



January 2018

Genomic Regulation Of Abiotic Stress Response In The Soil Nematode *Oscheius Tipulae*

Riley Mcglynn

Follow this and additional works at: <https://commons.und.edu/theses>

Recommended Citation

Mcglynn, Riley, "Genomic Regulation Of Abiotic Stress Response In The Soil Nematode *Oscheius Tipulae*" (2018). *Theses and Dissertations*. 2417.
<https://commons.und.edu/theses/2417>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, and Senior Projects at UND Scholarly Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UND Scholarly Commons. For more information, please contact zeinebyousif@library.und.edu.

GENOMIC REGULATION OF ABIOTIC STRESS RESPONSE IN THE SOIL
NEMATODE *OSCHEIUS TIPULAE*

by

Riley D. McGlynn
Bachelor of Arts, Concordia College, Moorhead, MN, 2012

A Dissertation
Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

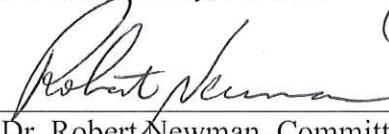
Grand Forks, North Dakota

December
2018

This dissertation, submitted by Riley McGlynn in partial fulfillment of the requirements for the Degree of Doctor of Philosophy from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.



Dr. Brain Darby, Advisor



Dr. Robert Newman, Committee Member



Dr. Turk Rhen, Committee Member



Dr. Rebecca Simmons, Committee Member



Dr. Archana Dhasarathy, Committee Member

This dissertation is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.



Grant McGimpsey
Dean of the School of Graduate Studies

December 5, 2018
Date

PERMISSION

Title Genomic Regulation of Abiotic Stress Response in the Soil Nematode
Oscheius tipulae

Department Biology

Degree Doctor of Philosophy

In presenting this dissertation in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my dissertation work, or in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this dissertation or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my dissertation.

Riley D. McGlynn
December 6, 2018

TABLE OF CONTENTS

LIST OF FIGURES.....	viii
LIST OF TABLES.....	x
ACKNOWLEDGMENTS.....	xiv
ABSTRACT.....	xv
CHAPTER	
I. LITERATURE REVIEW.....	1
Introduction.....	1
Phylum Nematoda.....	1
<i>Caenorhabditis elegans</i>	2
<i>Oscheius tipulae</i>	4
Heat Stress: Threats and Response.....	4
Desiccation Stress: Threats and Response	6
Freezing Stress: Threats and Response	10
Genetic Regulation.....	12
Genome Sequencing.....	17
Nematode Genomics.....	19
DNA Methylation.....	22
Research Goals.....	23
References.....	28

II.	DRAFT GENOME OF THE KJO STRAIN OF THE SOIL NEMATODE <i>OSCHEIUS TIPULAE</i>	37
	Abstract.....	37
	Introduction.....	38
	Materials and Methods.....	39
	Nematode Preparation.....	39
	Genomic DNA Extraction and Genome Assembly...	40
	Mitochondrial Genome.....	42
	Synteny.....	43
	Genome Annotation.....	43
	Positive Selection.....	44
	Results and Discussion.....	45
	Genome Assembly.....	46
	Genome Annotation.....	46
	Gene Content.....	48
	Mitochondrial Genome.....	49
	Synteny.....	51
	Positive Selection.....	54
	Conclusions.....	56
	References.....	58
III.	ABIOTIC STRESS RESPONSE TRANSCRIPTION PROFILES IN THE SOIL NEMATODE <i>OSCHEIUS TIPULAE</i>	61
	Abstract.....	61
	Introduction.....	62

Materials and Methods.....	67
Nematode Preparation.....	67
Stress Treatments.....	68
RNA Extraction.....	69
Sequencing.....	70
RNA-Seq.....	70
Positive Selection.....	72
Results and Discussion.....	72
Transcription Profiles.....	72
Upregulated Genes.....	76
Downregulated Genes.....	83
Positive Selection.....	90
Conclusions.....	93
References.....	96
IV. PRESENCE AND LOCATION OF DNA METHYLATION IN THE GENOMES OF THE SOIL NEMATODES <i>OSCHEIUS TIPULAE</i> AND <i>CAENORHABDITIS ELEGANS</i>	100
Abstract.....	100
Introduction.....	101
Materials and Methods.....	104
Nematode Growth and DNA Extraction.....	104
Library Preparation and Sequencing.....	105
Whole-Genome Bisulfite Sequencing Analysis.....	106
MethylCap-Seq Analysis.....	107

	Methyltransferase Presence.....	109
	Role of Methylation on Abiotic Stress Response....	109
	Results and Discussion.....	110
	Whole-Genome Bisulfite Sequencing.....	110
	MethylCap-Seq.....	111
	Methyltransferase Presence.....	116
	Role of Methylation on Abiotic Stress Response....	117
	Conclusions.....	119
	References.....	121
V.	EPILOGUE.....	124
	The <i>Oscheius tipulae</i> Genome.....	125
	Genomic Regulation of Abiotic Stress in <i>O. tipulae</i>	125
	The Presence and Role of DNA Methylation.....	126
	Conclusion.....	127
	APPENDICES.....	129

LIST OF FIGURES

Figure	Page
1.1. Simplified organizational models for genomic stress response centered on the one-gene-one-polypeptide hypothesis.....	13
1.2. Simplified organizational model for genomic stress response centered on the idea of pleiotropy	15
2.1. Synteny comparison between the <i>O. tipulae</i> and <i>C. elegans</i> genomes.....	52
2.2. Synteny comparison between the KJO draft genome scaffolds and the CEW1 draft genome scaffolds	53
2.3. Blast2GO Gene Ontology analysis of the predicted <i>O. tipulae</i> genes.....	55
2.4. Histogram of the dN/dS ratios of the nematode reciprocal orthologs.....	57
3.1. Heatmap of relative transcription levels across all treatment replicates.....	73
3.2. Multi-dimensional analysis of transcription pattern replicates.....	75
3.3 Appendix. RNA-Seq dispersion estimates.....	130
3.4 Appendix. RNA-Seq fitted dispersions	131
3.5. Venn diagram of genes upregulated in at least one of the three treatments.....	82
3.6. Venn diagram of genes downregulated in at least one of the three treatments.....	88
3.7 Appendix. Differentially expressed genes under the heat treatment.....	132
3.8 Appendix. Differentially expressed genes under the desiccation treatment.....	133
3.9 Appendix. Differentially expressed genes under the freezing treatment.....	134
3.10 Histogram of the dN/dS ratios for upregulated genes.....	91
3.11 Venn diagrams of the positively selected genes.....	92

3.12 Histogram of the dN/dS ratios for the downregulated genes..... 94

LIST OF TABLES

Table	Page
2.1 Genome assembly statistics for this <i>Oscheius tipulae</i> genome and various other nematode species.....	47
2.2 Genomic content makeup of the assembled KJO <i>O. tipulae</i> genome.....	50
3.1 Lists of a subset of the known molecular and physiological responses nematodes have to three abiotic stresses: heat, desiccation and freezing.....	66
3.2 Appendix. DAVID functional annotation results for all genes upregulated under at least one abiotic stress.....	135
3.3 Appendix. DAVID functional annotation results for all genes downregulated under at least one abiotic stress.....	138
3.4 Appendix. DAVID functional annotation results for all genes upregulated under heat stress.....	141
3.5 Appendix. DAVID functional annotation results for all genes upregulated under desiccation stress.....	144
3.6 Appendix. DAVID functional annotation results for all genes upregulated under freezing stress.....	147
3.7 Appendix. DAVID functional annotation results for all genes upregulated under only heat stress.....	150
3.8 Appendix. DAVID functional annotation results for all genes upregulated under only desiccation stress.....	153

3.9 Appendix. DAVID functional annotation results for all genes upregulated under only freezing stress.....	156
3.10 Appendix. DAVID functional annotation results for all genes upregulated under both heat and desiccation stresses.....	159
3.11 Appendix. DAVID functional annotation results for all genes upregulated under both heat and freezing stresses.....	162
3.12 Appendix. DAVID functional annotation results for all genes upregulated under both freezing and desiccation stresses.....	164
3.13 Appendix. DAVID functional annotation results for all genes upregulated under heat, freezing, and desiccation stresses.....	167
3.14 Appendix. DAVID functional annotation results for all genes downregulated under heat stress.....	170
3.15 Appendix. DAVID functional annotation results for all genes downregulated under desiccation stress.....	173
3.16 Appendix. DAVID functional annotation results for all genes downregulated under freezing stress.....	176
3.17 Appendix. DAVID functional annotation results for all genes downregulated under only heat stress.....	178
3.18 Appendix. DAVID functional annotation results for all genes downregulated under only desiccation stress.....	180
3.19 Appendix. DAVID functional annotation results for all genes downregulated under only freezing stress.....	183

3.20 Appendix. DAVID functional annotation results for all genes downregulated under both heat and desiccation stresses.....	186
3.21 Appendix. DAVID functional annotation results for all genes downregulated under both heat and freezing stresses.....	189
3.22 Appendix. DAVID functional annotation results for all genes downregulated under both desiccation and freezing stresses.....	190
3.23 Appendix. DAVID functional annotation results for all genes downregulated under heat, desiccation, and freezing stresses.....	193
4.1 Appendix. List of DNA methyltransferase genes used in reciprocal Blastp analyses.....	196
4.2 Annotation of hypermethylated region peak locations within the <i>O. tipulae</i> genome compared to total bp content in genome.....	113
4.3 Annotation of hypermethylated region peak locations within the <i>C. elegans</i> genome compared to total bp content in genome.....	115
4.4 Appendix. NCBI Blastp results from mapping DNA (5-cytosine-)- methyltransferase (DNMT) genes against <i>Oscheius tipulae</i> predicted genes.....	201
4.5 Appendix. NCBI Blastp results from blasting the <i>Oscheius tipulae</i> predicted genes obtained as a hit in initial blast against DNA (5-cytosine-)-methyltransferase (DNMT) genes.....	203
4.6 Appendix. NCBI Blastp results from mapping DNA N6- methyltransferase genes against <i>Oscheius tipulae</i> predicted genes.....	207

4.7 Appendix. NCBI Blastp results from blasting the <i>Oscieius tipulae</i> predicted genes obtained as a hit in initial blast against DNA N6-methyltransferase genes.....	209
4.8 Presence of hypermethylated-CG peaks within genes upregulated due to abiotic stress.....	118
4.9 Appendix DAVID functional annotation results for all genes both upregulated under at least one abiotic stress and hypermethylated.....	211

ACKNOWLEDGMENTS

I never would have been able to do all of this without those who have provided so much help and support over the years. First off, I'd like to thank my parents, David and Leslie McGlynn, for all the encouragement they've given me throughout my entire life. Their support has meant the world to me. I also thank my PhD advisor, Dr. Brian Darby, for all of his help, training, and guidance throughout my years at UND. I want to thank Dr. Rebecca Simmons, Dr. Robert Newman, Dr. Turk Rhen, Dr. Archana Dhasarathy, and Dr. Katherine Sukalski all for agreeing to be a part of my graduate committee and providing me with insightful input, guidance, and advice. I am also incredibly grateful to the graduate students I've worked alongside for offering their help when needed and for being a great group of friends as well. None of what I did could have been possible without the people around me, and for that, I will always be grateful.

ABSTRACT

Oscheius tipulae is a species of free-living soil nematode that can be found in ecosystems worldwide. Because of this, individuals must be able to respond to heat, freezing, and desiccation stresses in order to survive. They do this by producing a suite of cellular responses, some of which are necessary to survive multiple stresses, and some are stress-specific. While these cellular responses are well known, the ways in which they are regulated in a genome-wide context are not. In this project, multiple high throughput sequencing and bioinformatics analyses were utilized to answer this question. First, the *O. tipulae* genome was sequenced via Illumina HiSeq, assembled, and annotated. An RNA-Seq experiment was performed to determine transcription patterns within stress responses. Pooled nematode samples were subjected to heat, freezing, or desiccation stress prior to RNA sequencing and read mapping. Results showed that shared cellular responses were controlled by the upregulation of both shared and stress-specific genes. This suggests that the genome remains efficient by utilizing overlapping response genes and reinforcing them with stress-specific genes. Whole genome bisulfite sequencing and MethylCap-Seq analyses were performed to assess DNA cytosine methylation presence in *O. tipulae* and the model organism *Caenorhabditis elegans* and to determine its role in the abiotic stress response process in *O. tipulae*. Methylated cytosines were found in both *O. tipulae* and *C. elegans*, contradicting the historical belief that cytosine methylation is absent in nematodes. Changes in DNA methylation were not associated with the abiotic stress response as very few methylation sites were found within upregulated genes. This

project utilized new sequencing technologies and various bioinformatics programs to provide an in-depth look into the genome-wide responses to abiotic stress in *O. tipulae*.

CHAPTER I

LITERATURE REVIEW

Introduction

The soil nematode *Oscheius tipulae* is widely dispersed throughout the planet, and due to this, must face multiple abiotic stresses. Multiple studies have analyzed the responses to these stresses on the molecular and genetic levels, and have found that while each abiotic stress triggers a few unique survival responses, there is also extensive overlap in the way nematodes survive certain stresses. Genome sequencing technologies have become more affordable and more readily available, and a full genome-wide analysis across the multiple abiotic stress responses nematodes utilize can provide extensive amounts of information. A genome-wide approach can provide an overall view of how nematodes have evolved to survive abiotic stress and can promote understanding of how genomes respond to these stresses. Lastly, understanding the role of epigenetic regulation in nematodes is in the beginning stages. Previously believed to be absent in nematodes, DNA methylation could be involved in regulating gene transcription. While there is plenty already known about nematodes, there is still much more to understand regarding the patterns and mechanisms with which they are able to regulate gene transcription on a large-scale level.

Phylum Nematoda

Nematodes (Phylum: Nematoda) are a diverse group of multicellular roundworms that can be found living in every ecosystem on the planet. There are over 25,000

identified and described species of nematodes, and they make up approximately 80% of the abundance of multicellular animals on land and 90% in the deep sea (Danovaro et al., 2008; Z.-Q. Zhang, 2013). Populations of nematodes have been found in Antarctic tundra, hot water springs, and the deep-sea floor (Danovaro et al., 2008; Perry & Wharton, 2011). Species have also been found deep underground, kilometers under the Earth's surface (Borgonie et al., 2011). Due to the vast number of species and their adaption to each of the planet's ecosystems, nematodes have a wide variety of ecological niches. Both plants and animals can be parasitized by nematode species, and other species can live freely within soil and water.

Free-living soil nematodes feed on bacteria, fungi, and other microinvertebrates and can be found in incredibly high abundance as millions of individuals can be found within one square meter of soil. They are able to cycle the inorganic nitrogenous compounds they consume into ammonia, an organic form that can be absorbed and utilized by plants (Ingham, Trofymow, Ingham, & Coleman, 1985). This behavior gives them a key role in soil health and plant survival, and because they are found throughout many of the world's ecosystems, their impact is global.

Caenorhabditis elegans

The most well-known and well-studied species of nematode is the bacterivore species *Caenorhabditis elegans*. Populations of *C. elegans* can be easily grown, maintained, and manipulated under laboratory settings, making them a perfect candidate for laboratory experiments. Individuals are also very small and transparent, making them ideal for microscopy work, and the species has a rapid life cycle, allowing for the quick establishment of multiple generations. Throughout its history as a study subject, *C.*

elegans has had the developmental fate of every somatic cell detailed, providing an entire cell lineage for the organism (Kimble & Hirsh, 1979; Sulston, Schierenberg, White, & Thomson, 1983). Due to its ease of use under a microscope, the entire anatomy of the species has been well documented and is readily available online, but the largest development in its history as a study species came when the entire genome was sequenced by a group known as the The *C. elegans* Sequencing Consortium (1998). At approximately 100 million base pairs in length, the *C. elegans* genome was the first multicellular organism to have its complete genome sequenced and published. This quickly allowed it to be used for various types of genetics studies as the field of genomics truly began to grow.

Because it is so deeply studied and has been for so long, *C. elegans* is frequently used as a model organism. This means that studies performed with this species have been used to explain biological properties and phenomena for species other than itself. Many of the behaviors and biological properties of *C. elegans* have been assumed to be true and present not only across the entire phylum Nematoda, but also across all microinvertebrates. Studies done using the *C. elegans* model are used to provide insight into human-related issues and behaviors. The *C. elegans* model has been used to shed light on multiple neurodegenerative diseases such as Parkinson's disease, Alzheimer's disease, and Huntington's disease (Apfeld & Fontana, 2017; Bravo et al., 2018; Gama Sosa, De Gasperi, & Elder, 2012; Shen, Yue, Zheng, & Park, 2018; Zeng et al., 2016). It also is used as a model for health-related topics such as toxicology (Leung et al., 2008) and inflammation (Hendricks & Mylonakis, 2017). Researchers have also used *C. elegans* for studies of aging (Apfeld & Fontana, 2017; Pandey et al., 2018), obesity (Shen

et al., 2018), learning and memory (Rankin, Beck, & Chiba, 1990), and even the effects of e-cigarettes on humans (Panitz, Swamy, & Nehrke, 2015). Over time, more nematode species have begun to be studied in-depth, but none of them play as ubiquitous of a role as *C. elegans*.

Oscheius tipulae

One of these emerging model species, and the target species of this dissertation, is *Oscheius tipulae*. *O. tipulae* is a free-living species of soil nematode that can be found within the same taxonomic family as *C. elegans*: Rhabditidae. Because its primary feeding source is bacteria, populations can be easily grown and maintained in a laboratory setting. Populations of *O. tipulae* can be found in soils all over the globe, including those in central North America, Brazil, and the Tuscan Archipelago of Italy (Torrini et al., 2016). This species has primarily been used a model for nematode gonad development (Dichtel-Danjoy & Félix, 2004; Félix & Sternberg, 1997). Two strains of this species have been used to develop laboratory populations: the CEW1 strain from Brazil and the KJO strain from the Konza Prairie located south of Manhattan, Kansas, USA. This KJO strain was chosen for study in this dissertation because it faces a wide variety of abiotic stresses in its native Kansan soils. This makes it an excellent study species when it comes to nematode stress survival, and this dissertation will focus on three main abiotic stresses: heat stress, desiccation stress, and freezing stress.

Heat Stress: Threats and Response

Nematodes, like most organisms, have an optimal range of temperatures under which they can survive. Temperatures outside of that range have effects on the nematodes to which they must react. Specifically, high temperatures can alter nematode

development speed and life cycle (Anderson & Coleman, 1982). The biggest problem nematodes face under heat stress is protein denaturation and aggregation (Singer & Lindquist, 1998a). In order to combat this, nematodes will need to synthesize heat shock proteins (HSPs).

