1.33851E-06	B	<	E	1.5402
C08F11.12	2.87692E-09	B	<	E	2.6129

C08F11.12	6.19841E-10	B	<	P	2.7508
C08F11.12	2.56197E-05	E	>	M	1.5439
C08F11.12	5.81398E-06	M	<	P	1.6818
C09H10.2	3.01829E-05	B	>	P	1.487
C10G11.5b	0.000037191	B	<	P	0.9498
C16H3.2	4.63568E-07	B	<	M	1.4253
C16H3.2	1.85109E-08	M	>	P	1.645
C17F4.7	7.37437E-10	B	>	P	3.2222
C17F4.7	2.30523E-05	E	>	P	1.8361
C17F4.7	1.52652E-09	M	>	P	3.0794
C23G10.11	1.29937E-05	B	<	E	2.4646
C23H5.8a	0.000014679	E	<	M	1.1286
C23H5.8a	1.16067E-05	M	>	P	1.2004
C24B9.3	1.60089E-05	M	>	P	1.1558
C24B9.9	5.6088E-06	M	<	P	1.6141
C24F3.6	6.14981E-06	B	>	P	2.513
C24F3.6	8.42877E-08	E	>	P	3.4593
C24F3.6	1.1747E-06	M	>	P	2.8907
C31E10.7	1.73085E-05	M	>	P	1.3259
C32F10.4	2.29986E-05	B	<	E	1.3535
C32F10.4	3.82827E-06	B	<	P	1.47
C34F6.2	1.13341E-07	E	>	P	3.6832
C34F6.2	1.38534E-06	M	>	P	3.1423
C34H4.4	8.68619E-06	B	<	E	1.7735
C35B1.4	7.60558E-06	B	<	P	1.2422
C37A2.7	8.59153E-06	B	>	P	1.9546
C37A2.7	4.55716E-07	E	>	P	2.333
C37A2.7	2.64404E-06	M	>	P	2.079
C41G11.1	7.64868E-06	B	<	P	1.2331

C44E4.6	1.56017E-05	M	>	P	1.2043
C45B11.3	1.87112E-07	E	<	M	2.2555
C45B11.3	2.98875E-06	M	>	P	1.8708
C45G7.3	7.45056E-06	B	<	M	1.7127
C45G7.3	1.0618E-09	B	>	P	2.5388
C45G7.3	2.53538E-07	E	<	M	1.9848
C45G7.3	3.99665E-08	E	>	P	2.2666
C45G7.3	1.84068E-19	M	>	P	4.2514
C47B2.3b	2.57345E-05	B	>	P	1.5046
C47B2.3b	4.01878E-06	E	>	P	1.6954
C47B2.3b	2.26951E-07	M	>	P	1.985
C50F7.5	4.90355E-06	B	>	P	2.1242
C53B4.5	1.58604E-07	E	>	P	4.4321
C53B4.5	9.36824E-06	M	>	P	3.5018
C53B7.3	1.75644E-05	B	<	P	1.0355
C53B7.3	2.2331E-06	E	>	M	1.1841
C53B7.3	8.17571E-08	M	<	P	1.3916
C54C8.9	2.65395E-06	B	<	E	2.7701
C54C8.9	3.39124E-09	B	<	P	4.0576
C54C8.9	4.12108E-06	E	>	M	2.4913
C54C8.9	6.24023E-09	M	<	P	3.7788
C54F6.5	2.94043E-05	B	<	E	2.0676
C54G6.5	8.003E-06	B	<	E	1.2038
C54G6.5	2.8474E-09	B	<	M	1.7954
C54G6.5	4.97745E-08	E	>	P	1.5625
C54G6.5	3.44278E-11	M	>	P	2.1542
C55B7.4a	7.96486E-07	B	<	M	2.7392
C55B7.4a	1.47737E-07	M	>	P	2.7293
D1065.5	1.36279E-05	E	>	P	2.0424

D1086.6	3.85217E-07	B	<	E	2.2592
E04A4.7	3.55144E-05	B	>	P	1.385
E04A4.7	3.74513E-07	E	>	P	1.8714
E04A4.7	3.25256E-06	M	>	P	1.627
E04F6.3	1.02328E-09	B	<	M	2.177
E04F6.3	6.77984E-10	E	<	M	2.1986
E04F6.3	5.09765E-13	M	>	P	3.0424
F01D5.9	0.000029371	B	<	M	1.9947
F01D5.9	2.80477E-05	E	<	M	2.0612
F01D5.9	3.47514E-06	M	>	P	2.4526
F02A9.2	2.90937E-05	B	<	E	1.0994
F07C3.9	1.47792E-05	B	<	E	1.5286
F10D2.9	1.61011E-12	B	<	M	3.1712
F10D2.9	1.72746E-10	E	<	M	2.5928
F10D2.9	7.11004E-09	M	>	P	2.175
F11E6.3	3.56711E-05	B	<	M	1.1229
F11E6.3	6.265E-07	M	>	P	1.456
F11E6.5	8.02231E-06	E	<	M	1.5273
F11G11.11	7.41621E-06	E	>	P	3.402
F15E6.8	2.84592E-07	B	<	E	1.8967
F15H10.2	3.66429E-06	E	>	P	1.9353
F15H10.2	2.10634E-05	M	>	P	1.7242
F17E9.11	1.63782E-08	B	>	P	2.8973
F17E9.11	1.09038E-06	E	<	M	2.2193
F17E9.11	3.25763E-05	E	>	P	1.7041
F17E9.11	2.22574E-11	M	>	P	3.9234
F17E9.4	1.81663E-05	E	>	P	1.3639
F18A1.5	2.79535E-05	E	>	P	1.2994
F18H3.3b	3.75396E-05	E	>	P	1.7497