Heat shock proteins chaperone denatured proteins towards mechanisms that can either refold them into the correct conformation or degrade them completely (Wharton, 2011). By removing malfunctioning proteins, they prevent aggregation and increase cellular activity and productivity. In fact, HSPs perform this job so efficiently, that other molecular mechanisms responsible for this job under various other stresses (ie: proteasomes and certain nuclear granules) are inhibited from formation (Sampuda, Riley, & Boyd, 2017). Providing further molecular evidence, HSF-1, a transcription factor that mediates HSP transcription, has been shown to play a role in nematode heat stress response (Joo et al., 2016). HSPs are so important to a nematode's survival, both under stress and under normal conditions, that upon the silencing of certain HSP genes, other HSP genes begin to be over-transcribed as a response (Eckl, Sima, Marcus, Lindemann, & Richter, 2017).

Nematodes under heat stress will also increase trehalose levels by 90% (Jagdale & Grewal, 2003). Trehalose, a reducing disaccharide, is used as a reserve carbohydrate when energy demands increase, but more importantly, trehalose preserves the shape and activity of proteins, particularly enzymes, at higher temperatures (Honda, Tanaka, & Honda, 2010; Hottiger, Boller, & Wiemken, 1987; Hottiger, de Virgilio, Hall, Boller, & Wiemken, 1994; Lillie & Pringle, 1980; Singer & Lindquist, 1998b). It also reduces aggregation of proteins already denatured by higher temperatures and must be broken

down as temperatures levels decrease towards the nematodes optimal range because high trehalose levels can inhibit the refolding of the denatured proteins (Singer & Lindquist, 1998a). Trehalose also does not inhibit or interfere with the activity of enzymes as it protects them (Hottiger et al., 1994). Production of trehalase, the enzyme responsible for trehalose degradation, does not increase under high temperatures, indicating high trehalose levels are necessary during the heat shock response (Argüelles, 1997). Trehalose has also been shown to work synergistically with HSPs. By modifying the HSP C-terminal activation domain, trehalose can increase the effectiveness of the HSP, providing an even more efficient heat stress response (Bulman & Nelson, 2005).

Desiccation Stress: Threats and Response

A dry environment poses multiple threats to nematodes. The biggest threat is to the cell membrane as it can suffer phase changes as water levels drop (Crowe, Crowe, & Chapman, 1984). The membrane loses its fluidity and becomes more rigid, preventing the movement of membrane proteins. In doing so, membrane functions, most importantly transmembrane transport, see a drop in speed and efficacy.

Osmotic pressure brought on by desiccation will lead to rapid water loss across the membrane of the nematode, causing an individual to shrink and collapse. Dehydration also triggers the accumulation of reactive oxygen species (ROS) (Adhikari, Wall, & Adams, 2009). ROS are oxygen-containing molecules that are highly chemically reactive, and at high concentrations, can react and modify macromolecules such as proteins, lipids, and nucleic acids. Oxidation compromises the molecular function, leading to the failure of important biochemical processes necessary for survival. ROS accumulation causes the mitochondrial electron transport chain to malfunction, and this in turn creates more ROS.

Protein unfolding and aggregation is also common during desiccation stress (Potts, 1994). Protein unfolding causes a suite of intracellular problems including inhibition of enzymatic reactions, breakdown of the cell's physical structures, and the cessation of molecular transport and cell signaling. Protein aggregation further exacerbates this problem as large structures of denatured proteins begin to take up much of the cellular space. Protein aggregation is known to be harmful and is linked with multiple neurodegenerative diseases in humans (Ross & Poirier, 2004).

Nematodes utilize a wide set of behavioral and molecular responses to survive stress brought on by desiccation. Behaviorally, individual nematodes will coil in on themselves (Demeure, Freckman, & Van Gundy, 1979). Coiled nematodes will then come together to form larger groups during desiccation. Only those individuals on the outside of the group are likely to die from the dry conditions. This outer layer of dead nematodes then helps preserve the rest of the population by conserving the water within the group (Higa & Womersley, 1993).

Molecularly, levels of HSPs are increased under desiccation stress just as they are under heat stress (Adhikari et al., 2009). As under heat stress, HSPs act as molecular chaperones, helping transport denatured proteins toward organelles that can break them down and remove them from interfering with the remaining intracellular processes. Late embryogenesis abundant (LEA) proteins are also necessary for desiccation survival. LEA proteins not only play multiple roles in the stress response process, but they are so essential to surviving desiccation that the silencing of a single LEA gene is enough to significantly decrease nematode survival (Gal, Glazer, & Koltai, 2004).

One of the roles of LEA proteins is acting as a molecular chaperone to prevent protein aggregation and maintain protein activity. According to a study by Goyal et al. (2005), the enzyme citrate synthase, an enzyme that begins the citric acid cycle, was shown to maintain functionality under desiccation stress in the nematode *Aphelenchus avenae* when treated with LEA proteins. Without the LEA proteins, enzymatic activity dropped significantly. Protein aggregation was high under desiccation and nearly eliminated under low levels of LEA proteins. LEA proteins preserve protein structure and function under times of desiccation stress in studies of both pea (Grelet et al., 2005) and citrus plants (Sanchez-Ballesta, Rodrigo, Lafuente, Granell, & Zacarias, 2004).

Certain LEA proteins can be found in high amounts under normal conditions despite low mRNA levels. It is not until gene expression increases under desiccation stress that the LEA protein can be cleaved into multiple subunits that each act as anti-aggregants. This allows for the most rapid response because the proteins are present before the stress occurs, and it causes a maximum response because each single protein can create multiple anti-aggregants (Goyal, Pinelli, et al., 2005).

LEA proteins and LEA protein-like molecules also have ion binding properties that allow them to act as an antioxidant (Alsheikh, Svensson, & Randall, 2005; Hara, Fujinaga, & Kuboi, 2005; Heyen et al., 2002; Wise & Tunnacliffe, 2004). LEA proteins aid in ROS-induced stress by binding and stabilizing newly created ions and free radicals, preventing them from reacting with biochemically important cellular components. Certain LEA proteins even act as “radical scavenging proteins” that seek out ROS free radicals, allowing themselves to be reacted upon and degraded, preserving the structure and function of more essential macromolecules (Hara, Fujinaga, & Kuboi, 2004).

Genetic studies have also supported the idea that LEA proteins act as an important antioxidant in dealing with ROS under desiccation stress. Tyson et. al (2007) showed that under desiccation stress, levels of LEA-protein associated ESTs known to handle ROS accumulation were upregulated. LEA protein genes are also upregulated upon the introduction of hydrogen peroxide, a type of ROS (Desikan, Niell, & Hancock, 2000).

Organic osmolytes, molecules influencing osmosis, are also found in greater amounts under desiccation stress. Sugars, polyols, and certain amino acids fall under this category, but the most predominant and extensively studied osmolyte produced is trehalose, a glucose disaccharide (W. J. Welch & Brown, 1996). Lipid and glycogen levels drop under desiccation stress in order to create trehalose. At stable conditions, trehalose is used as an enzyme and food protectant, and trehalose levels can drop to 7% of their natural levels without repercussions to the nematode (Womersley & Higa, 1998). As trehalose levels increase, it associates with cell membranes. The hydroxyl group of the sugar forms a hydrogen bond with the phosphate group in the head of a phospholipid within the membrane bilayer (Behm, 1997). This separates the phospholipids in order to retain the membrane's liquid state and stabilizes water loss (Crowe et al., 1984; Higa & Womersley, 1993). Trehalose also prevents oxidative damage to the membrane and inhibits browning reactions under desiccation stress. This also preserves protein structures by stopping the decay of primary amines (Higa & Womersley, 1993).

LEA proteins and trehalose have also been shown to work closely together. In the nematode *Steinernema feltiae*, levels of Desc47, a protein with a partial structure and biochemical function similar to typical LEA proteins, increased as trehalose began to accumulate (Solomon, Salomon, Paperna, & Glazer, 2000). Synergy between LEA

proteins and trehalose is also found in in the nematode *Aphelenchus avenae*, where high levels of LEA proteins increased the efficacy of trehalose under water-related stress (Goyal, Walton, et al., 2005).

Freezing Stress: Threats and Response

As water both inside and outside the nematode becomes frozen, it is no longer biologically active, and because of this, there is much overlap in the specific biological dangers nematodes face under the freezing stress as under desiccation stress. Proteins lose structure and function while undergoing freezing temperatures (Ramløtv, 2000). As in desiccation stress, membrane structure is also changed, as it becomes less fluid and more gelatinous, causing problems with cellular transport and recognition. A unique threat brought on by freezing stress is intracellular freezing, a process that can be lethal to nematodes (Mazur, 1984). Low temperatures also slow movement of particles, drastically decreasing the rates of intracellular biological reactions.

In order to survive freezing, nematodes utilize a variety of behavioral and molecular methods. One mechanism utilized by nematodes is to simply resist the freezing of water. Ice nucleators are substances that are known to induce the formation of ice crystals as water temperature drops below freezing. Nematodes can inhibit this activity by producing and releasing inhibitory substances into the medium they inhabit that will prevent the nucleator from inducing the surrounding water to change from liquid to solid form (Wharton & Worland, 1998). By doing this, the intracellular and extracellular water that is essential for the nematode's survival remains a biologically available liquid.

Another method which allows for nematode survival in freezing temperatures is their ability to turn this freezing stress into a desiccation stress. Nematodes have been

shown to induce desiccation as a way of resisting internal freezing (Wharton, 2003).

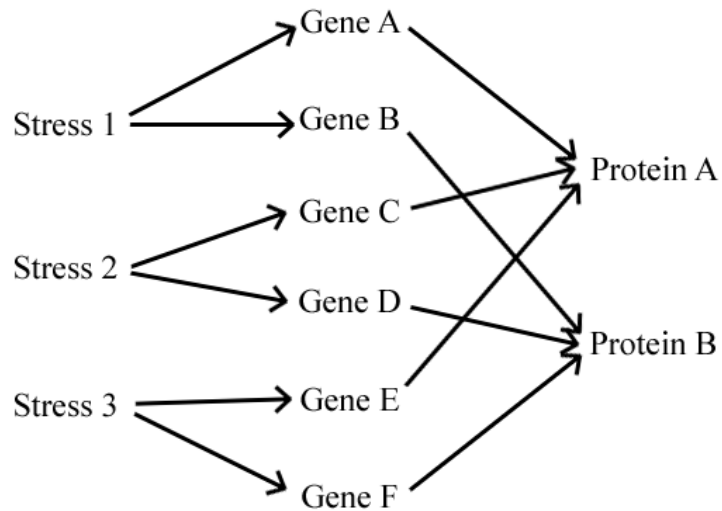
Water that has remained in a liquid phase despite being below its freezing point is called supercooled water. As this water is purposefully dehydrated across the nematode's permeable membrane, the internal water pressure increases. Once this water pressure matches that of the external medium, the threat of internal freezing is removed (Holmstrup, Bayley, & Ramløv, 2002).

A variety of molecular responses found under freezing stress are also found within the stress responses for either heat or desiccation. Firstly, nematodes will produce high levels of unsaturated and polyunsaturated fatty acids. These molecules can be used as a false hydration agent, binding to molecules within the membrane and allowing it to maintain its fluid structure. They are also used as an alternate energy source to ensure the nematode has the capacity to survive the stress (Margesin, Neuner, & Storey, 2007). Similarly, nematodes respond to freezing stress by producing high levels of LEAs and organic osmolytes, each providing the same essential roles in survival as they did under desiccation stress. Again, trehalose also plays an important role in freezing survival as levels can increase up to 350% the normal level (Jagdale & Grewal, 2003). It has also been shown that activity of trehalose-6-phosphate, an enzyme responsible for the formation of trehalose, significantly increases while under cold stress, whereas activity of trehalase, the enzyme responsible for breaking down trehalose, significantly decreased (Jagdale, Grewal, & Salminen, 2005). Lastly, it has also been shown that the presence and activity of heat shock proteins also increase while under cold stress (Martinez, Perez-Serrano, Bernadina, & Rodriguez-Caabeiro, 2001; C. Zhang & Guy, 2006). Functionally, these proteins act as to chaperone denatured proteins to their disposal.

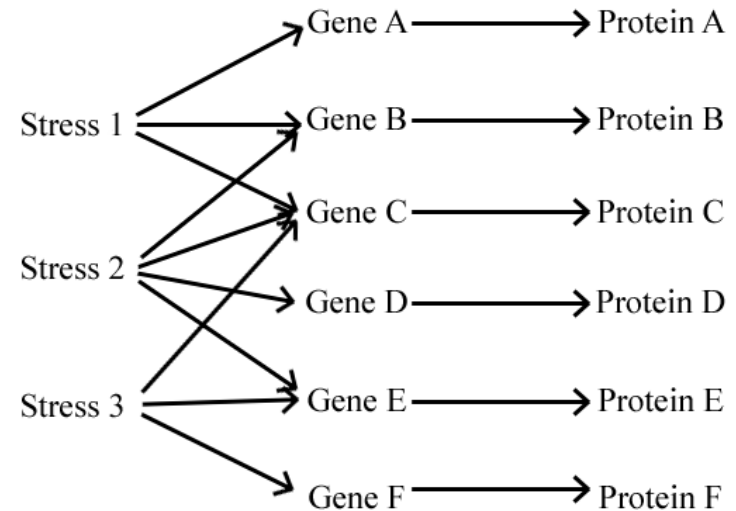
Genetic Regulation

The central dogma of biology states that the nucleotide sequences found in DNA are transcribed into messenger RNA (mRNA) before being translated into a protein made up of amino acids. Initially, it was proposed that each gene coded for one unique enzyme, which eventually became known as the “one-gene-one-enzyme hypothesis” (Beadle & Tatum, 1941). This name fell out of favor when scientists realized non-enzymatic proteins were also produced by the transcription and translation of DNA, but the one-codes-for-one idea still persisted. Since then, researchers have discovered molecular processes that allow for the modification of this central dogma in order to ensure gene translation is at its most efficient. When multiple genes are involved in a larger phenotypic function, such as a stress response, different genetic regulation mechanisms and genome-wide organizational models allow organisms to maximize efficiency at an even larger scale.

For example, under the one-gene-one-polypeptide hypothesis, a stress would signal the transcription of a single gene that would eventually code one protein. This would happen hundreds of times across the genome as each stress response gene would need its corresponding gene to be transcribed and translated individually. Figure 1.1 demonstrates the two organizational models possible in regards to stress response under the one-gene-one-polypeptide hypothesis: different stresses could each activate the transcription of its own suite of genes, or there can be overlap based on common response proteins that are needed. What these two models have in common is that each gene corresponds to only one phenotypic protein.



A



B

13 Figure 1.1 Simplified organizational models for genomic stress response centered on the one-gene-one-polypeptide hypothesis. A. Each stress activates the transcription of a unique suite of genes that in turn are responsible for the production of various proteins. Overlapping necessary protein products is possible. B. Each stress activates the transcription of a set of genes that may or may not be shared with other stress responses. Each gene is then responsible for the production of a unique protein.

Conversely, it has been shown that a single gene can lead to multiple phenotypic responses in a process known as pleiotropy (Figure 1.2). These phenotypic responses may or may not be similar in nature, and this process does not include the alternate splicing of genes, as is found in higher organisms. Instead, it is hypothesized that pleiotropy developed as genes with similar original functional and developmental functions were co-selected (Wagner, 1996). Over time, these genes became genetically integrated on a molecular level, which allowed natural selection to act upon them as one unit instead of multiple units (Cheverud, 1996). Genes that do not have similar functions will not be co-selected, meaning they will become neither genetically nor evolutionarily integrated. Pleiotropy will not develop in these genes.

Evolutionarily, the most important consequence of pleiotropy is that it reduces the cost of complexity in organisms (Wang, Liao, & Zhang, 2010). The cost of complexity is the idea that complex organisms will develop and adapt at a much slower rate than simpler organisms. More complex organisms have more traits under selection than simple organisms. This means that the effect one mutation will have is progressively diminished as the number of individual and unique traits increases (Orr, 2000). By having one gene control multiple phenotypes, the number of unique traits under selection decreases, thereby decreasing the cost of complexity and allowing complex organisms to adapt and develop quicker and more effectively.

Another possible mechanism of genomic regulation is a process called modular pleiotropy, or modularity. Modularity is a system where one gene produces multiple phenotypic responses, but those responses are functionally similar and are few in number (Wagner, 1996). Rather than one gene being responsible for dozens of varying

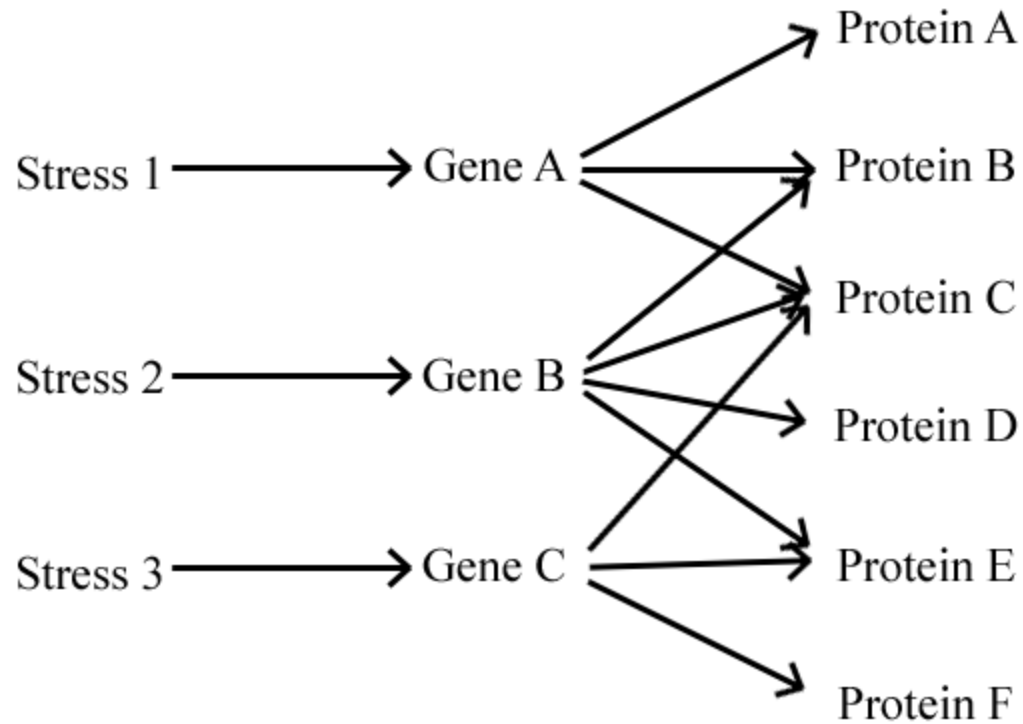


Figure 1.2 Simplified organizational model for genomic stress response centered on the idea of pleiotropy. Each stress activates the transcription of one gene (or suite of genes) that are each in turn responsible for the production of multiple phenotypic proteins. There may or may not be overlaps in the proteins that are created across the genes.

phenotypic responses, it may be responsible for fewer, more similar, responses. Two processes produce modularity within a genome. The first is called integration and is similar to the way in which pleiotropy develops: the co-selection of two functionally related genes. The second is called parcellation and it assumes the primitive state of the genome is pleiotropic. During parcellation, pleiotropic effects of functionally unrelated genes are selected against and eliminated. Essentially, both directional and stabilizing selection act upon a series of genes to organize them into functionally related gene modules (Wagner, 1996).

Modularity is a balance between the one-gene-one-polypeptide hypothesis and pleiotropy. It ensures that similar traits are unable to evolve independently from one another as there are more functional links within a module than across multiple modules (Wang et al., 2010; J. J. Welch & Waxman, 2003). Modularity limits the number of phenotypic traits that can be affected by a mutation, whether it is advantageous or deleterious. In order for the gene modules to evolve, the mutation must be advantageous for all resulting phenotypes. If the mutation is advantageous for some while deleterious for others, no evolution will take place (J. J. Welch & Waxman, 2003).

In practice, it is unlikely that these three forms of genome regulation are all mutually exclusive. It is possible that all three can be found within the same genome. For example, in a study done across the baker's yeast *Saccharomyces cerevisiae*, the nematode *C. elegans*, and the house mouse *Mus musculus*, it was found that there were generally low levels of pleiotropy but that the relationship between genes and their phenotypic traits was highly modular (Wang et al., 2010). It is therefore important to not

attempt to prove one of these three mechanisms correct, but to instead attempt to assess the extent at which they are present within the genome.

Genome Sequencing

The three-dimensional double helix structure of DNA was discovered in 1953 (Watson & Crick). This inspired researchers to begin to find ways to sequence DNA. This process involved two important steps: discovering which nucleotides were present and discovering the order in which they were located. RNA became the first target of nucleic acid sequencing as known RNase usage and treatment with selective ribonucleases allowed researchers to determine both composition and directionality. The first nucleic acids sequenced were a series of tRNAs (Holley et al., 1965), and rRNA quickly followed (Brownlee, Sanger, & Barrell, 1967).

The biggest breakthrough in DNA sequencing was established by Frederick Sanger in 1977. His method involved the use of radioactively labeled dideoxynucleotides that terminate the synthesis of DNA when incorporated into the strand (Sanger & Nicklen, 1977). Sanger sequencing allowed for faster and more reliable sequencing of DNA than the RNA methods, but it still could only sequence nucleotide chains no larger than 1 kilobase (kb). These sequences then had to be overlapped *in silico* in order to generate the complete sequence. This is known as shotgun sequencing. Sanger sequencing also helped establish the method of ‘sequencing-by-synthesis’ (SBS), where the incorporation of nucleotides by DNA polymerase during DNA synthesis is exploited in order to determine the sequence of DNA being created.

This combination of SBS and shotgun sequencing continued to be the method of choice as technology developed and sequencing techniques improved. The next milestone

in the history of DNA sequencing was the development of a process known as pyrosequencing. Pyrosequencing exploits the enzymes ATP sulfurylase and luciferase during pyrophosphate synthesis, which occurs as DNA polymerase synthesizes a DNA strand. Nucleotides are washed in turn over template DNA strands that are attached to a solid plate, and DNA polymerase incorporates them into the synthesizing DNA strand, pyrophosphate is released, ATP sulfurylase converts this into ATP, which is used as a luciferase substrate. Luciferase produces light that can be detected and measured where color presence and intensity correspond to nucleotide type and amount (Margulies et al., 2005). Pyrosequencing allows for mass parallel sequencing efforts as multiple reactions could be done concurrently, drastically increasing the amount of DNA that could be sequenced in one run. This method was eventually licensed to 454 Life Sciences and became known as 454 sequencing.

The latest development in DNA sequencing is an improvement of 454 sequencing that now belongs to Illumina. This method used two oligo adaptors that are added to either side of the target DNA molecule to be sequenced. These adaptors attach to complimentary sequences on a flowcell before nucleotides are washed over and DNA polymerase uses them to create a complimentary DNA strand, allowing the DNA to be sequenced. Eventually, the original strand is removed, and the newly sequenced strand must fold over and polymerize with neighboring adaptor sequences before sequencing can continue. This process is known as 'bridge amplification'. Pyrosequencing only produces short reads approximately 50-200 bp in length, but because this is occurring at either end of the same DNA sequence, these two sequencing reads can then be paired together. This is known as 'paired-end' data, and it improves the accuracy of steps the

genome assembly step by providing known distances between the two reads (Illumina, 2010). Illumina is currently the most popular form of DNA sequencing, providing multiple sequencing machines that allow for variations in read length, sequencing depth, and cost for the researcher.

Sequencing only produces short reads containing approximately 50-200 base pairs. In order to turn these reads into an entire genome, one must use a variety of bioinformatics programs. This means that the rest of the sequencing process must be done *in silico*. The simplest way this is done is by partially overlapping the reads, and as they overlap, they extend the sequence, one base pair at a time. In the process of this overlapping, three different algorithms may be used by various programs. These three algorithms are the Overlap/Layout/Consensus algorithm, the de Bruijn graph method, or the greedy algorithms that may use either of the previous two methods (Miller, Koren, & Sutton, 2010).

Nematode Genomics

C. elegans was the first multicellular organism to have its genome sequenced, and many more have followed, including other nematodes. Well over 50 nematode species have had their genome sequenced, and more are being published at a very fast rate. Each species is inevitably compared the *C. elegans* model genome, and what has been found has been fairly surprising. While most nematodes look very similar, small and transparent with slight changes in anatomy, they are found to be vastly different genetically. Genome sizes have been shown to range from 49 million base pairs (Mb) to approximately 370 Mb, and while some species have 6 pairs of chromosomes, others have shown to have 5

pairs. While analyzing the diversity of nematode genomes, the *C. elegans* model is traditionally used as a base.

The *C. elegans* genome is made up of six pairs of chromosomes, one of which is a pair of sex chromosomes. *C. elegans* is a hermaphroditic species with no Y chromosome. Males are formed in the presence of a single X chromosome; a hermaphrodite is formed in the presence of an XX pair. The genome is approximately 100Mb in length, and it contains around 20,000 protein-coding genes (The *C. elegans* Sequencing Consortium, 1998). The species is also a rare eukaryote known to have its genome organized into operons, genes that are transcribed as a single unit. It is estimated that 15% of the *C. elegans* genes are located in over 1,000 operons, each ranging from 2-8 genes in length (Blumenthal et al., 2002).

The *Prinstionchus pacificus* genome provides evidence that nematode species have diverse genomes, even when they are taxonomically closely related. The species is also a microscopic free-living nematode found within the same taxonomic order (Rhabditida) as *C. elegans*, but their genomes are quite different. While it has the same six chromosomes as *C. elegans*, the *P. pacificus* genome is 169 Mb in length, approximately 69% larger than that of *C. elegans*. It also contains ~45% more protein-coding genes than *C. elegans* with an estimated 29,000 (Dieterich et al., 2008).

Brugia malayi is another commonly studied nematode species due to its role as a human parasite known to cause Lymphahtic filariasis. *B. malayi* is found in a different taxonomic class than *C. elegans* (Scernentea, while *C. elegans* is located in Chromadorea), and the two have very different genetic makeups. The most notable difference is in their chromosomes. *B. malayi* have one less chromosome pair than *C.*

elegans, and their sex chromosome pair follows an XY/XX sex-determining system, rather than the hermaphroditic X0/XX system. At ~88 Mb in length, the *B. malayi* genome is approximately 12% smaller in size, but it contains ~42% less protein-coding genes with an estimated 11,500 (Ghedini et al., 2007).

The species *Oscheius tipulae* is found within the same taxonomic family, Rhabditidae, as *C. elegans*, and has been extensively studied as a model for nematode vulva development (Dichtel-Danjoy & Félix, 2004; Félix & Sternberg, 1997). Genetically, it is known that the *O. tipulae* genome works in operons similar to those found in *C. elegans* (Evans et al., 1997). A preliminary study using a mixture of fluorescent detection and mathematical modeling done on the genome of *O. tipulae* placed its statistics close to that of *C. elegans*. It was predicted that the genome is 100.8 Mb long, and contains approximately the same number of 20,000 protein-coding genes (Ahn & Winter, 2006). It was not until very recently that the genome on a specific *O. tipulae* strain, the CEW1 strain, was sequenced and published. The CEW1 strain, obtained from Brazilian soils, has a genome size of approximately 59 Mb in length and contains slightly fewer than 15,000 protein-coding genes (Besnard, Koutsovoulos, Dieudonne, Blaxter, & Felix, 2017). Both of these are much smaller than previous assumptions. While this is a necessary first step in understanding the genome of *O. tipulae*, more work can be done to ensure that these statistics are an accurate reflection of the actual genome size. Also, because this species is even more widely dispersed throughout the world's soils than *C. elegans*, further work can shed light on whether or not this species has conserved genome structure across the various populations throughout the world.

DNA Methylation

The methylation of DNA is part of the field of epigenetics, a field that has existed for over 70 years. The term “epigenotype” was first used in 1942 as a way to describe the molecular processes that connect the genotype to the phenotype (Waddington). By this, he meant the mechanisms that organisms can use to regulate how the genetic code needed is utilized, inhibited, and modified as the amino acid sequences it produces are needed. Seventy years later, much is known about the epigenome. Transcription factors, double-stranded RNA interference, histone modification, and the methylation of DNA are all ways in which the expression of individual genes can be regulated.

The role that epigenetics play in nematodes is currently poorly understood. *C. elegans* has been shown to utilize double-stranded RNA to silence gene expression (Fire et al., 1998), but DNA methylation has traditionally been thought to be nonexistent. One marker of DNA methylation is the presence of the DNA methyltransferase proteins. These enzymes are what is ultimately responsible for physically transferring methyl groups onto cytosines within the DNA, and these enzymes are also highly conserved across species where DNA methylation is heavily used (Kumar et al., 1994). Methyl binding domains (MBDs) bind and interact with methylated CpG regions in DNA and are also used as a diagnostic for DNA methylation presence. There has been no indication of MBD orthologs in *C. elegans* and out of many, only a select few have been found in other invertebrates (Hendrich & Tweedie, 2003). In these invertebrates that utilize DNA methylation, the methylation pattern is more modular than the global methylation pattern found within vertebrates (Tweedie, Charlton, Clark, & Bird, 1997).

Since then, more DNA methylation work has been done on nematodes. Cytosine methylation has been found in the species *Trichinella spiralis* (Gao et al., 2012), but this species is taxonomically very different than *O. tipulae*. Greer et al. (2015) have used a series of antibody studies, immunofluorescence, mass spectrometry, and DNA sequencing to show that there is methylation of adenine (6mA) in *C. elegans*, a trait that is more commonly studied in prokaryotes. Currently, there has yet to be any extensive work done on the *O. tipulae* genome, despite the fact that DNA methylation studies are now available to be done via high-throughput sequencing.

Whole-genome bisulfite sequencing (WGBS) uses the chemical properties of both DNA and a bisulfite treatment to detect DNA methylation. During the bisulfite treatment, non-methylated cytosines are converted into uracil. Once sequenced, they are read as thymine, and upon mapping back to the untreated genomic sequence, any cytosines that remain should theoretically be methylated (Illumina, 2014). MethylCap-Seq is another way in which high throughput sequencing can detect methylation. In this process, genomic DNA is randomly fragmented in pieces that average 300 bp in length. The fragments are then exposed to MBD proteins bound to a magnetic bead. If the fragment contains a methylation site, the protein will bind to it. A magnet is then used to remove the captured DNA, and the unbound, non-methylated DNA is washed away. The captured DNA is then sequenced, and these sequences represent areas of methylation (Brinkman et al., 2010). These methods have not been extensively used to analyze DNA methylation in nematodes.

Research Goals

The purpose of this project was to determine how abiotic stress responses are regulated within the genome of *Oscheius tipulae*. In order to do this, the genome was sequenced and annotated, including the prediction of protein-coding gene sequences and locations. Next, after subjecting nematodes to heat, freezing, or desiccation stress, an RNA-Seq experiment was performed, providing read mapping estimates to act as a measure of gene transcription, which could be compared to a control sample and other treatments in order to understand the transcription patterns involved. Lastly, WGBS and MethylCap-Seq analyses were performed in order to assess the presence, location, and influence of DNA methylation on the genome in general, and more specifically, in the genomic responses to abiotic stress.

Chapter II - Draft Genome of the KJO Strain of the Soil Nematode *Oscheius tipulae*

In my second chapter, I generated an annotated draft genome assembly of the free-living soil nematode species *Oscheius tipulae*. The KJO strain, obtained from the soils of the Konza Prairie located south of Manhattan, Kansas, USA, was chosen. This genome assembly will hopefully be further groundwork in establishing *O. tipulae* as a satellite model organism, one closer in taxonomic relation to *C. elegans* than previously well-studied nematode species.

Much of what is known about this species is related to its development, primarily that of its vulva (Dichtel-Danjoy & Félix, 2004; Félix & Sternberg, 1997; Louvet-Vallee, Kolotuev, Podbilewicz, & Felix, 2003; Sommer, 2005). While this information is useful, it is also isolated, needing a further support system in the organism's genome. By establishing a well-assembled draft genome, it can provide a bank of information that contains both answers and new questions with which *O. tipulae* can be understood.

For years, the results of studies performed on *C. elegans* have been extrapolated and assumed to be true for not only other nematode species, but also microinvertebrates as a whole. By establishing *O. tipulae* as a satellite model organism, this need for assumption-based understanding is decreased, and *C. elegans*' role as a behavioral and genetic model organism can be evaluated.

Recent genome sequencing work has been performed on the CEW1 strain of *O. tipulae* obtained from Brazil (Besnard et al., 2017). This sequencing of the KJO strain will build on this information by providing a comparison between the KJO and CEW1 strains that will shed light on possible variations in genome structure across populations isolated across continents.

Chapter III – Abiotic Stress Response Transcription Profiles in the Soil Nematode *Oscheius tipulae*

In my third chapter, I used the annotated genome and RNA-Seq data to analyze the ways in which *O. tipulae* regulates its responses to heat, freezing, and desiccation stress on a genome-wide level. This study used laboratory techniques to mimic the environmental extremes this species faces in its natural habitat. Multiple previous nematode stress studies have been done on species other than *O. tipulae*, and a good portion of them involve seeing how a species responds to a stress it does not actually encounter in the wild (Ali & Wharton, 2013; Grewal, Gaugler, & Wang, 1996; Jagdale & Grewal, 2003). Other studies that subject nematodes to stresses they frequently face are focused on a single niche stress that the species is specialized in surviving. They document the molecular phenotypic responses to the stress, i.e. the molecular compounds that are being produced in order to physically combat the lethal intracellular issues

brought on by the stress (Raymond & Wharton, 2013; Wharton, 2003). Those studies tend to only focus on a single or a select number of genes or genetic properties (Adhikari, Wall, & Adams, 2010; Bulman & Nelson, 2005).

This study was unique in that it was genome-wide. Rather than focus on a particular set of genes known to be responsible for stress survival in a single condition, this study looked to open that up, studying the transcription levels of every gene found in the *O. tipulae* genome across three stresses and a control. The three abiotic stresses that were studied were heat, freezing, and desiccation stress, all of which *O. tipulae* is known to face within the soils of Kansas. This study provides an analysis of the transcriptional methods utilized by the *O. tipulae* genome in order to understand how these stress responses are regulated on a genome-wide level.

Chapter IV – Presence and Location of DNA Methylation in the Genomes of the Soil Nematodes *Oscheius tipulae* and *Caenorhabditis elegans*

In the fourth chapter of this dissertation, I used whole-genome bisulfite sequencing and MethylCap-Seq to determine and describe the role that DNA methylation plays in the regulation in the *O. tipulae* genome. DNA methylation, traditionally understood to occur on CpG islands within the genome, allows for the turning on and turning off of genes. By understanding the prevalence of DNA methylation within the *O. tipulae* genome, we can begin to extend our understanding on the physical properties that control and are responsible for not only the genetic regulation process as whole for this species, but also even more specifically for the transcription profiles for the abiotic stress responses detailed in Chapter III of this dissertation.

Historically, DNA methylation was thought to be absent in Nematoda as it has not been reported in *C. elegans*. This is further supported by the lack of DNA methyltransferase enzyme genes in the model organism (Wenzel, Palladino, & Jedrusik-Bode, 2011). Recent studies have begun to challenge this idea with cytosine methylation found in the parasitic species *Trichinella spiralis* (Gao et al., 2012) and adenine methylation being found in *C. elegans* (Greer et al., 2015). By utilizing newer research methods, this study can supplement these more recent findings and challenge the established beliefs regarding DNA methylation in nematodes.