F19C7.1	8.48546E-06	M	>	P	1.8644
F21F8.4	1.45204E-06	B	>	E	3.4414
F21F8.4	1.9429E-07	B	>	P	3.6504
F21F8.7	8.217E-06	B	>	P	2.1087
F22A3.6	0.000017171	B	>	E	2.0601
F22A3.6	1.10837E-11	B	>	P	4.1502
F22A3.6	1.39203E-05	E	>	P	2.0901
F22A3.6	6.72874E-10	M	>	P	3.4329
F22B5.3	2.8239E-07	B	<	E	1.8203
F23A7.4	4.72901E-06	M	>	P	2.0354
F25B5.4c	2.37308E-05	B	>	P	1.6659
F25B5.4c	1.35903E-06	E	>	P	1.935
F25B5.4c	1.76388E-06	M	>	P	1.8724
F25E2.3	3.88055E-06	M	>	P	2.884
F25H2.10	4.11281E-06	B	>	P	2.4612
F25H2.10	4.92832E-09	E	>	P	3.5945
F25H2.10	3.85612E-08	M	>	P	3.2043
F25H5.8	2.71735E-05	B	<	E	1.6449
F26F12.1	1.42012E-06	E	>	P	3.6862
F26F12.1	3.68093E-05	M	>	P	2.8884
F27C8.4	3.7321E-06	B	>	E	1.4914
F27C8.4	2.89629E-08	B	>	P	1.8922
F27C8.4	1.41551E-05	M	>	P	1.3104
F28A12.4	2.11811E-05	B	<	M	1.6363
F28G4.1	1.00039E-08	B	<	M	2.1823
F28G4.1	3.0519E-07	E	<	M	1.7006
F28G4.1	1.54695E-06	M	>	P	1.6134
F28H7.3	1.46581E-06	B	>	E	1.9777
F28H7.3	2.05801E-08	B	>	P	2.4188

F28H7.3	4.42952E-06	M	>	P	1.7752
F31E3.3	2.03025E-06	B	>	P	1.3734
F31E3.6	1.92946E-05	B	>	P	1.393
F32A5.5a	3.70618E-08	B	<	E	2.2655
F32A5.5a	2.26737E-05	E	>	M	1.5391
F32D1.5	1.79169E-06	B	>	P	1.5952
F35C8.5	1.36419E-05	B	<	E	1.0086
F35C8.5	9.89934E-06	B	<	M	1.0018
F38E11.2	1.25114E-06	B	<	E	2.4176
F41F3.3	3.20758E-07	B	<	E	2.9018
F41F3.4	1.03149E-05	E	>	P	2.7586
F41F3.4	3.58197E-06	M	>	P	3.1077
F41H10.7	2.15045E-05	M	>	P	1.7459
F42F12.6	9.33314E-06	E	>	P	1.6371
F44C4.3	9.14268E-06	B	>	P	3.3209
F46F11.2	2.55549E-06	B	>	P	1.7616
F49C12.11	3.93198E-05	E	>	P	0.9248
F49C12.7	3.33823E-06	B	>	P	2.1042
F52B11.4	3.28097E-05	E	>	P	1.4689
F52C6.12	1.05586E-05	M	>	P	1.8008
F52H3.7a	2.84109E-05	E	>	P	1.8165
F53H4.2	6.93821E-06	B	<	P	1.5547
F54D7.2	8.9828E-06	B	>	P	1.995
F54D7.2	1.30535E-05	E	<	M	1.8521
F54D7.2	1.99685E-06	M	>	P	2.1732
F54D8.1	1.80541E-07	B	<	E	1.6517
F54D8.1	1.87964E-05	B	<	M	1.1972
F54H12.6	8.29483E-06	B	>	P	2.2273
F54H12.6	6.276E-08	E	>	P	2.9493