References

- Adhikari, B. N., Wall, D. H., & Adams, B. J. (2009). Desiccation survival in an Antarctic nematode: molecular analysis using expressed sequenced tags. *BMC Genomics*, 10, 69. doi:10.1186/1471-2164-10-69
- Adhikari, B. N., Wall, D. H., & Adams, B. J. (2010). Effect of slow desiccation and freezing on gene transcription and stress survival of an Antarctic nematode. *Journal of Experimental Biology*, 213(11), 1803-1812. doi:10.1242/jeb.032268
- Ahn, I. Y., & Winter, C. E. (2006). The genome of *Oscheius tipulae*: determination of size, complexity, and structure by DNA reassociation using fluorescent dye. *Genome*, 49(8), 1007-1015. doi:10.1139/g06-045
- Ali, F., & Wharton, D. A. (2013). Cold tolerance abilities of two entomopathogenic nematodes, *Steinernema feltiae* and *Heterorhabditis bacteriophora*. *Cryobiology*, 66(1), 24-29. doi:10.1016/j.cryobiol.2012.10.004
- Alsheikh, M. K., Svensson, J. T., & Randall, S. K. (2005). Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. *Plant, Cell and Environment*, 28, 1114-1122.
- Anderson, R. V., & Coleman, D. C. (1982). Nematode Temperature Responses: A Niche Dimension in Populations of Bacterial-feeding Nematodes. *Journal of Nematology*, 14(1), 69-76.
- Apfeld, J., & Fontana, W. (2017). Age-Dependence and Aging-Dependence: Neuronal Loss and Lifespan in a *C. elegans* Model of Parkinson's Disease. *Biology (Basel)*, 7(1). doi:10.3390/biology7010001
- Argüelles, J. C. (1997). Thermotolerance and trehalose accumulation induced by heat shock in yeast cells of *Candida albicans*. *FEMS Microbiology Letters*, 146, 65-71.
- Beadle, G. W., & Tatum, E. L. (1941). Genetic control of biochemical reactions in *Neurospora*. *PNAS*, 27, 499-506.
- Behm, C. A. (1997). The Role of Trehalose in the Physiology of Nematodes. *International Journal for Parasitology*, 27(2), 215-229.
- Besnard, F., Koutsovoulos, G., Dieudonne, S., Blaxter, M., & Felix, M. A. (2017). Toward Universal Forward Genetics: Using a Draft Genome Sequence of the Nematode *Oscheius tipulae* To Identify Mutations Affecting Vulva Development. *Genetics*, 206(4), 1747-1761. doi:10.1534/genetics.117.203521
- Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., . . . Kim, S. K. (2002). A global analysis of *Caenorhabditis elegans* operons. *Nature*, 417(6891), 851-854. doi:10.1038/nature00830

- Borgonie, G., Garcia-Moyano, A., Litthauer, D., Bert, W., Bester, A., van Heerden, E., . . . Onstott, T. C. (2011). Nematoda from the terrestrial deep subsurface of South Africa. *Nature*, 474(7349), 79-82. doi:10.1038/nature09974
- Bravo, F. V., Da Silva, J., Chan, R. B., Di Paolo, G., Teixeira-Castro, A., & Oliveira, T. G. (2018). Phospholipase D functional ablation has a protective effect in an Alzheimer's disease Caenorhabditis elegans model. *Sci Rep*, 8(1), 3540. doi:10.1038/s41598-018-21918-5
- Brinkman, A. B., Simmer, F., Ma, K., Kaan, A., Zhu, J., & Stunnenberg, H. G. (2010). Whole-genome DNA methylation profiling using MethylCap-seq. *Methods*, 52(3), 232-236. doi:10.1016/j.ymeth.2010.06.012
- Brownlee, G. G., Sanger, F., & Barrell, B. G. (1967). Nucleotide Sequence of 5S-ribosomal RNA from Escherichia coli. *Nature*, 215, 735.
- Bulman, A. L., & Nelson, H. C. (2005). Role of trehalose and heat in the structure of the C-terminal activation domain of the heat shock transcription factor. *Proteins*, 58(4), 826-835. doi:10.1002/prot.20371
- Cheverud, J. M. (1996). Developmental Integration and the Evolution of Pleiotropy. *American Zoologist*, 36, 44-50.
- Consortium, T. C. e. S. (1998). Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology. *Science*, 282(5396), 2012-2018. doi:10.1126/science.282.5396.2012
- Crowe, J. H., Crowe, L. M., & Chapman, D. (1984). Preservation of membranes in anhydrobiotic organisms - the role of trehalose. *Science*, 223(4637), 701-703.
- Danovaro, R., Gambi, C., Dell'Anno, A., Corinaldesi, C., Fraschetti, S., Vanreusel, A., . . . Gooday, A. J. (2008). Exponential decline of deep-sea ecosystem functioning linked to benthic biodiversity loss. *Current Biology*, 18(1), 1-8. doi:10.1016/j.cub.2007.11.056
- Demeure, Y., Freckman, D. W., & Van Gundy, S. D. (1979). Anhydrobiotic Coiling of Nematodes in Soil. *Journal of Nematology*, 11(2), 189-195.
- Desikan, R., Niell, S. J., & Hancock, J. T. (2000). Hydrogen Peroxide-Induced Gene Expression in Arabidopsis thaliana. *Free Radical Biology & Medicine*, 28(5).
- Dichtel-Danjoy, M.-L., & Félix, M.-A. (2004). The two steps of vulval induction in Oscheius tipulae CEW1 recruit common regulators including a MEK kinase. *Developmental Biology*, 265(1), 113-126. doi:10.1016/j.ydbio.2003.09.010
- Dieterich, C., Clifton, S. W., Schuster, L. N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., . . . Sommer, R. J. (2008). The Pristionchus pacificus genome

provides a unique perspective on nematode lifestyle and parasitism. *Nature Genetics*, 40(10), 1193-1198. doi:10.1038/ng.227

- Eckl, J., Sima, S., Marcus, K., Lindemann, C., & Richter, K. (2017). Hsp90-downregulation influences the heat-shock response, innate immune response and onset of oocyte development in nematodes. *PLoS One*, 12(10), e0186386. doi:10.1371/journal.pone.0186386
- Evans, D., Zorio, D., MacMorris, M., Winter, C. E., Lea, K., & Blumenthal, T. (1997). Operons and SL2 trans-splicing exist in nematodes outside the genus *Caenorhabditis*. *PNAS*, 94, 9751-9756.
- Félix, M. A., & Sternberg, P. W. (1997). Two nested gonadal inductions of the vulva in nematodes. *Development*, 124, 253-259
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(19), 806-811.
- Gal, T. Z., Glazer, I., & Koltai, H. (2004). An LEA group 3 family member is involved in survival of *C. elegans* during exposure to stress. *FEBS Letters*, 577(1-2), 21-26. doi:10.1016/j.febslet.2004.09.049
- Gama Sosa, M. A., De Gasperi, R., & Elder, G. A. (2012). Modeling human neurodegenerative diseases in transgenic systems. *Hum Genet*, 131(4), 535-563. doi:10.1007/s00439-011-1119-1
- Gao, F., Liu, X., Wu, X.-P., Wang, X.-L., Gong, D., Lu, H., . . . Liu, M. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. *Genome Biology*, 13(R100), 1-13.
- Ghedini, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., . . . Scott, A. L. (2007). Draft Genome of the Filarial Nematode Parasite *Brugia malayi*. *Science*, 317, 1756-1760.
- Goyal, K., Pinelli, C., Maslen, S. L., Rastogi, R. K., Stephens, E., & Tunnacliffe, A. (2005). Dehydration-regulated processing of late embryogenesis abundant protein in a desiccation-tolerant nematode. *FEBS Lett*, 579(19), 4093-4098. doi:10.1016/j.febslet.2005.06.036
- Goyal, K., Walton, L. J., & Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. *Biochemical Journal*, 388, 151-157.
- Greer, E. L., Blanco, M. A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corrales, D., . . . Shi, Y. (2015). DNA Methylation on N6-Adenine in *C. elegans*. *Cell*, 161(4), 868-878. doi:10.1016/j.cell.2015.04.005

- Grelet, J., Benamar, A., Teyssier, E., Avelange-Macherel, M. H., Grunwald, D., & Macherel, D. (2005). Identification in pea seed mitochondria of a late-embryogenesis abundant protein able to protect enzymes from drying. *Plant Physiol*, *137*(1), 157-167. doi:10.1104/pp.104.052480
- Grewal, P. S., Gaugler, R., & Wang, Y. (1996). Enhanced cold tolerance of the entomopathogenic nematode *Steinernema feltiae* through genetic selection. *Annals of Applied Biology*, *129*, 335-341.
- Hara, M., Fujinaga, M., & Kuboi, T. (2004). Radical scavenging activity and oxidative modification of citrus dehydrin. *Plant Physiol Biochem*, *42*(7-8), 657-662. doi:10.1016/j.plaphy.2004.06.004
- Hara, M., Fujinaga, M., & Kuboi, T. (2005). Metal binding by citrus dehydrin with histidine-rich domains. *J Exp Bot*, *56*(420), 2695-2703. doi:10.1093/jxb/eri262
- Hendrich, B., & Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends in Genetics*, *19*(5), 269-277. doi:10.1016/s0168-9525(03)00080-5
- Hendricks, G., & Mylonakis, E. (2017). Expanding the nematode model system: The molecular basis of inflammation and infection recovery in *C. elegans*. *Virulence*, *8*(3), 244-245. doi:10.1080/21505594.2016.1239011
- Heyen, B. J., Alsheikh, M. K., Smith, E. A., Torvik, C. F., Seals, D. F., & Randall, S. K. (2002). The calcium-binding activity of a vacuole-associated, dehydrin-like protein is regulated by phosphorylation. *Plant Physiol*, *130*(2), 675-687. doi:10.1104/pp.002550
- Higa, L. M., & Womersley, C. Z. (1993). New insights into the anhydrobiotic phenomenon: The effects of trehalose content and differential rates of evaporative water loss on the survival of *Aphelenchus avenae*. *Journal of Experimental Zoology*, *267*(2), 120-129.
- Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., . . . Zamir, A. (1965). Structure of a Ribonucleic Acid. *Science*, *147*, 1462-1465.
- Holmstrup, M., Bayley, M., & Ramløv, H. (2002). Supercool or dehydrate? An experimental analysis of overwintering strategies in small permeable arctic invertebrates. *PNAS*, *99*(8), 5716-5720.
- Honda, Y., Tanaka, M., & Honda, S. (2010). Trehalose extends longevity in the nematode *Caenorhabditis elegans*. *Aging Cell*, *9*, 558-569.
- Hottiger, T., Boller, T., & Wiemken, A. (1987). Rapid changes of heat and desiccation tolerance correlated with changes of trehalose content in *Saccharomyces cerevisiae* cells subjected to temperature shifts. *FEBS Letters*, *220*(1), 113-115.

- Hottiger, T., de Virgilio, C., Hall, M. N., Boller, T., & Wiemken, A. (1994). The role of trehalose synthesis for the acquisition of thermotolerance in yeast. *European Journal of Biochemistry*, 219, 187-193.
- Illumina. (2010). *Illumina Sequencing Technology* (770-2007-002). Retrieved from San Diego, CA:
- Illumina. (2014). *Whole-Genome Bisulfite Sequencing* (470-2014-004). Retrieved from
- Ingham, R. E., Trofymow, J. A., Ingham, E. R., & Coleman, D. C. (1985). Interactions of Bacteria, Fungi, and their Nematode Grazers: Effects on Nutrient Cycling and Plant Growth. *Ecological Monographs*, 55(1), 119-140. doi:10.2307/1942528
- Jagdale, G. B., & Grewal, P. S. (2003). Acclimation of entomopathogenic nematodes to novel temperatures: trehalose accumulation and the acquisition of thermotolerance. *International Journal for Parasitology*, 33(2), 145-152. doi:10.1016/s0020-7519(02)00257-6
- Jagdale, G. B., Grewal, P. S., & Salminen, S. O. (2005). Both Heat-Shock and Cold-Shock Influence Trehalose Metabolism in an Entomopathogenic Nematode. *The Journal of Parasitology*, 95(5), 988-994.
- Joo, H. J., Park, S., Kim, K. Y., Kim, M. Y., Kim, H., Park, D., & Paik, Y. K. (2016). HSF-1 is involved in regulation of ascarioside pheromone biosynthesis by heat stress in *Caenorhabditis elegans*. *Biochem J*, 473(6), 789-796. doi:10.1042/BJ20150938
- Kimble, J., & Hirsh, D. (1979). The Postembryonic Cell Lineages of the Hermaphrodite and Male Gonads in *Caenorhabditis elegans*. *Developmental Biology*, 70, 396-417.
- Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R. J., & Wilson, G. G. (1994). The DNA (cytosine-5) methyltransferases. *Nucleic Acids Research*, 22(1), 1-10.
- Leung, M. C., Williams, P. L., Benedetto, A., Au, C., Helmcke, K. J., Aschner, M., & Meyer, J. N. (2008). *Caenorhabditis elegans*: an emerging model in biomedical and environmental toxicology. *Toxicol Sci*, 106(1), 5-28. doi:10.1093/toxsci/kfn121
- Lillie, S. H., & Pringle, J. R. (1980). Reserve Carbohydrate Metabolism in *Saccharomyces cerevisiae*: Responses to Nutrient Limitation. *Journal of Bacteriology*, 143(3), 1384-1394.
- Louvet-Vallee, S., Kolotuev, I., Podbilewicz, B., & Felix, M. A. (2003). Control of Vulval Competence and Centering in the Nematode *Oscheius* sp. 1 CEW1. *Genetics*, 163, 133-146.

- Margesin, R., Neuner, G., & Storey, K. B. (2007). Cold-loving microbes, plants, and animals--fundamental and applied aspects. *Naturwissenschaften*, 94(2), 77-99. doi:10.1007/s00114-006-0162-6
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., . . . Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), 376-380. doi:10.1038/nature03959
- Martinez, J., Perez-Serrano, J., Bernadina, W. E., & Rodriguez-Caabeiro, F. (2001). Stress response to cold in *Trichinella* species. *Cryobiology*, 43(4), 293-302. doi:10.1006/cryo.2001.2363
- Mazur, P. (1984). Freezing of living cells: mechanisms and implications. *American Journal of Physiology - Cell Physiology*, 247, C125-142.
- Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, 95(6), 315-327. doi:10.1016/j.ygeno.2010.03.001
- Orr, H. A. (2000). Adaption and the Cost of Complexity. *Evolution*, 54(1), 13-20.
- Pandey, S., Tiwari, S., Kumar, A., Niranjana, A., Chand, J., Lehri, A., & Chauhan, P. S. (2018). Antioxidant and anti-aging potential of Juniper berry (*Juniperus communis* L.) essential oil in *Caenorhabditis elegans* model system. *Industrial Crops and Products*, 120, 113-122. doi:10.1016/j.indcrop.2018.04.066
- Panitz, D., Swamy, H., & Nehrke, K. (2015). A *C. elegans* model of electronic cigarette use: Physiological effects of e-liquids in nematodes. *BMC Pharmacol Toxicol*, 16, 32. doi:10.1186/s40360-015-0030-0
- Perry, R. N., & Wharton, D. A. (2011). Preface. In R. N. Perry & D. A. Wharton (Eds.), *Molecular and Physiological Basis of Nematode Survival*: CAB International.
- Potts, M. (1994). Desiccation Tolerance of Prokaryotes. *Microbiological Reviews*, 58(4), 755-805.
- Ramløv, H. (2000). Aspects of natural cold tolerance in ectothermic animals. *Human Reproduction*, 15, 26-46.
- Rankin, C. H., Beck, C. D. O., & Chiba, C. M. (1990). *Caenorhabditis elegans*: a new model system for the study of learning and memory. *Behavioural Brain Research*, 37, 89-92.
- Raymond, M. R., & Wharton, D. A. (2013). The ability of the Antarctic nematode *Panagrolaimus davidi* to survive intracellular freezing is dependent upon nutritional status. *Journal of Comparative Physiology B*, 183(2), 181-188. doi:10.1007/s00360-012-0697-0

- Ross, C. A., & Poirier, M. A. (2004). Protein aggregation and neurodegenerative disease. *Nat Med*, *10 Suppl*, S10-17. doi:10.1038/nm1066
- Sampuda, K. M., Riley, M., & Boyd, L. (2017). Stress induced nuclear granules form in response to accumulation of misfolded proteins in *Caenorhabditis elegans*. *BMC Cell Biol*, *18*(1), 18. doi:10.1186/s12860-017-0136-x
- Sanchez-Ballesta, M. T., Rodrigo, M. J., Lafuente, M. T., Granell, A., & Zacarias, L. (2004). Dehydrin from Citrus, Which Confers in Vitro Dehydration and Freezing Protection Activity, Is Constitutive and Highly Expressed in the Flavedo of Fruit but Responsive to Cold and Water Stress in Leaves. *Journal of Agricultural and Food Chemistry*, *52*, 1950–1957.
- Sanger, F., & Nicklen, S. C., A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Biochemistry*, *74*(12), 5463-5467.
- Shen, P., Yue, Y., Zheng, J., & Park, Y. (2018). *Caenorhabditis elegans*: A Convenient In Vivo Model for Assessing the Impact of Food Bioactive Compounds on Obesity, Aging, and Alzheimer's Disease. *Annu Rev Food Sci Technol*, *9*, 1-22. doi:10.1146/annurev-food-030117-012709
- Singer, M. A., & Lindquist, S. (1998a). Multiple Effects of Trehalose on Protein Folding In Vitro and In Vivo. *Molecular Cell*, *1*, 639-648.
- Singer, M. A., & Lindquist, S. (1998b). Thermotolerance in *Saccharomyces cerevisiae*: the Yin and Yang of trehalose. *Trends in Biotechnology*, *16*, 460-468.
- Solomon, A., Salomon, R., Paperna, I., & Glazer, I. (2000). Desiccation stress of entomopathogenic nematodes induces the accumulation of a novel heat--stable protein. *Parasitology*, *121*, 409-416.
- Sommer, R. J. (2005). Evolution of development in nematodes related to *C. elegans*. *WormBook*, 1-17. doi:10.1895/wormbook.1.46.1
- Sulston, J. E., Schierenberg, E., White, J. G., & Thomson, J. N. (1983). The Embryonic Cell Lineage of the Nematode *Caenorhabditis elegans*. *Developmental Biology*, *100*, 64-119.
- Torrini, G., Mazza, G., Strangi, A., Barabaschi, D., Landi, S., Mori, E., . . . Roversi, P. F. (2016). *Oscheius tipulae* in Italy: Evidence of an Alien Isolate in the Integral Natural Reserve of Montecristo Island (Tuscany). *Journal of Nematology*, *48*(1), 8-13.
- Tweedie, S., Charlton, J., Clark, V., & Bird, A. (1997). Methylation of Genomes and Genes at the Invertebrate-Vertebrate Boundary. *MOLECULAR AND CELLULAR BIOLOGY*, *17*(3), 1469-1475.

- Tyson, T., Reardon, W., Browne, J. A., & Burnell, A. M. (2007). Gene induction by desiccation stress in the entomopathogenic nematode *Steinernema carpocapsae* reveals parallels with drought tolerance mechanisms in plants. *International Journal for Parasitology*, *37*(7), 763-776. doi:10.1016/j.ijpara.2006.12.015
- Waddington, C. H. (1942). The Epigenotype. *Endeavor*, 18-20.
- Wagner, G. P. (1996). Homologues, Natural Kinds and the Evolution of Modularity. *American Zoologist*, *36*, 36-43.
- Wang, Z., Liao, B. Y., & Zhang, J. (2010). Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences*, *107*(42), 18034-18039. doi:10.1073/pnas.1004666107
- Watson, J. D., & Crick, F. H. (1953). Molecular Structure of Nucleic Acids. *Nature*, *171*, 737-738.
- Welch, J. J., & Waxman, D. (2003). Modularity and the cost of complexity. *Evolution*, *57*(8), 1723-1734.
- Welch, W. J., & Brown, C. R. (1996). Influence of molecular and chemical chaperones on protein folding. *Cell Stress & Chaperones*, *1*(2), 109-115.
- Wenzel, D., Palladino, F., & Jedrusik-Bode, M. (2011). Epigenetics in *C. elegans*: facts and challenges. *Genesis*, *49*(8), 647-661. doi:10.1002/dvg.20762
- Wharton, D. A. (2003). Freezing survival and cryoprotective dehydration as cold tolerance mechanisms in the Antarctic nematode *Panagrolaimus davidi*. *Journal of Experimental Biology*, *206*(2), 215-221. doi:10.1242/jeb.00083
- Wharton, D. A. (2011). Cold Tolerance. In R. N. Perry & D. A. Wharton (Eds.), *Molecular and Physiological Basis of Nematode Survival*: CAB International.
- Wharton, D. A., & Worland, M. R. (1998). Ice Nucleation Activity in the Freezing-Tolerant Antarctic Nematode *Panagrolaimus davidi*. *Cryobiology*, *36*, 279-286.
- Wise, M. J., & Tunnacliffe, A. (2004). POPP the question: what do LEA proteins do? *Trends Plant Sci*, *9*(1), 13-17. doi:10.1016/j.tplants.2003.10.012
- Womersley, C. Z., & Higa, L. M. (1998). Trehalose: Its Role in the Anhydrobiotic Survival of *Ditylenchus Myceliophagus*. *Nematologica*, *44*(3), 269-281.
- Zeng, Y., Guo, W., Xu, G., Wang, Q., Feng, L., Long, S., . . . Pei, Z. (2016). Xyloketal-derived small molecules show protective effect by decreasing mutant Huntingtin protein aggregates in *Caenorhabditis elegans* model of Huntington's disease. *Drug Des Devel Ther*, *10*, 1443-1451. doi:10.2147/DDDT.S94666

Zhang, C., & Guy, C. L. (2006). In vitro evidence of Hsc70 functioning as a molecular chaperone during cold stress. *Plant Physiol Biochem*, 44(11-12), 844-850. doi:10.1016/j.plaphy.2006.09.012

Zhang, Z.-Q. (2013). Animal biodiversity: An update of classification and diversity in 2013. *Zootaxa*, 3703(1), 5-11. doi:10.11646/zootaxa.3703.1.3

CHAPTER II
DRAFT GENOME OF THE KJO STRAIN OF THE SOIL NEMATODE
OSCHEIUS TIPULAE

Abstract

Oscheius tipulae is a species of free-living nematode that is found in soils throughout the world. These microscopic roundworms cycle nutrients in large amounts to enable plant growth. Previous studies have provided estimates of the size and composition of the *O. tipulae* genome, but no extensive sequencing has been done to assemble it. In this study, the *O. tipulae* genome was sequenced using Illumina high-throughput sequencing using paired-end and long-insert library preparation. Assembly was completed using a variety of bioinformatics programs. This provided an assembled genome with a size of approximately 60 Mb and approximately 20,000 protein-coding genes. The assembled *O. tipulae* and *C. elegans* genomes were directly compared and found to vary in both structure and genetic content. This was further supported by a comparison of each species' protein-coding gene sequences. This study adds to the existing nematode genomics library, providing an extensive assembly in the *Oscheius* genus, as well as begins the process of establishing *O. tipulae* as a satellite model organism to *C. elegans*.

Introduction

Free-living soil nematodes are found in incredibly high numbers on a global scale, and they play an important role in nutrient cycling and plant health. Nematodes digest inorganic nitrogen compounds from the bacteria they consume and excrete the excess nitrogen as ammonia, which can then be taken in and utilized by plants (Ingham, Trofymow, Ingham, & Coleman, 1985). The most commonly studied species of free-living soil nematode is *Caenorhabditis elegans*. *C. elegans* has been used as a model organism for microinvertebrate genetics, development, and behavior. The complete genome of *C. elegans* has been sequenced by the *C. elegans* Sequencing Consortium and is used as a model for both nematode genomes and microinvertebrate genomes (1998).

Oscheius tipulae is another species of free-living soil nematode, and it is found within the same taxonomic family, Rhabditidae, as *C. elegans*. Both species most commonly reproduce as self-fertilizing hermaphrodites and can be easily grown in culture in a laboratory setting. While *O. tipulae* has not been as extensively studied as *C. elegans*, it has been used in studies on vulva development and in various comparative studies across nematode species (Louvet-Vallee, Kolotuev, Podbilewicz, & Felix, 2003; Sommer & Sternberg, 1995). *C. elegans* shares a more recent common ancestor with *O. tipulae* than it does with any other commonly studied species of nematode outside the *Caenorhabditis* genus. Despite this, it has been shown that *O. tipulae* has much higher levels of genetic diversity than *C. elegans* and is more widespread in soils across the world (Baille, Barriere, & Felix, 2008).

While most molecular work done with *O. tipulae* has been at the level of individual genes, previous work has been done on the size and composition of the

genome of the CEW1 strain of *O. tipulae* which was established in Sao Paulo, Brazil in 1992. It is estimated to be similar in size to the *C. elegans* genome: six chromosomes, approximately 100 megabases (Mb), and composed of approximately 20,000 protein-coding genes (Ahn & Winter, 2006). Recently, a draft genome of the CEW1 strain of *O. tipulae* has been sequenced, and the statistics achieved by the project did not fully match those initially estimated. The 191 scaffolds that were assembled only comprised a total length of approximately 59 Mb, and only slightly less than 15,000 protein-coding genes were detected (Besnard, Koutsovoulos, Dieudonne, Blaxter, & Felix, 2017).

There is currently room for work to supplement this data. A second strain of *O. tipulae* has been reared under laboratory settings. This strain is the KJO strain and was originally obtained from the Konza Prairie in Manhattan, Kansas. In this study, both the genome and transcriptome of the KJO strain of *O. tipulae* were sequenced and annotated using Illumina high-throughput sequencing and various bioinformatics techniques. The data obtained from this sequencing project will provide in-depth detail and knowledge on the genomic profile of the strain. This will be used to complement previous sequencing work done with the species, as well as allow for further comparisons between *O. tipulae* and other nematode species. In turn, this will further establish the role of *O. tipulae* as a satellite model organism of *C. elegans*.

Materials and Methods

Nematode Preparation

Nematodes of the KJO strain of *O. tipulae* were grown at 18°C on 60x15 mm non-vented dishes of Nematode Growth Media and seeded with the OP50 strain of *Escherichia coli* as described by Brenner (1974). Once the bacteria had been completely

eaten, the nematodes were washed from plates into 15 mL conical tubes with M9 Buffer. These tubes were stored at 6°C for 1-7 days, allowing time for the remaining bacteria within the digestive tract of the nematodes to be digested. Nematode samples were then pooled into one sample in order to insure a high enough concentration of DNA would be extracted for sequencing.

Genomic DNA Extraction and Genome Assembly

Genomic DNA was extracted from the collected nematode sample using the PureLink DNA Genomic Mini Kit (Life Technologies). In order to ensure quality standards needed for sequencing, the sample was visualized via gel electrophoresis on a 1% agarose gel. The concentration and optical density of the sample were then obtained via a NanoDrop ND-1000 spectrophotometer (NanoDrop products, Wilmington, DE) ($A_{260}/A_{280} > 1.8$). One DNA sample was then sent to the University of Missouri Core DNA Facility for library preparation via the Illumina's Genomic PCR-free Library Preparation (TruSeq). DNA concentration was externally confirmed via Qubit Fluorometric Quantitation (Thermo Fisher Scientific) before undergoing high-throughput sequencing via Illumina HiSeq 2000.

This resulted in 100-bp paired-end reads which were then pooled, trimmed for quality using default parameters, and used for the *de novo* assembly into contigs using CLC Genomics Workbench 9 (CLC bio, Cambridge, MA, USA). This assembly was done using an automatic estimated bubble size of 50, a word size of 64, and an auto-detected estimated paired distance of 185-593. Scaffolding was automatically performed during the assembly, and the minimum contig size was set to 20 bp. Mismatch, insertion,

and deletion costs were 2, 3, and 3, respectively. A length fraction of 0.5 and a similarity fraction of 0.98 were used.

Contigs were then BLAST-searched (blastn) to the *E. coli* genome. Contigs that matched with an E-value less than 0.01 along the entirety of its length were assumed to belong to *E. coli* and not *O. tipulae* and were removed from the analysis. This blastn had an expected E-value of 0.01 and a word size of 11. Match/mismatch scores were set to 2/-3, and existence/extension gap costs were set to 5/2.

A second DNA sample was obtained using the same methods as previously detailed and sent to MOgene (Kansas City, Missouri) for Nextera mate-pair 4-kb and 10-kb insert library preparation. High-throughput sequencing was then performed via Illumina HiSeq 2500 at the University of Missouri Core DNA Facility, producing 50-bp paired-end reads. These long-insert reads were then used to scaffold together, with extension, the previously assembled contigs using SSPACE (Boetzer, Henkel, Jansen, Butler, & Pirovano, 2011). The 4-kb inserts were used first to scaffold the assembled contigs, and the 10-kb inserts were used to assemble the results of the 4-kb scaffolding. A 25% error in insert sizes was used in both analyses, and default settings were used for the remaining parameters. GapFiller was then used to clarify any unknown bases that remained in the scaffolds (Boetzer & Pirovano, 2012). Default parameters were used, except for the following changes: maximum difference between the gap size and the number of closed nucleotides was set to 100, the number of reads required to trim off the start was set to 5, and the minimum contig overlap to merge sequences was set to 30.

The quality of this assembly was assessed using two programs: BUSCO v2.0.1 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015; Waterhouse et al., 2017)

and QUAST v4.3 (Gurevich, Saveliev, Vyahhi, & Tesler, 2013). The BUSCO assessment was done using the metazoa_odb9 Nematoda lineage in genome mode. The QUAST assessment used the known *C. elegans* genome assembly as a reference and the remaining parameters were set at the defaults.

Mitochondrial Genome

Prior to quality assessment, the assembled scaffolds were compared to nineteen known nematode mitochondrial sequences using blastn in order to locate the KJO mitochondrial genome. The parameters for this were the default parameters utilized by the blastn executable function. A majority of the genomes were from the *Caenorhabditis* genus, meaning they were fairly similar taxonomically. Once it was shown that only one of the *O. tipulae* scaffolds matched, and that it matched to all nineteen of the other mitochondrial sequences, it was removed from the *O. tipulae* draft genome assembly for further analysis.

Based on the blast results, the putative mitochondrial scaffold needed to be rearranged. Using the coordinates and their relative order in the *C. elegans* mitochondrial genome, the scaffold was split into four parts and manually pieced together in the proper order. Alignments were then performed in Geneious 11.1.4 (<https://www.geneious.com>) in order to assess the accuracy of the manually reassembled mitochondrial genome when compared to the original assembled scaffold. The alignments were done using the Geneious aligner, creating a global alignment with free end gaps. The cost matrix was 65% similarity with a gap open penalty of 10 and extension penalty of 3. In these alignments, the original scaffold and the reassembled sequence were each aligned to the

known *C. elegans*, *C. brenneri*, *C. briggsae*, and *Litoditis aff. marina PmIII* mitochondrial sequences.

Synteny

A large-scale synteny analysis was performed using SyMAP v4.2 from the University of Arizona (Soderlund, Nelson, Shoemaker, & Paterson, 2006). The 100 largest assembled scaffolds were mapped to the six *C. elegans* chromosomes. The nucleotide sequences were used for each, and default parameters were used, including an anchor value of 7. The same analysis was performed comparing the 100 largest assembled KJO scaffolds to the 25 largest CEW1 scaffolds.

To further assess synteny between the KJO and CEW1 genomes, the assembled KJO scaffolds were aligned to the 191 CEW1 scaffolds via MUMmer v3.23 (Kurtz et al., 2004). In MUMmer, the nucmer program was run under default parameters. In order to detect SNPs, the show-snps program was used, excluding indels and SNPs from alignments with ambiguous mapping.

Genome Annotation

Detection and annotation of interspersed repeats was performed using RepeatMasker v4.0.7 (Smit, Hubley, & Green, 2013-2015). The analysis was run on the assembled scaffolds using the following parameters: the NCBI search engine, *C. elegans* as the assumed species, and the slowest possible run speed to ensure the maximum amount of repeats were included in the annotation. This analysis also ignored low complexity and simple repeats, as these were located using Phobos. RNA pseudo genes were also left out of this analysis.

Location and annotation of tandem repeat sequences was performed using Phobos v3.3.12 (Mayer, 2006-2010). Analyses were run for the identification of both microsatellites (2-10 bp) and minisatellites (11-100 bp), and for each class of tandem repeat, default parameters were used.

Gene prediction on the assembled scaffolds was performed *de novo* via Augustus, using the default parameters (Stanke & Morgenstern, 2005). Gene annotation was done using Blast2GO (Götz et al., 2008). The predicted *O. tipulae* genes were compared to the amino acid sequences of the known *C. elegans* protein-coding genes via blastx. The following parameters were used: HSP length cutoff of 33, word size of 3, and E-value cutoff of 1E-3. Mapping and annotation processes were then ran in Blast2GO. An InterPro analysis was also performed in Blast2GO using the HMMPfam, SuperFamily, HMMPanther, and MobiDBLite applications.

Positive Selection

Complete gene lists of the following nematode species were obtained from NCBI for the analysis: *Caenorhabditis brenneri*, *Caenorhabditis briggsae*, *Caenorhabditis japonica*, *Caenorhabditis remanei*, and *Pristionchus pacificus*. For this analysis, the predicted KJO genes were also obtained and used for *O. tipulae*. These six species will be referred to as “target species”. The *Caenorhabditis elegans* gene sequences were also obtained. A series of blastx analysis were performed in CLC Genomics Workbench 10 (CLC bio, Cambridge, MA, USA) using the following parameters: E-value cutoff of 10.0, word size of 3, maximum number of hits as 10, the matrix BLOSUM62, an existence gap cost of 11, and an extension gap cost of 1. In order to find reciprocal blast results, the nucleotide sequences of the *C. elegans* genes were BLAST-searched to the amino acid

genes sequences for each of the other obtained nematode species. The inverse of this was then performed. The nucleotide sequences for the six other species were BLAST-searched to the *C. elegans* amino acid gene sequences.

The blastx comparisons were then separated by the species involved and compared to one another. In a given pair of comparisons, if a *C. elegans* gene had a best hit within the target species gene list, and that same target species gene landed the same *C. elegans* gene as its top hit, the two genes were deemed to be reciprocal orthologs and remained in the overall analysis. Genes without a reciprocal ortholog were removed. This was performed for all six comparisons between *C. elegans* and the target species. The six sets of reciprocal orthologs were then compared. If a *C. elegans* gene had reciprocal orthologs in all six comparisons, it remained. Genes that did not have reciprocal orthologs in all six comparisons were removed.

Sets of orthologs, each containing a sequence from all seven species, were then aligned using MUSCLE under the default parameters (Edgar, 2004). In order for further analysis to continue, the aligned sequences for each gene were trimmed at the end so that their total length was a multiple of three. Phylogenetic trees were then created for each gene set using RAxML version 8.2.11 (Stamatakis, 2014). The GTRCAT substitution model was used, while the remaining parameters remained at the default. Positive selection testing was performed using the codeml function within the PAML version 4.9d package of programs (Yang, 2007). The parameters remained at their default. From these results, genes with dN/dS ratios greater than 1 were deemed to be positively selected, and genes with dN/dS ratios less than 1 were not.

Results and Discussion

Genome Assembly

Initial assembly of the 100 bp paired-end reads resulted in 2,000 contigs, which was then reduced to 1,904 once those that had a strong match to the *E. coli* genome were removed. After scaffolding and nucleotide clarification via SSPACE and GapFiller, respectively, the assembly was comprised of 1,508 scaffolds, and 1 of these was determined to be the mitochondrial genome and removed. This resulted in an assembly made up of 1,507 scaffolds of at least 200 bp in length.

The QUAST assessment indicated an N50 scaffold length of 163,494, and the total genome length was 60,646,666 bp. This size is smaller than the 100 Mb estimate from Ahn and Winter (2006). This number is also shorter than most when comparing it across other sequenced nematode genomes (Table 2.1). The BUSCO analysis for genome completion determined the genome is 89.9% complete as it contained 828 complete and 37 fragmented BUSCOs from a total of 982, providing evidence that the assembly is still fairly complete for its smaller size. Additional assembly statistics and their comparisons with other assembled nematode genomes can be found in Table 2.1.

Genome Annotation

The RepeatMasker analysis found 1,112 interspersed repeat elements totaling 231,657 bp. This total makes up 0.38% of the total assembled genome. A majority of the repeat elements were DNA transposons, which were located 636 times. There were 333 retroelements found, and the remaining 143 interspersed repeats were unclassified. Of the DNA transposons, a large majority (597) were a part of the Tc1-IS630-Pogo family. The 333 retroelements were broken down into long interspersed nuclear elements (LINEs)

Table 2.1 Genome assembly statistics for this *Oscheius tipulae* genome and various other nematode species.

Species	Genome size (~ Mb)	GC content (%)	Protein-coding genes (count)	Completeness (BUSCO %)
<i>Brugia malayi</i> ^a	88	30.5	11,515	96.7
<i>Caenorhabditis briggsae</i> ^b	108	37.4	22,405	97.7
<i>Caenorhabditis elegans</i> ^c	100	36.0	20,239	98.6
<i>Dictyocaulosis viviparous</i> ^d	161	34.8	14,171	71.2
<i>Heterorhabditis bacteriophora</i> ^e	77	32.2	21,250	87.1
<i>Oscheius tipulae</i>	60	44.5	20,402	89.9
<i>Pristionchus pacificus</i> ^f	169	43	29,201	90.8

a. Ghedin et al., 2007
b. Stein et al., 2003
c. C elegans Sequencing Consortium, 1998
d. McNulty et al., 2016
e. Bai et al., 2013
f. Dieterich et al., 2008

and long-terminal repeat (LTR) elements. These were found 256 and 77 times, respectively.

This estimate of 0.38% of the *O. tipulae* genome being made up of interspersed repeats, including transposable elements, is fairly low compared to other nematode species. The *C. elegans* genome is made up of an estimated 12% transposable elements (Sijen & Plasterk, 2003). It has been shown that transposable element detection based on homology is underestimated in the assemblies of new genomes when the phylogenetic distance between the two species is great (Platt, Blanco-Berdugo, & Ray, 2016). *O. tipulae* and *C. elegans* are within the same taxonomic family, but as mentioned further, their genomes lack intensive homology and have great molecular divergence. This could be leading to a vast underrepresentation of transposable elements in the analysis.