F54H12.6	1.33339E-06	M	>	P	2.4621
F55B11.2	1.94866E-06	B	<	E	2.218
F55F3.3	7.53427E-06	M	>	P	2.0833
F56C9.7	1.98479E-06	M	>	P	2.1419
F56H9.2	3.13844E-05	M	<	P	1.7379
F57B1.3	1.92298E-06	E	>	P	1.9242
F57B1.3	5.8441E-07	M	>	P	1.8483
F57F5.1	3.38185E-11	B	>	P	4.5718
F57F5.1	9.1736E-07	E	>	P	2.8595
F57F5.1	5.44778E-11	M	>	P	4.4288
F58B3.1	1.18515E-05	B	>	E	1.814
F58B3.1	7.3E-23	B	>	P	4.9792
F58B3.1	1.16876E-05	E	<	M	1.8029
F58B3.1	1.31137E-12	E	>	P	3.1652
F58B3.1	1.44E-22	M	>	P	4.9681
F58B3.2	7.53715E-06	B	>	P	2.8135
F58F9.7	3.29882E-09	B	<	M	1.9388
F58F9.7	3.35331E-11	E	<	M	2.4348
F58F9.7	2.92241E-10	M	>	P	2.1192
F59B1.2	5.33388E-07	B	>	P	2.5473
F59B1.2	2.39643E-05	E	>	P	1.9859
F59B1.2	9.46824E-06	M	>	P	2.0601
H02I12.1	6.8626E-06	B	>	P	1.8553
H22K11.1	1.39324E-05	B	>	P	2.0938
H27M09.4	5.99753E-06	B	<	E	1.3881
H27M09.4	2.98643E-05	B	<	M	1.2084
JC8.8	1.28166E-05	B	<	E	1.881
JC8.8	2.74022E-05	E	>	M	1.6699
K01D12.12	4.03126E-05	M	>	P	1.0743

K02F3.4	1.48863E-06	M	>	P	1.537
K03H1.4	3.87349E-05	B	<	E	1.4966
K03H1.4	1.34071E-07	B	<	P	2.1367
K03H1.4	2.289E-07	M	<	P	1.9122
K03H6.2	1.02566E-05	E	>	P	1.4487
K07A1.6	1.64762E-09	B	<	E	1.6759
K07A1.6	1.68501E-08	B	<	P	1.4593
K07A1.6	1.18436E-06	E	>	M	1.1634
K07A1.6	2.45331E-05	M	<	P	0.9468
K07F5.11	2.09604E-08	E	>	P	2.2652
K07F5.9	1.91212E-05	E	>	P	2.3769
K07F5.9	1.15139E-05	M	>	P	2.498
K07H8.6	0.00002566	M	>	P	1.5734
K08B4.6	1.22034E-11	B	<	E	3.6191
K08B4.6	1.42404E-09	B	<	P	2.8718
K08B4.6	4.06119E-09	E	>	M	2.7359
K08B4.6	1.43991E-06	M	<	P	1.9885
K09F5.2	9.00621E-06	E	<	M	1.9507
K09F5.2	1.58113E-06	M	>	P	2.0071
K10D2.4	2.38291E-05	B	>	P	0.9228
K11G9.6	2.01835E-13	B	<	E	5.1396
K11G9.6	7.89859E-11	B	<	M	3.9076
K11G9.6	1.88151E-10	B	<	P	3.7164
K12G11.3	2.14155E-10	B	<	E	3.4823
K12G11.3	8.75986E-09	B	<	P	2.8544
K12G11.3	3.91442E-11	E	>	M	3.6534
K12G11.3	1.99652E-09	M	<	P	3.0254
K12H4.7a	6.21629E-07	B	>	P	2.7359
K12H4.7a	1.94E-06	M	>	P	2.4811

LLC1.2	5.11212E-06	M	>	P	1.8974
LLC1.3	3.49362E-05	E	>	P	1.2638
M03A8.1	4.13583E-07	B	<	M	1.5478
M03A8.1	4.68482E-09	E	<	M	1.9573
M03A8.1	1.19719E-09	M	>	P	2.0024
M03F4.2b	3.34113E-05	B	>	P	1.8426
M03F4.2b	1.73296E-07	E	>	P	2.5454
M03F4.2b	5.31191E-07	M	>	P	2.363
M60.4	1.95424E-08	B	<	E	1.9689
M60.4	1.74124E-06	E	>	M	1.5051
M88.1	3.36336E-05	M	>	P	1.9776
R06B10.3	1.71013E-05	B	>	P	1.4529
R07B1.10	4.00579E-05	B	>	P	1.9106
R07B1.10	1.30239E-07	M	>	P	2.6821
R09A8.4	1.92378E-05	E	>	P	2.0856
R09B3.3	9.55595E-06	B	>	P	1.241
R09H10.5	7.7355E-08	M	>	P	1.9833
SR3 - rbc	8.50201E-06	E	<	P	1.0081
SR3 - rbc	1.89322E-05	M	<	P	1.0356
T01B7.7	1.02903E-05	E	>	P	2.6893
T01C3.4	1.55691E-10	B	>	E	2.5767
T01C3.4	1.76048E-12	B	>	M	2.9355
T01C3.4	1.63889E-12	B	>	P	2.9475
T03E6.7	9.41257E-07	M	>	P	1.9846
T05A1.2	3.54356E-06	E	>	P	3.2731
T05A1.2	4.09989E-06	M	>	P	3.2949
T08A9.12	1.46833E-05	B	>	P	2.6997
T08A9.12	7.67887E-06	M	>	P	2.839
T08A9.8	2.87919E-08	B	>	P	3.4421