The Phobos analysis detected 133,368 tandem repeats. Of these, 132,418 of them were microsatellite (2-10 bp) loci, amassing to 1,642,662 bp in length. The remaining 950 repeats were minisatellite (11-100 bp) loci, totaling 69,651 bp in length. With a combined length of 1,712,313 bp, tandem repeat sequences make up 2.82% of the assembled genome. This falls nearly perfectly in line with the estimate from the *C. elegans* genome where it is estimated that tandem repeats make up 2.7% of the genome (The *C. elegans* Sequencing Consortium, 2008). This estimate shows that a very small portion of this genome is made up of by tandem repeats.

Gene Content

Using the Augustus gene prediction software, it is predicted that there are 20,402 protein-coding genes within this *O. tipulae* genome assembly, falling in line with previous estimations (Ahn & Winter, 2006). In total, these genes are comprised of

44,957,717 bp, meaning the average gene size is approximately 2,204 bp and that genes comprise approximately 74.13% of the overall assembled genome. At 17,734,451 bp, the total intron content is 29.24% of the genome. This also means that 39.45% of the total gene content is intronic regions. Conversely, coding sequences (CDS) made up 22,842,686 bp of the total genome. This means approximately half (~50.8%) of the genetic content was made of CDS. This is also approximately 37.67% of the total genome content. Further detail regarding the annotation and genome content is provided by Table 2.2.

Mitochondrial Genome

The scaffold believed to contain the KJO mitochondrial genome contained a region of 4,297 bp of unknown nucleotide. This was believed to be an artifact of the paired-end reads during the original genome assembly. Upon the rearrangement, this region was removed from the scaffold. Prior to the rearrangement, the total length of the scaffold was 20,551 bp. After rearrangement, it was 13,185 bp. This new length falls much closer in line to other known nematode genomes, all of which are somewhere in the 13,000s or 14,000s.

The Geneious alignments of each resulted in similar levels of percent identity. The original scaffold had showed 53.2% identity with the *C. elegans* mitochondrial sequence, and the reassembled sequence had 53.1% identity. The important difference was that the original scaffold was unable to align across its entirety to the *C. elegans* sequence. Only approximately 10,000 bp were able to be aligned. In the alignment with the reassembled mitochondrial sequence, the alignment included the entire sequence, providing support for the fact that the reassembled sequence is the correct mitochondrial

Table 2.2 Genomic content makeup of the assembled KJO *O. tipulae* genome.

Genomic Feature	Total Content (bp)	Percent of Total Genome
Total size	60,646,666	-
Genes	44,957,717	74.1%
(CDS)	(22,842,686)	(37.7%)
(introns)	(17,734,451)	(29.2%)
Intersperced repeats	213,657	0.4%
Tandem repeats	1,712,313	2.8%
(microsatellites [2-10 bp])	(1,642,662)	(2.7%)
(minisatellites [11-100 bp])	(69,651)	(0.1%)

genome. This reassembled sequence was then used for further alignments, and the percent identities found in alignments to the *C. brenneri* and *C. briggsae* mitochondrial sequences were 52.6% and 76.8%, respectively. At 83.3%, the alignment with the highest percent identity was that with the nematode species *Litoditis aff. marina PmIII*. All of these also showed an alignment across the entire length of the mitochondrial sequence, further supporting this sequence as the true mitochondrial genome.

Synteny

The synteny analysis comparing the 100 largest assembled *O. tipulae* scaffolds with the six *C. elegans* chromosomes showed little evidence of any large-scale synteny or conserved gene order between the two genomes. A majority of the top largest *O. tipulae* scaffolds have no clear ortholog in any of the six the *C. elegans* chromosomes (Figure 2.1). Across nematode species, this is not uncommon, as multiple nematode genome assemblies have found similar results (Cotton et al., 2014; Ghedin et al., 2007; Jex et al., 2011). Further, this indicates that even within the same taxonomic family (Rhabditidae), nematode genomes are not structurally similar. This is supported by the fact that *C. elegans* and another member of the *Oscheius* genus, *O. myriophila*, are estimated to have greater molecular divergence than a human has with zebra fish (Kiontke et al., 2004).

The synteny analysis between the KJO and CEW1 strains of *O. tipulae* show much more genetic similarity than the KJO strain does with *C. elegans*. The same 100 KJO scaffolds from the *O. tipulae* analysis show much greater synteny with the CEW1 scaffolds, with a majority of the scaffolds having a specific match within at least one of the CEW1 scaffolds (Figure 2.2). This indicates strength in this genome sequencing, as it shows extensive similarities with a genome of the same species, but it also indicates there

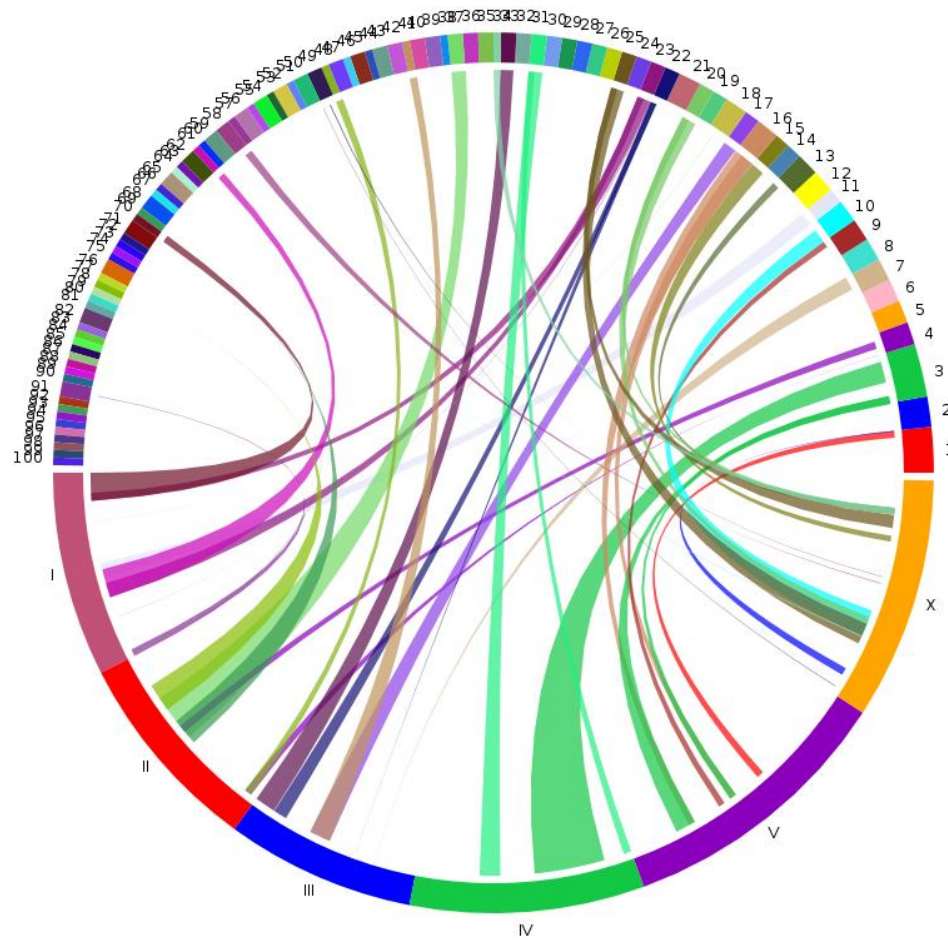


Figure 2.1 Synteny comparison between the *O. tipulae* and *C. elegans* genomes. The circle is comprised of the 100 largest assembled *O. tipulae* scaffolds (top half of circle) and the 6 *C. elegans* chromosomes (bottom half of circle). Bridges connecting sequences are similar nucleotide sequences. The more similarity, the higher the synteny. Small amounts of mapping indicate no large-scale synteny between the two species.

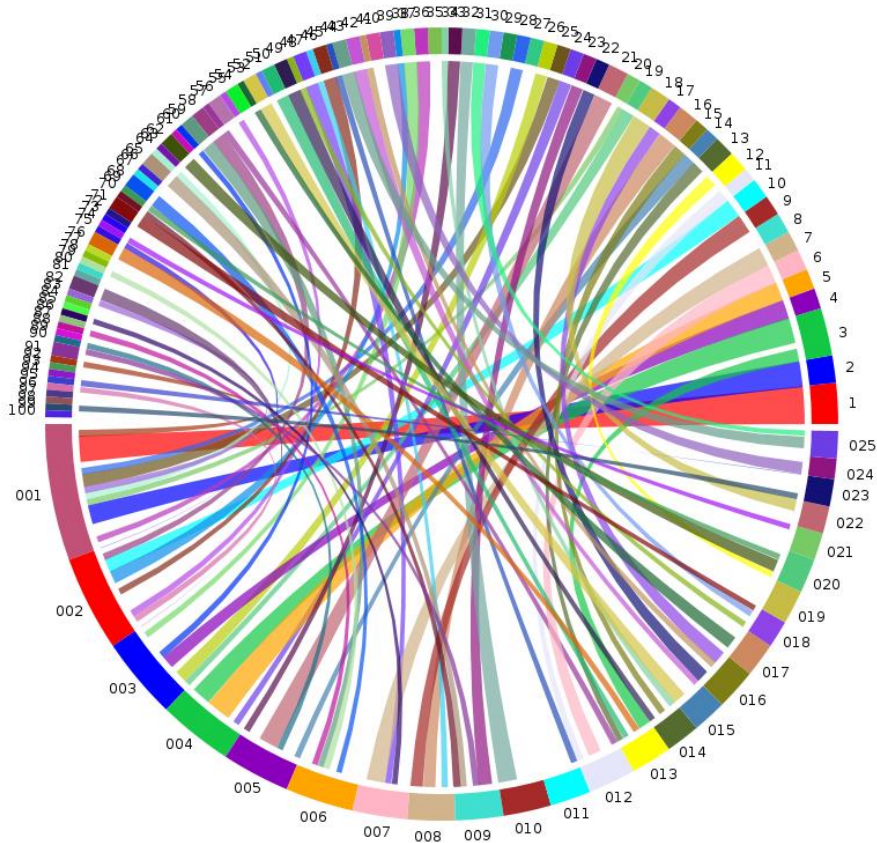


Figure 2.2 Synteny comparison between the KJO draft genome scaffolds and the CEW1 draft genome scaffolds. The circle is comprised of the 100 largest assembled *O. tipulae* KJO scaffolds (top half of circle) and the 25 largest *O. tipulae* CEW1 scaffolds (bottom half of circle). Bridges connecting sequences are similar nucleotide sequences. The more similarity, the higher the synteny. Extensive overlaps across the two genomes indicate high synteny, meaning high similarity of genomic sequences.

could be some genetic differences between the two strains as the similarities do not encompass the entirety of all scaffolds analyzed.

Excluding insertions and deletions, the MUMmer analysis showed there are 673,604 SNPs between the assembled KJO genome and the published CEW1 genome. This means, on average, there is slightly more than 1 SNP for every 100 bp in the KJO genome. As in the synteny analysis, this proves that the two strains have highly similar genome sequences, but it also proves that they are not identical.

Of the predicted 20,402 genes, 13,537 (66.4%) had a BLAST result match with a known *C. elegans* gene. Of those genes with a hit, 12,285 genes were able to match to at least one Gene Ontology (GO) term. The breakdown of the level 2 Biological Processes GO Terms of these annotated genes is incredibly similar to not only other nematode genomes, but also the genomes of other microinvertebrates (Figure 2.3). This also indicates that a *C. elegans* ortholog was not found for approximately one-third of the predicted protein-coding genes (6,865 in total). This provides further evidence to the idea that the *O. tipulae* genome lacks large-scale synteny with and varies greatly from that of *C. elegans*. While there is a large portion of genes that are conserved across these organisms, due to the high molecular divergence and early estimated evolutionary divergence, *O. tipulae* still looks to be highly unique in its genetic makeup.

Positive Selection

After isolating the genes that had reciprocal orthologs in all seven nematode species, only 3,810 genes remained. This is approximately 18.7% of the predicted *O. tipulae* genes. This means that of the ~20,000 genes in each of the *C. elegans* and *O. tipulae* genomes, less than 20% of them were found to have reliable orthologs in these

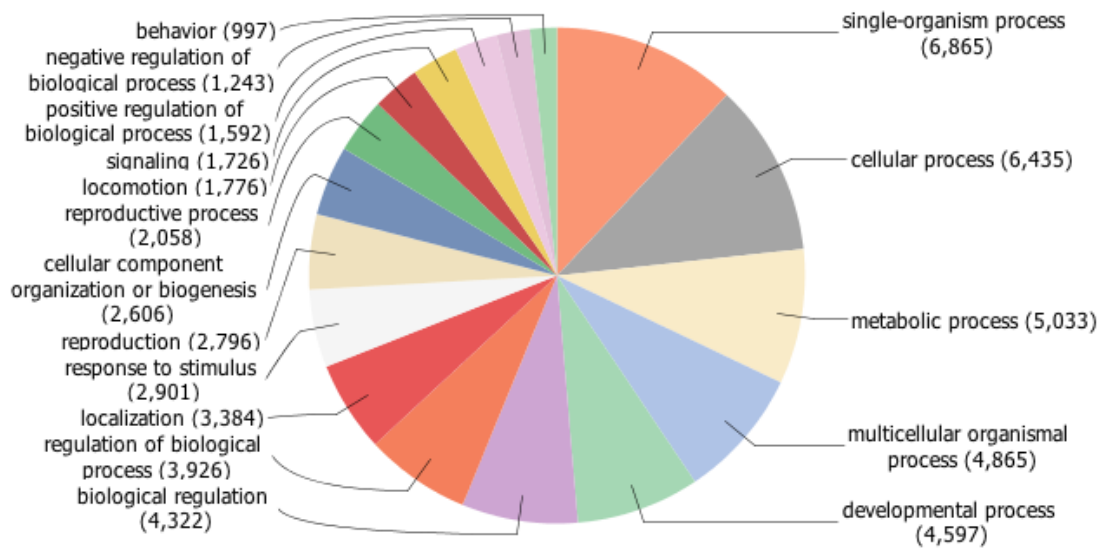


Figure 2.3 Blast2GO Gene Ontology analysis of the predicted *O. tipulae* genes. The categories included in the analysis are level 2 Biological Process GO terms. This breakdown of GO terms is consistent with other known microinvertebrate genomes.

rhabditid nematodes. This further supports the idea that nematode genomes greatly differ from one another and are incredibly diverse. The dN/dS ratios ranged from less than 0.01 to over 10.0, and 979 genes were shown to be positively selected with a dN/dS ratio greater than 1 (Figure 2.4). This means that only approximately 4.8% of the total gene count for *O. tipulae* and approximately 25.7% of the total number of orthologs between the seven species are actively under selective pressures.

Conclusions

In this study, a high quality draft genome of the soil nematode *Oscheius tipulae* was assembled using Illumina high-throughput sequencing, including the use of large-insert libraries for scaffolding. While the total size of the genome is smaller than previous estimates, other statistics, including BUSCO completeness, provide evidence that this genome assembly is strong. This is further supported by the estimated amount of protein-coding genes. These genes present a full makeup of the genetic profile of *O. tipulae*. While approximately two-thirds of these genes had orthologs in the *C. elegans* genome, the approximate one-third that did not suggest that while these two species may be in the same taxonomic family, their genomes are quite different. This was further supported by a synteny analysis showing a lack of conserved gene order or any large-scale synteny. This assembly analysis helps fill in gaps in the understanding of nematode genomes and is the first step in establishing *Oscheius tipulae* as a satellite model organism in the world of nematode studies.

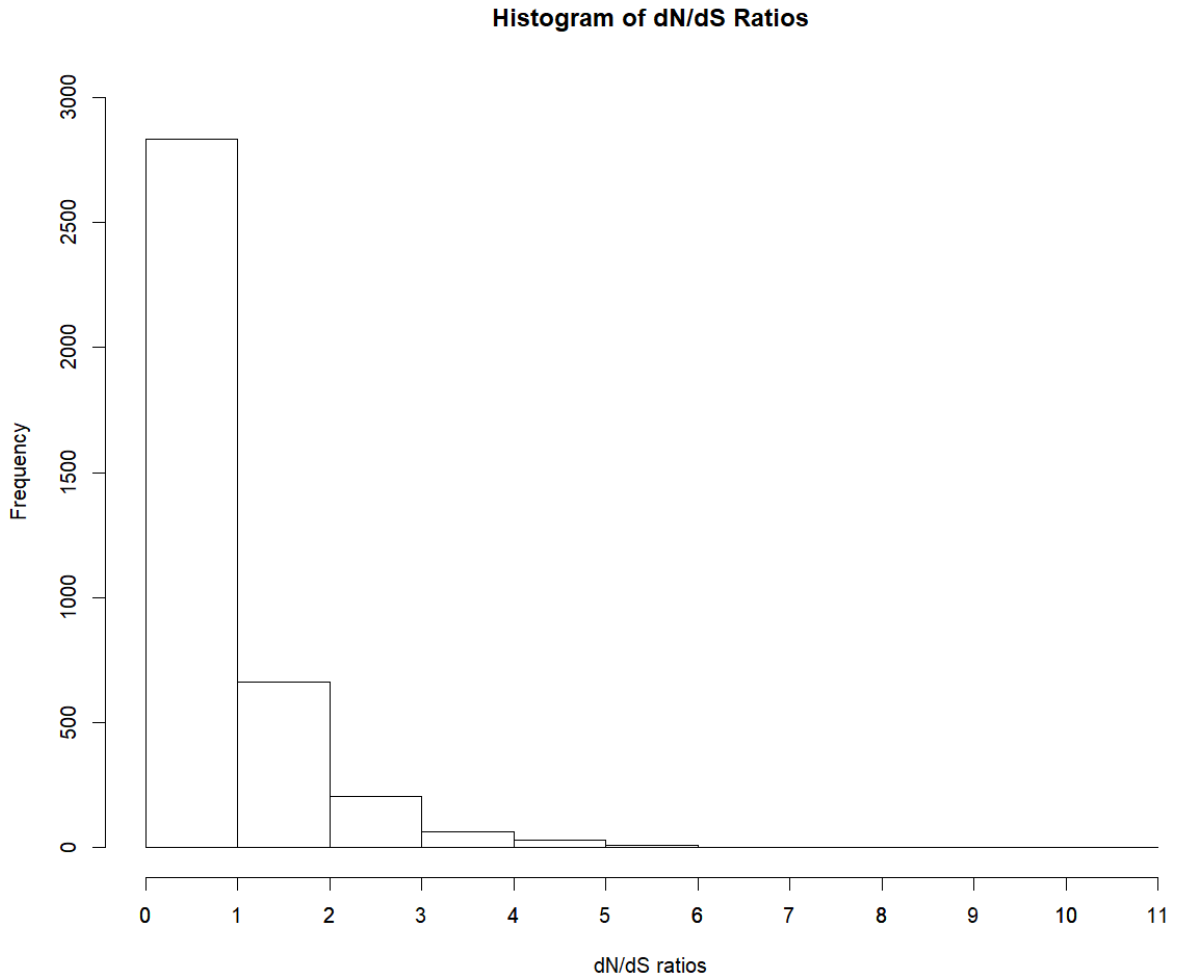


Figure 2.4 Histogram of the dN/dS ratios of the nematode reciprocal orthologs. Reciprocal orthologs were obtained from the genomes of *Caenorhabditis brenneri*, *Caenorhabditis briggsae*, *Caenorhabditis elegans*, *Caenorhabditis japonica*, *Caenorhabditis remanei*, *Pristionchus pacificus*, and the predicted KJO *O. tipulae* genes. A large majority of the genes have ratios < 1 , indicating they are not under positive selection. The remaining with a ratio > 1 are assumed to be under positive selection.