T08A9.8	7.57325E-08	M	>	P	3.3086
T08G5.10	1.76978E-07	B	<	E	3.0595
T08G5.10	0.000021323	E	>	P	2.1706
T10H4.12	1.95254E-06	B	>	P	2.1573
T12D8.5	3.28574E-05	B	>	M	1.7442
T12D8.5	2.89878E-06	B	>	P	2.0488
T14F9.3	0.000027874	B	<	P	1.9736
T16G1.6	6.61681E-06	B	<	P	2.0902
T18H9.2	1.60984E-07	B	>	P	2.588
T18H9.2	9.83667E-07	M	>	P	2.2077
T21C9.9	0.000029721	E	>	P	2.4224
T21C9.9	6.64965E-06	M	>	P	2.8165
T21H3.1	5.02429E-06	B	>	P	1.9977
T21H3.1	4.12164E-06	M	>	P	1.9977
T22F3.4	4.76394E-06	B	>	P	1.5562
T22G5.2	1.57994E-05	M	>	P	1.5109
T23G11.3	7.7087E-06	B	>	P	1.2584
T24B8.5	3.63573E-05	B	>	E	1.922
T28C6.6	1.07506E-05	B	<	E	1.3176
VW02B12L.1	1.89182E-08	B	>	P	2.2873
VW02B12L.1	2.52207E-05	M	>	P	1.4847
W02A2.1	2.65662E-06	M	>	P	1.8093
W02D3.5	5.08219E-06	E	<	M	1.4333
W02D3.5	1.09349E-05	M	>	P	1.3473
W02D3.7	5.24388E-06	B	<	M	1.8918
W02D3.7	1.98917E-08	E	<	M	2.7802
W02D3.7	4.07947E-07	M	>	P	2.172
W02D9.6	2.40246E-05	B	<	E	1.8385
W02D9.6	5.24631E-06	E	>	M	2.014

W03F9.4	2.76898E-06	E	<	M	2.5063
W03F9.4	6.43635E-08	M	>	P	3.2124
W04E12.8	6.05511E-06	M	>	P	3.8919
W05B2.6	5.71673E-06	E	>	P	1.9502
W05B2.6	1.6693E-06	M	>	P	2.0635
W07B8.5	7.94686E-09	B	>	P	2.6449
W07B8.5	4.45555E-08	M	>	P	2.2962
W09C5.6a	1.49229E-05	B	>	P	2.0219
W09C5.6a	1.82019E-07	E	>	P	2.6329
W09C5.6a	9.87474E-07	M	>	P	2.3648
Y105E8A.4	1.74615E-05	B	<	M	1.2615
Y105E8A.4	9.96637E-06	E	<	M	1.2916
Y105E8A.4	4.53465E-07	M	>	P	1.5604
Y106G6D.4	8.30118E-06	B	<	P	0.9897
Y106G6D.4	8.24879E-07	M	<	P	1.0903
Y22F5A.4	2.41172E-05	E	<	M	1.7742
Y24D9A.4b	1.76697E-05	E	>	P	1.6648
Y24D9A.4c	2.75563E-05	E	>	P	1.7828
Y32F6A.5	1.33317E-08	B	>	E	1.4445
Y32F6A.5	5.10043E-07	B	>	P	1.2088
Y34B4A.6	2.34042E-05	M	>	P	1.6291
Y37E3.7	5.98898E-06	E	>	P	1.683
Y37E3.7	8.56841E-06	M	>	P	1.6243
Y38F2AR.9	1.30625E-05	M	>	P	1.6923
Y38H6C.1	9.96499E-09	B	>	P	3.2242
Y38H6C.1	5.88609E-07	E	<	M	2.4892
Y38H6C.1	1.94833E-11	M	>	P	4.0911
Y38H6C.17	5.02231E-06	B	>	P	2.7552
Y39A1C.3	2.52872E-06	M	>	P	1.6514

Y39B6A.1	2.05764E-06	B	>	E	2.5671
Y39B6A.1	4.15255E-10	B	>	P	3.8897
Y39B6A.1	2.21627E-07	E	<	M	2.9684
Y39B6A.1	8.81016E-11	M	>	P	4.2909
Y45F10C.2	3.82109E-08	B	<	E	2.4436
Y45F10C.2	4.7001E-09	B	<	P	2.6378
Y45F10C.2	1.27588E-05	E	>	M	1.7045
Y45F10C.2	1.90286E-06	M	<	P	1.8986
Y45F10C.4	2.69304E-05	B	<	E	1.5035
Y48B6A.2	2.1789E-07	E	>	P	2.6026
Y48B6A.2	3.2668E-06	M	>	P	2.1997
Y48G8AL.8a	1.39751E-05	B	>	P	2.0448
Y48G8AL.8a	2.97312E-07	E	>	P	2.5814
Y48G8AL.8a	6.52345E-07	M	>	P	2.4388
Y49E10.18	3.41058E-06	B	>	E	2.0545
Y49E10.18	2.66129E-05	E	<	M	1.9154
Y4C6B.6	9.99332E-07	B	>	P	1.8874
Y4C6B.6	4.03273E-07	M	>	P	2.0835
Y53H1B.2	2.66224E-05	B	<	E	1.4121
Y53H1B.2	9.37163E-06	B	<	P	1.4185
Y54G11A.5b	2.51977E-05	M	>	P	1.586
Y54G11A.6	4.43571E-07	M	>	P	1.5413
Y54G2A.6	1.03737E-05	E	>	P	2.2535
Y55B1AR.1	1.14935E-06	M	>	P	1.9522
Y56A3A.20	3.36776E-05	B	>	P	1.4001
Y57A10C.6	1.24803E-06	B	<	M	2.1269
Y57A10C.6	8.26063E-08	E	<	M	2.6368
Y57A10C.6	3.38332E-09	M	>	P	3.0985
Y60A3A.18	3.98488E-05	B	<	M	0.9073

Y60A3A.18	3.45787E-06	M	>	P	1.0097
Y62E10A.1	7.58377E-06	B	>	P	1.9532
Y62E10A.1	9.33511E-08	E	>	P	2.5189
Y62E10A.1	2.50349E-06	M	>	P	2.0681
Y65B4BR.1	3.09202E-06	B	>	E	1.8778
Y65B4BR.1	2.41496E-05	B	>	P	1.5826
Y65B4BR.1	1.13765E-06	E	<	M	2.0196
Y65B4BR.1	1.06633E-05	M	>	P	1.7245
Y69A2AR.18a	0.000021051	E	>	P	1.7747
Y71F9AL.13a	1.99961E-05	B	>	P	2.3026
Y71F9AL.13a	6.57617E-08	E	>	P	3.2273
Y71F9AL.13a	2.56677E-07	M	>	P	2.965
Y73B3A.18a	4.51272E-07	B	>	P	1.953
Y73B3A.18a	7.74583E-13	E	>	P	3.1464
Y73B3A.18a	1.8153E-11	M	>	P	2.7695
Y75B8A.4	4.41872E-07	E	<	M	2.7974
Y82E9BR.3	7.67592E-06	E	>	P	2.1979
Y82E9BR.3	3.95147E-06	M	>	P	2.1667
ZC395.5	1.01162E-07	B	<	E	2.3565
ZC395.5	1.77544E-05	B	<	P	1.6828
ZC395.5	7.09063E-06	E	>	M	1.7846
ZC64.2	2.70603E-05	B	<	E	1.3184
ZC64.2	0.000017647	E	>	P	1.2212
ZK1193.1	1.25623E-05	E	>	P	2.6431
ZK1193.1	3.44384E-05	M	>	P	2.3377
ZK180.5c	4.12431E-06	E	>	P	2.2022
ZK622.3a	2.78684E-05	B	>	M	1.0509
ZK813.2	7.78914E-08	B	<	E	2.1254
ZK813.2	1.64016E-07	B	<	P	1.9781

ZK813.2	5.86578E-06	E	>	M	1.5938
ZK813.2	0.000018847	M	<	P	1.4464
ZK892.2	2.89796E-05	E	<	M	1.8711
ZK892.2	1.29038E-07	M	>	P	2.5517

Appendix B - Responding bacteria taxa

Table 9.1 List of bacteria taxa responding to the addition of nitrogen fertilizer

Taxa have been identified by BLAST to the Ribosome Database Project (RDP) and given Similarity scores (S). Fold change in response to disturbance is given.

Contig	Fold change in Disturbed	S	Phyla	Class	Order	Family	Genus
4569	3.34037	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
107	3.23484	0.97	Nitrospira	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira
1413	2.97236	0.79	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
1867	2.93851	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
4406	2.54963	0.94	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
3623	2.51578	0.95	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes
9230	2.48472	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
4025	2.42066	0.94	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
2024	2.25175	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
5598	2.09293	0.87	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Dokdonella
3289	2.02415	0.76	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp10
8224	1.99126	0.94	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiales	
8306	1.89055	0.95	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiales	
4991	1.87601	0.82	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae	
3434	1.77398	0.88	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
13027	1.75277	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
8042	1.67662	0.97	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium
12964	1.66142	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
9332	1.64937	0.94	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
9736	1.63438	0.94	Proteobacteria	Alphaproteobacteria	Alphaproteobacteria		
1108	1.61437	0.9	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriales	
1103	1.59445	0.9	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
4919	1.59148	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
7194	1.58286	0.96	Proteobacteria	Alphaproteobacteria	Alphaproteobacteria		
4951	1.56972	0.94	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidisphaera
5502	1.56834	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp7
1974	1.56645	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
11279	1.54305	0.98	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus c
5335	1.52552	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
9320	1.52175	0.88	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
12326	1.49685	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3