References

- Ahn, I. Y., & Winter, C. E. (2006). The genome of *Oscheius tipulae*: determination of size, complexity, and structure by DNA reassociation using fluorescent dye. *Genome*, 49(8), 1007-1015. doi:10.1139/g06-045
- Bai, X., Adams, B. J., Ciche, T. A., Clifton, S., Gaugler, R., Kim, K. S., . . . Grewal, P. S. (2013). A lover and a fighter: the genome sequence of an entomopathogenic nematode *Heterorhabditis bacteriophora*. *PLoS One*, 8(7), e69618. doi:10.1371/journal.pone.0069618
- Baïlle, D., Barriere, A., & Felix, M. A. (2008). *Oscheius tipulae*, a widespread hermaphroditic soil nematode, displays a higher genetic diversity and geographical structure than *Caenorhabditis elegans*. *Molecular Ecology*, 17(6), 1523-1534. doi:10.1111/j.1365-294X.2008.03697.x
- Besnard, F., Koutsovoulos, G., Dieudonne, S., Blaxter, M., & Felix, M. A. (2017). Toward Universal Forward Genetics: Using a Draft Genome Sequence of the Nematode *Oscheius tipulae* To Identify Mutations Affecting Vulva Development. *Genetics*, 206(4), 1747-1761. doi:10.1534/genetics.117.203521
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, 27(4), 578-579. doi:10.1093/bioinformatics/btq683
- Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology*, 13(R56), 1-9.
- Brenner, S. (1974). The Genetics of *Caenorhabditis elegans*. *Genetics*, 77, 71-94.
- Consortium, T. C. e. S. (1998). Genome Sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science*, 282(5396), 2012-2018. doi:10.1126/science.282.5396.2012
- Cotton, J. A., Lilley, C. J., Jones, L. M., Kikuchi, T., Reid, A. J., Thorpe, P., . . . Urwin, P. (2014). The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biology*, 15(R43), 1-17.
- Dieterich, C., Clifton, S. W., Schuster, L. N., Chinwalla, A., Delehaunty, K., Dinkelacker, I., . . . Sommer, R. J. (2008). The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nature Genetics*, 40(10), 1193-1198. doi:10.1038/ng.227
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, 32(5), 1792-1797. doi:10.1093/nar/gkh340

- Ghedini, E., Wang, S., Spiro, D., Caler, E., Zhao, Q., Crabtree, J., . . . Scott, A. L. (2007). Draft Genome of the Filarial Nematode Parasite *Brugia malayi*. *Science*, *317*, 1756-1760.
- Götz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., . . . Conesa, A. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res*, *36*(10), 3420-3435. doi:10.1093/nar/gkn176
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086
- Ingham, R. E., Trofymow, J. A., Ingham, E. R., & Coleman, D. C. (1985). Interactions of Bacteria, Fungi, and their Nematode Grazers: Effects on Nutrient Cycling and Plant Growth. *Ecological Monographs*, *55*(1), 119-140. doi:10.2307/1942528
- Jex, A. R., Liu, S., Li, B., Young, N. D., Hall, R. S., Li, Y., . . . Gasser, R. B. (2011). *Ascaris suum* draft genome. *Nature*, *479*(7374), 529-533. doi:10.1038/nature10553
- Kiontke, K., Gavin, N. P., Raynes, Y., Roehrig, C., Plano, F., & Fitch, D. H. (2004). *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. *PNAS*, *101*(24), 9003-9008.
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5*(2), R12.11-R12.19.
- Louvet-Vallee, S., Kolotuev, I., Podbilewicz, B., & Felix, M. A. (2003). Control of Vulval Competence and Centering in the Nematode *Oscheius* sp. 1 CEW1. *Genetics*, *163*, 133-146.
- Mayer, C. (2006-2010). Phobos 3.3.11. Retrieved from http://www.rub.de/ecoevo/cm/cm_phobos.htm
- McNulty, S. N., Strube, C., Rosa, B. A., Martin, J. C., Tyagi, R., Choi, Y. J., . . . Mitreva, M. (2016). *Dictyocaulus viviparus* genome, variome and transcriptome elucidate lungworm biology and support future intervention. *Scientific Reports*, *6*, 20316. doi:10.1038/srep20316
- Platt, R. N., 2nd, Blanco-Berdugo, L., & Ray, D. A. (2016). Accurate Transposable Element Annotation Is Vital When Analyzing New Genome Assemblies. *Genome Biology and Evolution*, *8*(2), 403-410. doi:10.1093/gbe/evw009
- Sijen, T., & Plasterk, R. H. A. (2003). Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature*, *426*(6964), 310-314.

- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212. doi:10.1093/bioinformatics/btv351
- Smit, A. F. A., Hubley, R., & Green, P. (2008-2015). RepeatModeler Open-1.0. Retrieved from www.repeatmasker.org
- Smit, A. F. A., Hubley, R., & Green, P. (2013-2015). RepeatMasker Open-4.0. Retrieved from <http://www.repeatmasker.org>
- Soderlund, C., Nelson, W., Shoemaker, A., & Paterson, A. (2006). SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res*, *16*(9), 1159-1168. doi:10.1101/gr.5396706
- Sommer, R. J., & Sternberg, P. W. (1995). Evolution of Cell Lineage in Vulval Equivalence Group of Rhabditid Nematodes. *Developmental Biology*, *167*, 61-74.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033
- Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, *33*(Web Server issue), W465-467. doi:10.1093/nar/gki458
- Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., . . . Waterston, R. H. (2003). The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol*, *1*(2), E45. doi:10.1371/journal.pbio.0000045
- Waterhouse, R. M., Seppey, M., Simao, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., . . . Zdobnov, E. M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*. doi:10.1093/molbev/msx319
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, *24*(8), 1586-1591. doi:10.1093/molbev/msm088

CHAPTER III
ABIOTIC STRESS RESPONSE TRANSCRIPTION PROFILES IN THE SOIL
NEMATODE *OSCHEIUS TIPULAE*

Abstract

In order to survive, an organism must be able to respond to the abiotic stresses that ecosystem places on it. *Oscheius tipulae*, a soil nematode that can be found in ecosystems all over the planet, must face multiple of these stresses. While much is known about the molecular and genetic levels in regard to these responses, it is not known how these genes work alongside each other across the entirety of the genome. In this study, an RNA-Seq analysis was performed using *O. tipulae* nematodes that were subjected to heat stress, desiccation stress, and freezing stress. Comparisons showed that, while there is an overlap in molecular responses needed to survive these stresses, the differential expression of genes responsible for those molecular responses was not completely shared, nor were they entirely unique. For any given shared molecular response, each stress involved had a unique suite of genes differentially expressed, but there was also some overlap in the genes used as well. On a large-scale genome-wide level, the nematode genome has been simplified by including these shared, more generalized, genes while also ensuring survival by establishing an alternate set of stress-specific genes. This not only sheds light on the ways in which nematodes can survive

these stresses, but it also allows for the better understanding of how whole genomes regulate their gene expression.

Introduction

Oscheius tipulae is a species of free-living soil nematode. Located within the same taxonomic family as the model organism *Caenorhabditis elegans*, the two have much in common: the ability to be easily grown under laboratory conditions, a diet primarily made of bacteria, and the tendency to be found in soils across the world. *O. tipulae* has actually been shown to have a wider range. Laboratory strains have been cultivated from populations in Brazil and Kansas, United States, and the species has also been located in the soils of Italian islands (Torrini et al., 2016). Because the species can be found over such a large range, *O. tipulae* individuals must be able to survive and persist in the unique abiotic stresses these ecosystems pose.

The KJO strain of *O. tipulae* was extracted from the soils of the Konza Prairie in Manhattan, Kansas, United States. Temperatures there can vary from well below freezing temperatures (0°C) to temperatures hovering around 100°C, all while experiencing periods of high and low precipitation. Because of this, three key abiotic stresses the KJO strain must face are heat stress, freezing stress, and desiccation stress. Over time, the species has developed a series of molecular and behavioral responses in order to survive these stresses.

In order to combat heat stress, nematodes must manage the biggest threat posed by increased temperatures: the denaturation and aggregation of proteins (Singer & Lindquist, 1998). In order to combat this, they increase their intracellular levels of heat shock proteins (HSPs) and organic osmolytes, such as trehalose. HSPs combat this

problem by chaperoning denatured proteins toward enzymes that can fold them into their proper formation, allowing them to retain their original function and prevent them from aggregating within the cell cytoplasm (Wharton, 2011). There is also extensive genetic control of the production of HSPs under heat stress. If a select few HSP genes are silenced, other HSP genes can compensate and become more highly transcribed (Eckl, Sima, Marcus, Lindemann, & Richter, 2017), and HSP-specific transcription factors have been shown to have a significant role in the heat shock response (Joo et al., 2016). Organic osmolytes are small solutes within the cell that influence osmosis and maintain cell shape. The organic osmolyte most prominent in the heat shock response is trehalose, which sees a 90% increase in nematodes under heat stress (Jagdale & Grewal, 2003). Trehalose is able to perform two roles during heat stress survival: it is able to act as an alternate source of carbohydrate under the increased energy demands (Lillie & Pringle, 1980), and it is able to help preserve the structures and functions of enzymes (Honda, Tanaka, & Honda, 2010; Hottiger, Boller, & Wiemken, 1987; Hottiger, de Virgilio, Hall, Boller, & Wiemken, 1994; Singer & Lindquist, 1998). Like HSPs, this allows the enzymes to function correctly and prevent intracellular aggregation.

Desiccation stress poses three main threats to nematodes. The first threat is a change in membrane shape, as it loses its fluid plastic form and becomes more rigid and gelatinous (Crowe, Crowe, & Chapman, 1984). Proteins can also denature and aggregate upon dehydration (Potts, 1994). The final threat is the accumulation of highly-reactive reactive oxygen species (ROS) that can alter the structures of cellular macromolecules (Adhikari, Wall, & Adams, 2009). In order to prevent these threats, nematodes will attempt to preserve water by coiling in on themselves, allowing less water to be lost

across the cuticle before they clump together, sacrificing those on the outside to preserve water for those in the center of the clump (Demeure, Freckman, & Van Gundy, 1979; Higa & Womersley, 1993). Trehalose is also produced under desiccation stress. It is able to stabilize the membrane by binding to the phosphate group and separating the phospholipids, allowing for a more fluid structure (Behm, 1997; Crowe et al., 1984; Higa & Womersley, 1993). As under heat stress, trehalose also acts to prevent the denaturation and aggregation of proteins, but it is not alone in this role (Higa & Womersley, 1993). Nematodes also use HSPs and late embryogenesis abundant (LEA) proteins to do this as well (Adhikari et al., 2009; Goyal, Walton, & Tunnacliffe, 2005). LEA proteins also have a second function. Their ion-binding properties allow them to act as an antioxidant, stabilizing the harmful reactive properties of ROS (Alsheikh, Svensson, & Randall, 2005; Hara, Fujinaga, & Kuboi, 2005; Tunnacliffe & Wise, 2007). They also seek out ROS, causing themselves to be reacted upon and damaged, preserving more important macromolecules (Hara, Fujinaga, & Kuboi, 2004).

As water becomes frozen, it cannot be biologically utilized by the nematode. Because of this, many of the physiological threats a nematode faces under freezing stress are also faced under desiccation stress. Nematodes not only have to deal with a similar phase change within their membrane, but proteins, again, begin to be denatured and aggregate (Ramløv, 2000). Unique to freezing stress is the threat of intracellular freezing, a process that can cause complete ruptures of cellular membranes (Mazur, 1984; Muldrew & McGann, 1994). In response, nematodes prevent intracellular water from freezing by using inhibitory molecules to prevent the behavior of ice nucleators, molecules known to trigger the shift from liquid water into ice (Wharton & Worland,

1998). Another strategy is to remove the possibility of intracellular freezing by forcing intracellular water out through the membrane, inducing desiccation (Holmstrup, Bayley, & Ramløtv, 2002; Wharton, 2003). When it comes to molecular responses, fatty acid levels increase and interact with the membrane to preserve fluidity (Margesin, Neuner, & Storey, 2007). Levels of LEAs, organic osmolytes, trehalose, and HSPs also increase, preventing protein denaturation and aggregation (Jagdale & Grewal, 2003; Jagdale, Grewal, & Salminen, 2005; Martinez, Perez-Serrano, Bernadina, & Rodriguez-Caabeiro, 2001; Zhang & Guy, 2006).

As Table 3.1 demonstrates, the three physiological responses to these three stresses have elements exclusive to each as well as significant overlaps. For instance, HSPs and trehalose are shown to be utilized in all three responses, whereas LEA proteins are found in the desiccation and freezing responses, but not in the heat response. Behavioral responses tend to show less overlap than the molecular responses. While these physiological responses and genetic properties are well known and detailed, what these processes look like across a genome are not. The question remains: how is the transcription of these genes regulated on a genome-wide level?

It may be possible that no matter the stress, similar genes are being upregulated as their corresponding protein is needed. For example, the exact same HSP genes are being transcribed under all three stresses as HSPs are needed for each response. Another possibility is that each stress response transcribes its own suite of genes. For example, while the formation of trehalose is necessary for each stress response, the genomes may transcribe a unique trehalose-6-phosphate synthase gene, a necessary enzyme for the formation of trehalose, depending on the stress. Both of these methods of genomic

Table 3.1 Lists of a subset of the known molecular and physiological responses nematodes have to three abiotic stresses: heat, desiccation and freezing. There are both stress-specific and overlapping responses. How these responses, as well as all of the other, are regulated on a genome-wide level is unknown.

Heat	Desiccation	Freezing
Heat shock proteins	Heat shock proteins	Heat shock proteins
Trehalose/organic osmolytes	Trehalose/organic osmolytes	Trehalose/organic osmolytes
	Late embryogenesis proteins	Late embryogenesis proteins
	Antioxidants	Antioxidants
		Induce desiccation
		Inhibit ice nucleators
	Coiling and clumping	

regulation fall under the one-gene-one-polypeptide hypothesis originally proposed by Beadle and Tatum (1941). Another possibility is that of pleiotropy, where the transcription of one gene is responsible for the production of a variety of molecular responses. If this were the case, each stress response may trigger only a small set of genes, each corresponding to multiple of the necessary physiological responses. This acts a way of reducing the cost of complexity within the genome (Wang, Liao, & Zhang, 2010). A final possible mechanism of genomic regulation that could be present here is that of modular pleiotropy, or modularity. Under modularity, one gene is responsible for the production of multiple polypeptides, but those polypeptides are few in number and similar in function (Wagner, 1996). This method can be seen as a hybrid of pleiotropy and the one-gene-one-polypeptide theory.

The purpose of the experiments done in this chapter is to determine the ways in which *O. tipulae* organizes and genomically regulates the transcription of the genes responsible for its survival under heat, desiccation, and freezing stresses. By utilizing the extensive knowledge obtained from the *O. tipulae* genome previously, and an in-depth RNA-Seq analysis, transcription levels for every gene within the genome will be calculated and compared across all three stress responses and a control. This information not only provides an explanation for this specific nematode species, but it also provides an extensive look into the most efficient mechanisms for genome regulation in microinvertebrates.

Materials and Methods

Nematode Preparation

O. tipulae nematodes were obtained from the Konza Prairie in Manhattan, Kansas and cultures were established within the laboratory. Nematodes were grown at 18°C on 60x15 mm non-vented dishes of Nematode Growth Media and seeded with the OP50 strain of *Escherichia coli* (Brenner, 1974) . Once the bacteria had been completely consumed, the nematodes were washed off plates with M9 buffer solution and pooled into a 15 mL conical tube and stored at 6°C for 1-7 days. This allowed for the remaining bacteria in the digestive tract to be broken down in order to minimize its DNA presence in the final sequencing results. Nematodes were then either used for DNA extraction, subjected to one of the three experimental groups: cold, desiccation, heat, or subjected to the control.

Stress Treatments

Freezing Treatment

M9 was either added or removed to the solution in the 15 mL tube in order to obtain a concentration of 200 individual nematodes per 200 µL solution. Then, 200 µL of solution was transferred into 24 individual 0.5 mL tubes. These tubes were then capped and placed at 6°C in order to allow for acclimation to the stress. After 48 hours, the samples were then treated to -20°C for 4 hours. The samples were then thawed at room temperature and pooled into a 15 mL tube. This process was repeated four times in order to obtain four biological replicates.

Desiccation Treatment

M9 solution was removed from the sample in order to obtain a concentration of 200 individual nematodes per 20 µL solution. Then, 20 µL was added to 24 individual 0.5 mL tubes, and the samples were placed at 18°C uncapped. This allowed for the remaining

M9 solution to evaporate, providing the desiccation stress. After the 48-hour treatment time, 200 μ L of M9 solution was added to each sample before they were all pooled into a single 15 mL conical tube. This process was repeated four times, allowing for four individual replicates.

Heat Treatment

M9 was either added or removed to the solution in the 15 mL tube in order to obtain a concentration of 200 individual nematodes per 200 μ L solution. Then, 200 μ L of solution was transferred into 24 individual 0.5 mL tubes. These samples were placed in a 30°C incubator for 24 hours to allow for acclimation to the warmer temperature. They were then placed in a 35°C bead bath for 8 hours. After treatment, all the samples were pooled into a 15 mL conical tube. This process was repeated four times, ensuring four individual replicates.