3876	1.48525	0.84	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
3108	1.4654	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp2
9417	1.46525	0.87	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	Acidisphaera
6534	1.44828	0.95	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria		
7470	1.44587	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
4726	1.44462	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
3483	1.43928	0.87	Proteobacteria	Betaproteobacteria	Betaproteobacteria		
8440	1.36658	0.89	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiales	
6652	1.34772	0.97	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
5582	1.34611	0.92	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
4160	1.33073	0.86	Proteobacteria	Betaproteobacteria	Betaproteobacteria		
13026	1.27422	0.88	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
11659	1.26622	0.91	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
8408	1.2435	0.95	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
2122	1.23207	0.72	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
10870	1.21551	0.9	Bacterium				
11186	1.20239	0.87	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
1042	1.20079	0.89	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriales	
2118	1.19188	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
9878	1.17239	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
4411	1.16746	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp2
2809	-1.13401	0.72	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
11553	-1.14196	0.94	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Opitutaceae	Opitutus
6049	-1.14947	0.88	Proteobacteria	Betaproteobacteria	Betaproteobacteria		
9787	-1.15374	0.93	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiales	
11518	-1.16392	0.94	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	Verrucomicrobium
568	-1.16402	0.85	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
486	-1.16532	0.88	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Crenotrichaceae	Terrimonas
3196	-1.16842	0.85	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Virgisporangium
10113	-1.17306	0.89	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae	
6563	-1.17373	0.8	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
7217	-1.17604	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
7584	-1.17924	0.87	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Opitutaceae	Opitutus
6560	-1.18304	0.96	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylocystaceae	Methylosinus
5188	-1.18821	0.96	Chloroflexi	Chloroflexi			
11246	-1.22674	0.67	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiales	
11568	-1.23754	0.89	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
5717	-1.2481	0.67	Firmicutes	Clostridia	Clostridiales	"Ruminococcaceae"	
8642	-1.24923	0.83	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
3094	-1.24984	0.78	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
6141	-1.25425	0.8	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
936	-1.2577	0.98	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
9040	-1.28089	0.94	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Actinoalloteichus
45	-1.29104	0.81	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
3744	-1.30802	0.89	Proteobacteria	Betaproteobacteria	Betaproteobacteria		
5220	-1.32446	0.96	Chloroflexi	Anaerolineae	Caldilineales	Caldilineacea	
1300	-1.3303	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
3751	-1.3459	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
10483	-1.351	0.81	Proteobacteria	Deltaproteobacteria	Myxococcales	Myxococcales	

542	-1.37755	0.8	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
1332	-1.40474	0.88	Proteobacteria	Deltaproteobacteria	Deltaproteobacteria		
8776	-1.4052	0.88	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp7
6509	-1.40957	0.95	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria		
2665	-1.48798	0.95	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacterineae	
6454	-1.52139	0.74	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
7561	-1.57249	0.76	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
3859	-1.60847	0.89	Proteobacteria	Betaproteobacteria	Betaproteobacteria		
9849	-1.60872	0.94	Proteobacteria	Alphaproteobacteria	Rhodospirillales		
1670	-1.62096	0.96	Chloroflexi	Chloroflexi			
3883	-1.63887	0.95	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
2454	-1.64915	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp5
1112	-1.65066	0.92	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
2161	-1.6684	0.91	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
1097	-1.67118	0.95	Chloroflexi	Anaerolineae	Caldilineales	Caldilineacea	Caldilinea
3898	-1.89685	0.92	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
528	-1.9206	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
2239	-2.10443	0.83	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
3629	-2.24677	0.84	Proteobacteria	Gammaproteobacteria	Gammaproteobacteria		
1045	-2.41574	0.87	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriales	
215	-2.45155	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6

Table 9.2 List of bacteria taxa responding to tillage

Taxa have been identified by BLAST to the Ribosome Database Project (RDP) and given Similarity scores (S). Fold change in response to disturbance is given.

Contig	Fold change in Distured	S	Phyla	Class	Order	Family	Genus
5923	3.21669	0.72	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
11669	2.69609	0.96	Proteobacteria	Deltaproteobacteria	Deltaproteobacteria		
2665	2.5562	0.95	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacterineae	
7545	1.96967	0.95	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	
2927	1.90531	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
10511	1.88362	0.94	Proteobacteria	Deltaproteobacteria	Deltaproteobacteria		
5502	1.82304	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp7
11659	1.78106	0.91	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
8623	-2.16557	0.96	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
3990	-2.23622	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
12516	-2.28745	0.88	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
9851	-2.58509	0.85	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
12341	-2.70342	0.96	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
10420	-2.70966	0.96	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
6976	-2.74664	0.9	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
11543	-3.56028	0.9	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	

Table 9.3 List of bacteria taxa responding to conversion from prairie to agriculture

Taxa have been identified by BLAST to the Ribosome Database Project (RDP) and given Similarity scores (S). Fold change in response to disturbance is given.

Contig	Fold change in Distured	S	Phyla	Class	Order	Family	Genus
1312	1.50671	0.94	Bacteroidetes	Sphingobacteria	Sphingobacteriales		
10902	1.49409	0.9	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	
2158	1.38842	0.81	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
1008	1.28623	0.89	Bacteroidetes	Sphingobacteria	Sphingobacteriales		
208	1.17254	0.96	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
8337	1.07928	0.87	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
1988	1.03536	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
2182	0.98282	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
8091	0.98244	0.75	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
2275	0.96015	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
262	0.93592	0.98	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
5512	0.93269	0.94	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	
5385	0.90125	0.84	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
206	0.89895	0.83	Bacteroidetes	Sphingobacteria	Sphingobacteriales		
8405	0.89569	0.77	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
9320	0.8885	0.88	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
2357	0.888	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp5
9174	0.84625	0.95	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	
7368	0.84313	0.88	Bacteria				
1815	0.83625	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp5
8158	0.82962	0.9	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	
12068	0.82516	0.94	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
1411	0.81604	0.83	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
4726	0.81187	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
5878	0.8073	0.88	Bacteria				
11914	0.79191	0.89	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
9851	0.78982	0.85	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
4549	0.77005	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
4678	0.76798	0.89	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Hydrocarboniphaga
790	0.76729	0.95	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Crenotrichaceae	
2939	0.762	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
4493	0.75484	0.95	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
320	0.75107	0.82	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae	Niastella
9138	0.73975	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
1054	0.73176	0.89	Bacteroidetes	Sphingobacteria	Sphingobacteriales		
246	0.70687	0.79	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp7
9230	0.69216	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
2272	0.69081	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6

8919	0.68705	0.89	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Actinoplanes
10533	0.68362	0.97	Planctomycetes	Planctomycetacia	Planctomycetales	Planctomycetaceae	
768	0.67572	0.95	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Sphingobacteriales	
1783	0.6721	0.69	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
393	0.66907	0.95	Bacteroidetes	Sphingobacteria	Sphingobacteriales		
2831	0.66836	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp22
576	0.66833	0.81	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
11228	0.66129	0.61	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
12047	0.65675	0.89	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
2266	0.65284	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
6429	0.64903	0.37	Proteobacteria	Alphaproteobacteria			
7753	0.6458	0.96	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
2140	0.6425	0.85	Bacteria				
2454	0.64223	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp5
8437	0.63467	0.73	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
6580	0.63422	0.95	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
6919	0.6327	0.86	Bacteria				
2938	0.63132	0.86	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
3386	0.61993	0.89	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	
5612	0.61931	0.88	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	Conexibacter
3300	0.61281	0.92	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Crenotrichaceae	Terrimonas
11766	0.61129	0.81	Proteobacteria	Deltaproteobacteria	Myxococcales	Polyangiaceae	Byssovorax
11431	0.60893	0.88	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
2552	0.60477	0.87	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Pseudonocardia
1382	0.60012	0.7	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Crenotrichaceae	Chitinophaga
5246	0.58846	0.79	Proteobacteria	Deltaproteobacteria			
2988	0.58784	0.83	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
11183	0.58081	0.68	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
735	0.58054	0.94	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp5
6618	0.57777	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
4327	0.57208	0.67	Chloroflexi	Anaerolineae	Caldilineales	Caldilineacea	
3840	0.57192	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
9010	0.57068	0.81	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
3314	0.57059	0.8	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
3353	0.56541	0.96	Bacteroidetes	Sphingobacteria	Sphingobacteriales		
1867	0.56442	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
10617	0.56004	0.94	Bacteria				
173	0.55869	0.9	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
12934	0.55837	0.94	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
1133	0.55776	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
2286	0.55706	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
11817	0.55119	0.86	Proteobacteria	Deltaproteobacteria	Myxococcales	Polyangiaceae	Byssovorax
12328	0.54968	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
7612	0.5482	0.68	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
4406	0.54805	0.94	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
1351	0.54601	0.84	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp13
2598	0.5413	0.95	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae	Niastella
4837	0.53664	0.95	Bacteria				
2194	0.53652	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1

1720	0.53418	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp5
2428	0.53314	0.84	Proteobacteria	Alphaproteobacteria			
7470	0.53097	0.91	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
10335	0.52971	0.88	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
11968	0.52733	0.83	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
2304	0.52345	0.83	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
4356	0.51913	0.95	Chloroflexi	Anaerolineae			
5709	0.51895	0.76	Bacteria				
3223	0.51406	0.92	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
7082	0.51043	0.87	Proteobacteria	Betaproteobacteria			
4216	0.50906	0.81	Proteobacteria	Gammaproteobacteria			
4569	0.50874	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
6879	-0.55494	0.96	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
2900	-0.5555	0.95	Proteobacteria	Alphaproteobacteria	Rhizobiales		
8485	-0.55607	0.9	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
11266	-0.55953	0.72	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
9202	-0.56028	0.95	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium
8150	-0.56255	0.91	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacterineae	
3007	-0.56455	0.84	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
7055	-0.56704	0.79	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
12990	-0.56842	0.95	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
12307	-0.56998	0.93	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
6594	-0.57379	0.88	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
3115	-0.57388	0.96	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
1878	-0.57822	0.91	Proteobacteria	Alphaproteobacteria			
7338	-0.58041	0.94	Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	Amycolatopsis
4869	-0.58168	0.95	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	
2141	-0.5823	0.72	Bacteria				
7456	-0.58289	0.92	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	
1901	-0.58407	0.95	Chloroflexi	Chloroflexi	Chloroflexales		
8109	-0.5841	0.95	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
1524	-0.58432	0.82	Proteobacteria	Betaproteobacteria			
3464	-0.58528	0.82	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
6563	-0.58556	0.8	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
1827	-0.58645	0.82	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
3847	-0.58677	0.89	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
2413	-0.58807	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp5
6885	-0.59674	0.89	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
2128	-0.59817	0.81	Actinobacteria	Actinobacteria	Actinomycetales	Thermomonosporaceae	Actinomadura
4725	-0.59982	0.94	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
10511	-0.60204	0.94	Proteobacteria	Deltaproteobacteria			
8539	-0.60232	0.78	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
7694	-0.60602	0.81	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	
6135	-0.60631	0.8	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae	
7132	-0.6073	0.94	Proteobacteria	Betaproteobacteria			
4117	-0.60853	0.95	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
4757	-0.61474	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
6468	-0.61649	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp3
3528	-0.61911	0.95	Proteobacteria	Alphaproteobacteria			

12399	-0.62061	0.86	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Knoellia
7545	-0.62172	0.95	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	
12522	-0.6259	0.94	Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Knoellia
6565	-0.62593	0.88	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacterineae	
9930	-0.62681	0.95	Chloroflexi	Anaerolineae			
9676	-0.63162	0.94	Proteobacteria	Alphaproteobacteria	Rhodospirillales		
4263	-0.639	0.84	Bacteroidetes	Sphingobacteria	Sphingobacteriales	Flexibacteraceae	
3518	-0.64196	0.88	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
1987	-0.65487	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp1
2538	-0.66488	0.88	Proteobacteria	Gammaproteobacteria			
6532	-0.67041	0.88	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
3207	-0.67682	0.86	Bacteria				
1548	-0.67748	0.94	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Phenylobacterium
2756	-0.68424	0.95	Bacteria				
2545	-0.68449	0.93	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Rhodoplanes
7217	-0.69301	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
5212	-0.7001	0.94	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
10807	-0.70339	0.96	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
3037	-0.70407	0.95	Actinobacteria	Actinobacteria			
12923	-0.7168	0.86	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
4661	-0.72272	0.89	Bacteria				
5625	-0.73085	0.94	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bradyrhizobium
2665	-0.73879	0.95	Actinobacteria	Actinobacteria	Actinomycetales	Propionibacterineae	
6107	-0.74199	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
5982	-0.74454	0.96	Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas
7061	-0.74619	0.93	Proteobacteria	Alphaproteobacteria			
1711	-0.74641	0.91	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
5467	-0.74968	0.95	Bacteria				
8623	-0.77051	0.96	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
7466	-0.77106	0.86	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
6490	-0.77129	0.87	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
12341	-0.77686	0.96	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
3786	-0.77947	0.88	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Subdivision 3	
747	-0.78372	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
6470	-0.78588	0.95	Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae	
779	-0.79051	0.92	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
1996	-0.80307	0.96	Bacteria				
5874	-0.81385	0.95	Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas
1311	-0.81406	0.89	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
940	-0.81986	0.96	Chloroflexi	Chloroflexi	Chloroflexales	Chloroflexaceae	Roseiflexus
12892	-0.83764	1	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
7415	-0.84798	0.83	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
5175	-0.84883	0.95	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	
4626	-0.86399	0.92	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
7346	-0.86753	0.87	Actinobacteria	Actinomycetales			
425	-0.87192	0.89	Bacteria				
11130	-0.88722	0.93	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
967	-0.88867	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
1171	-0.91131	0.95	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6

6450	-0.91249	0.94	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
2171	-0.92535	0.95	Actinobacteria	Actinobacteria			
568	-0.93297	0.85	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp6
12147	-0.95543	0.94	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
8945	-0.96541	0.93	Proteobacteria	Betaproteobacteria			
9120	-0.98722	0.81	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
8711	-0.99744	0.77	Acidobacteria	Acidobacteria	Acidobacteriales	Acidobacteriaceae	Gp4
5204	-1.0012	0.88	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiales	
12388	-1.07636	0.95	Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Xiphinematobacteriaceae	
1383	-1.08533	0.76	Nitrospira	Nitrospira	Nitrospirales	Nitrospiraceae	Nitrospira
8867	-1.19448	0.86	Proteobacteria	Alphaproteobacteria	Rhodospirillales		
3491	-1.2032	0.9	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Dactylosporangium
6428	-1.21204	0.95	Actinobacteria	Actinobacteria	Rubrobacterales	Rubrobacteraceae	