Control

For the initial solution, a concentration of 200 individual nematodes per 200 μ L solution was obtained by either removing or adding M9 solution. Then 200 μ L of this solution was pipetted into 24 individual 0.5 mL tubes. These samples were capped and placed at 18°C, the same temperature at which the nematode populations grew, for 48 hours. Afterwards, samples were pooled in a 15 mL conical tube, and like the treatment samples, this was repeated 4 times in order to obtain 4 biological replicates.

RNA Extraction

The RNA extraction protocol was the same across all treatment replicates. The 15 mL conical tubes containing the treated nematodes were centrifuged and excess M9 solution was removed until there was less than 0.5 mL remaining. The 100 μ L portion at

the bottom of the tube that contained all the nematodes was removed and added to 1 mL of TRIzol Reagent (Thermo Fisher Scientific). Each sample was then stored at -80°C.

Each sample went through three freeze-thaw cycles in order to help break the cuticle and allow for maximum RNA extraction. RNA was then extracted from each sample using Direct-zol RNA Miniprep Kit (Zymo Research) with an elution volume of 50 µL. Sample quality was assessed for each individual replicate through via NanoDrop and the Qubit 3.0 Fluorometer (Thermo Fisher Scientific). Once the concentration was confirmed to be above the University of Minnesota's minimum threshold, each sample was visualized on a 1% agarose gel. If RNA presence was clear, the sample was then deemed to be ready for sequencing.

Sequencing

The samples were submitted to the University of Minnesota Genomics Center in two separate runs. The first run consisted of all replicates for the treated samples, 12 in total. These samples underwent library preparation, including receiving individual tags for each replicate, and were sequenced together on one lane of the Illumina HiSeq 2500. Single-read 50 bp sequencing reads were obtained for each of the 12 replicates. The second sample consisted of the four control replicates. These also underwent library preparation and single-lane sequencing at the University of Minnesota using the Illumina HiSeq 2500. The only difference was that the reads obtained were 50 bp paired-end reads.

RNA-Seq

RNA-Seq analysis was performed using CLC Genomics Workbench 9. The sequencing reads for each stress were then mapped to the predicted protein-coding genes

obtained in Chapter II of this dissertation. The number of reads mapped to each gene represents transcription levels for that gene. Mismatch, insertion, and deletion costs were set to 2, 3, and 3, respectively, and length and similarity fractions were set to 0.5 and 0.98, respectively. Mapping was performed on both strands with no global alignment and max number of hits per read was set to 1. This analysis was done on all 12 replicates across each of the four treatment groups (cold, heat, desiccation, and control).

Statistical analysis of differentially expressed genes between the three treatments and the control were performed using the edgeR R package (Chen, Lun, & Smyth, 2014; Lun, Chen, & Smyth, 2016; Robinson & Oshlack, 2010) and the DESeq2 R package (Love, Huber, & Anders, 2014). The input for each analysis was the table of raw read counts obtained from the RNA-Seq analysis in CLC Genomics Workbench. Because this data is count based, and there a low number of replicates for each treatment, a negative binomial distribution must be assumed, and the chosen software programs do this.

The pairwise differential expression analysis was run three times in Blast2GO, each comparing one treatment to the control. Genes with count-per-million (CPM) values ≥ 1 in at least 4 of the replicates were kept, and the rest were discarded. Normalization was done using the trimmed mean of M-values (TMM), which normalizes gene counts based on overall library sizes and minimizes the perceived log-fold change between genes. The analysis in edgeR was done using these same parameters, and the analysis using DESeq2 was done using the default parameters.

Lists of upregulated and downregulated genes for each of three treatments were analyzed. Genes were then sorted into one of seven categories: found in heat only, desiccation only, freezing only, heat and desiccation but not freezing, heat and freezing

but not desiccation, desiccation and freezing but not heat, and heat and desiccation and freezing. This was done for both the upregulated and downregulated genes. Each gene list was then entered into DAVID 6.8 (Huang, Sherman, & Lempicki, 2009a, 2009b) for functional annotation clustering. The following annotation categories were used for the clustering: level 4 GO terms, direct GO terms, KEGG pathway terms, and terms from the InterPro database. A “high” stringency level was also used in order to increase the confidence that the genes belonged in their respective cluster.

Positive Selection

In order to determine if the genes involved in abiotic stress response are positively selected, those genes that were found to be upregulated or downregulated under at least one of the three stresses were compared to the list of positively selected genes obtained in Chapter II of this dissertation. Lists of upregulated and downregulated positively selected genes were generated for each of the seven categories of stress overlap.

Results and Discussion

Transcription Profiles

The heatmap produced (Figure 3.1) compares each of the 16 replicates with one another across all 20,402 genes in order to obtain a large-scale genome-wide comparison between each transcription profile. Figure 3.1 shows that the genome-wide responses to desiccation stress and freezing stress are more closely related with one another than either is to heat or the controls. Because the two stresses share both similar threats and similar molecular responses, this close relation in transcription profiles is to be expected. Interestingly, while all of the replicates for the heat, desiccation, and freezing stresses showed internal consistency by grouping with replicates of their own type, this was not

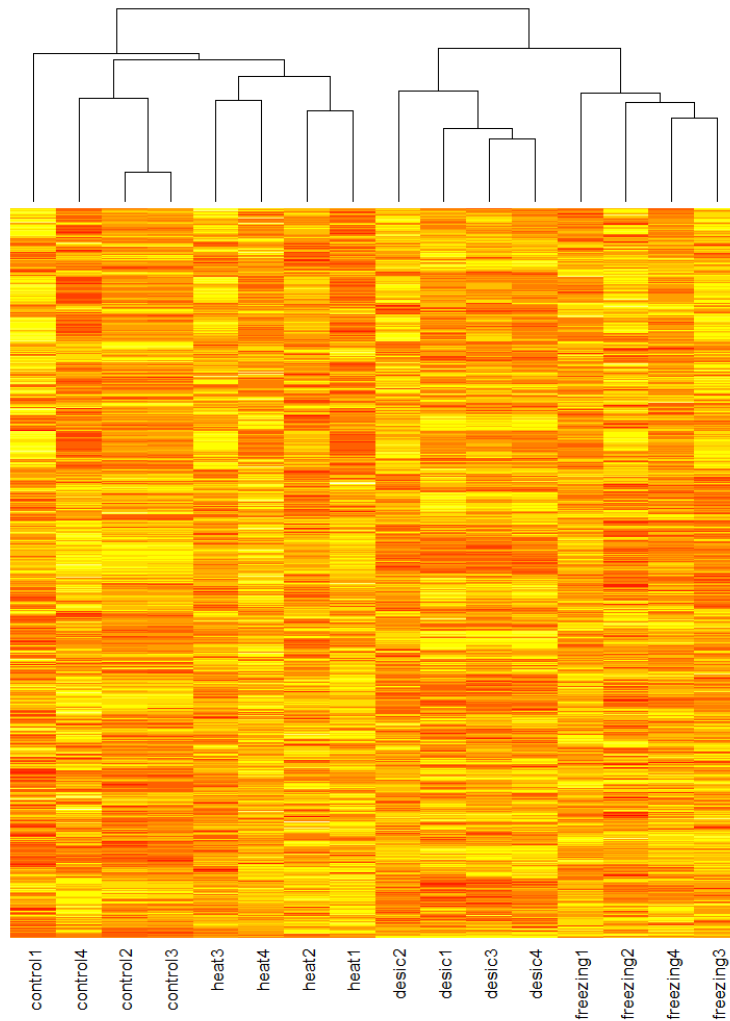


Figure 3.1 Heatmap of relative transcription levels across all treatment replicates. In total, there were 16 replicates, 4 for each of the 4 treatments: heat stress, desiccation stress, freezing stress, and the control. The brighter the shade of yellow indicates higher transcription levels, and the darker shade of red indicates lower transcription levels. The data indicates internal consistency within treatments and shows that the genomic responses to desiccation and freezing stresses are more closely related to each other than either are to the heat response or the control.

the case for one of the control samples, with it being more distantly related to the other three control replicates than those three replicates were to the heat samples. A multi-dimensional analysis also supports the relatedness of the transcription pattern replicates (Figure 3.2). Dispersion estimates and fitted dispersion estimates post-normalization can be found in Appendix Figures 3.3 and 3.4, respectively.

Across the entire experiment, 3,123 genes were upregulated in at least one of the treatments, and 3,073 genes were downregulated in at least one treatment (Figure 3.5 and Figure 3.6). These numbers make up 15.5% and 15.1% of the predicted protein-coding genes, respectively. Of these genes, 76 were found to be upregulated in at least one treatment *and* downregulated in at least one treatment. The log-fold changes and the differentially expressed genes for each of the three treatments can be visualized in Appendix Figures 3.7-3.9.

The upregulated genes fell into 157 DAVID functional annotation categories (Appendix Table 3.2). Of those, many known responses to these three stresses can be found. Many upregulated clusters contain genes involved with development, either of the larvae or the genitalia and sex differentiation. The process of locating, transporting, and breaking down denatured proteins is also a common annotation found within the upregulated clusters. Lastly, the transcription and translation processes were also found to represent multiple upregulated clusters. The genes that were downregulated in at least one treatment were grouped into 172 DAVID functional annotation clusters (Appendix Table 3.3). The three biological processes most represented within these clusters look to be the nervous system, the muscular system, and transport across a membrane.

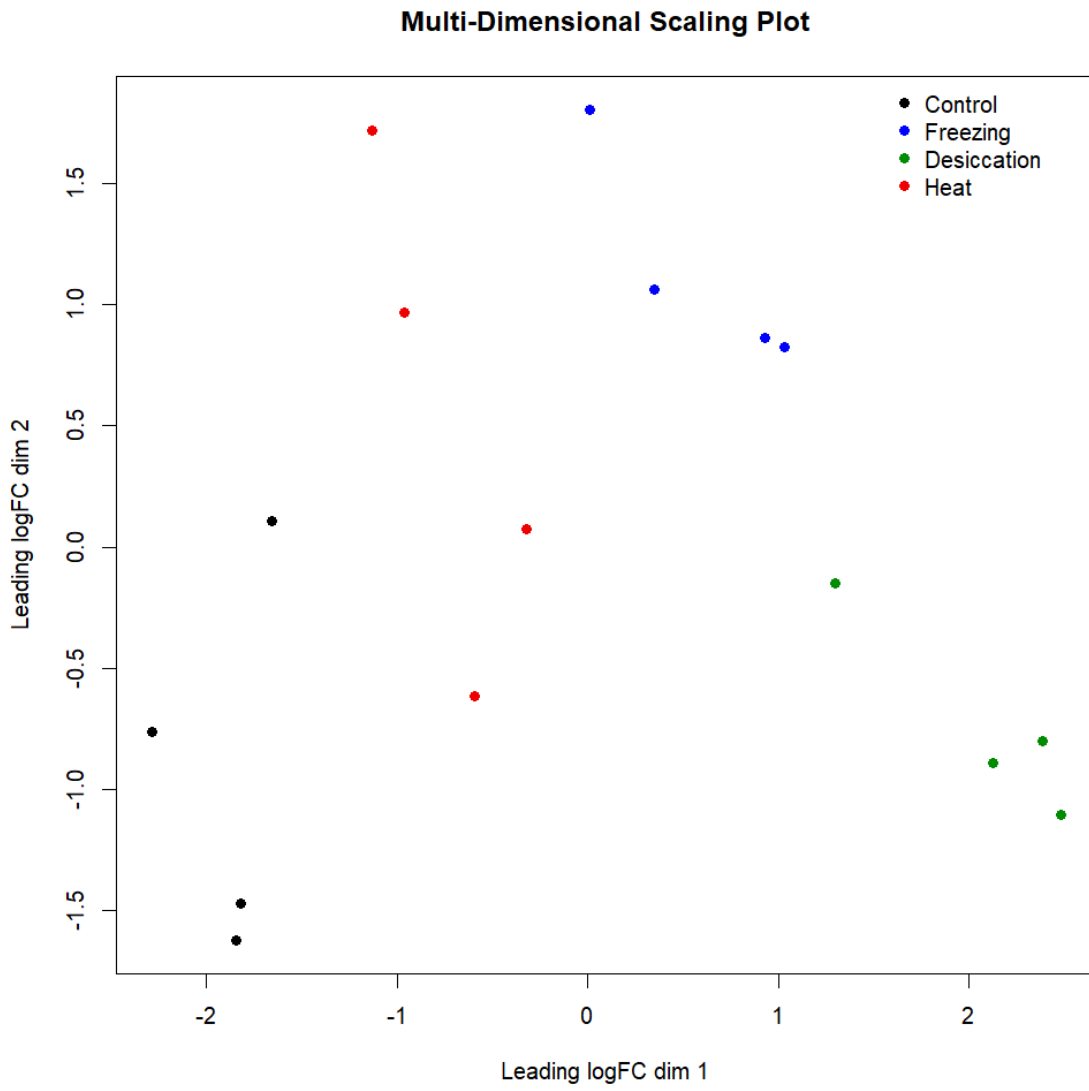


Figure 3.2 Multi-dimensional analysis of transcription pattern replicates. Clusters indicate internal consistency within treatments.

At this level, it appears that the upregulated genes are falling into annotations of processes that are known to be important during stress, whereas the downregulated genes have functions that are more involved with the upkeep of everyday life for the nematodes. This shows how as nematodes begin to face stress, they prioritize the genes necessary for survival while foregoing transcription of the genes that will not immediately aid in survival. The numbers of upregulated and downregulated genes being nearly equal shows how the number of genes actively being transcribed within the nematodes remains constant while this prioritization of stress response genes is occurring. Still, this does not answer the question of stress-specific transcription profiles, and in order to understand those, the upregulated and downregulated genes must be analyzed within the context of the three stresses.

Upregulated Genes

Heat Treatment

Of the 20,402 genes in the *O. tipulae* genome, 559 (~2.7%) were upregulated under heat stress. In the DAVID functional annotation analysis, these genes were sorted into 44 clusters (Appendix Table 3.4). Of the ten clusters with the highest DAVID enrichment scores, three (Clusters 1, 4, and 7) are involved with the translation and protein formation process. This is likely due to the nematode's need to produce new proteins that will aid in the molecular response and survival under heat stress. Another theme that reappears in the clusters is development (Clusters 2 and 10). As heat stress is known to alter and influence nematode development (Bird, 1972; Wong & Mai, 1973), this cluster is not unexpected. Two of the ten top clusters are involved with the response

and destruction of unfolded proteins (Clusters 5 and 8), which is a known response to heat stress since protein denaturation is known to occur under warm temperatures.

The remaining three clusters are uniquely found only once within the clusters with the ten highest Enrichment Scores. Cluster 3 is made up of genes involved with myosin and muscle, Cluster 6 contains genes that play a role in ribonucleotide binding, and Cluster 9 is comprised of genes related to neuropeptide hormone signaling.

Interestingly, there is one annotation expected to be present that is not. Across every annotation cluster found in the heat stress experiment, there is no annotation showing that heat shock proteins are upregulated under heat stress, despite the fact that they are known to be a very important response to heat stress. It might be possible that these genes were instead grouped into a separate cluster and given a different, possibly less specific annotation.

Desiccation Treatment

Under desiccation stress, 2,266 genes were upregulated, making up approximately 11% of the total number of genes across the genome. This is over 4x the number of genes upregulated under heat stress, indicating the threats posed under desiccation stress are greater both in count and magnitude. Upon analysis in DAVID, these genes fell into over 100 functional annotation categories (Appendix Table 3.5).

As under heat stress, clusters involved with translation and protein assembly (Clusters 1, and 4) were found within the ten clusters with the highest enrichment scores. More specifically, Cluster 1 is involved with the ribosome, while Cluster 4 is involved in the endoplasmic reticulum membrane, an organelle containing ribosomes that is substantially involved with the assembly of proteins. Also similar to the heat stress

clusters, development also was represented within the top ten clusters (Clusters 2, 5, 6, and 9). The increase in the need for the formation of new proteins is consistent with the idea that necessary stress response proteins are needed to survive desiccation stress.

Cluster 3 was made up of genes related to the protein chaperonin. This is used to chaperone unfolded proteins to either destruction or re-folding, but it could also be used to chaperone peptides together to form the larger quaternary structure of the protein complex. Both of these would be needed under desiccation stress, decreasing the levels of denatured and aggregated proteins while also quickening the rate of assembly of the proteins required to survive.

Cluster 8 contains genes involved with neuropeptide hormone signaling, another overlap between the desiccation response functional annotation and that of the heat response. A second annotation shared with the heat response is one involved with the binding of ribonucleotides (Cluster 10). Lastly, no cluster unique to desiccation stress is found in Cluster 7, which is comprised of genes involved in the defense response to organisms.

Similar to the heat stress results, there is one annotation that while expected to be present, is absent. Antioxidants and late embryogenesis proteins are known to be essential in minimizing the damage reactive oxygen species cause while under desiccation stress, but this annotation is not found among any of the clusters found to be upregulated under desiccation stress. Interestingly, while not found under heat stress, there is a cluster of heat shock protein genes found in this analysis (Cluster 29).

Freezing Treatment

The number of genes upregulated under freezing stress was 1,823, fewer than in desiccation but closer in number than under heat stress. Out of the 20,402 genes found within the *O. tipulae* genome, this number makes up almost 9% of them, and the DAVID functional analysis clustered the upregulated genes into 115 annotation clusters (Appendix Table 3.6).

As in both heat and desiccation stress, freezing stress poses the threat of protein denaturation and aggregation. The two functional annotation clusters with the highest Enrichment Scores both contained genes that were related to the activity of proteasomes, a protein complex containing enzymes responsible for the breakdown of proteins. Theoretically, this would solve the problem of protein aggregation by destroying the denatured proteins that have begun to aggregate. One of the top ten clusters are involved with the transcription and translation processes (Clusters 9), showing the increased demand for the creation of certain proteins.

Clusters 4 and 6 are similar in that they both contain genes involved with development. While the latter cluster is mostly associated with larval development, the former is more specifically involved with the development of the collagen cuticle that forms the outmost protective layer of the nematode that separates it from its environment. This cuticle can have functional overlaps with the genes found in Cluster 7, those involved in the defense response to other organisms. This cluster's annotation also was found in the top clusters upregulated under desiccation stress.

On a cellular level, clusters involved with structural elements are also found to be upregulated under freezing stress with both the cell cortex (Cluster 8) and tubulin (Cluster 10) represented. Tubulin in particular is a main component of microtubules

which themselves are an important element of the cytoskeleton. The cytoskeleton is responsible for maintaining the shape and structure of a cell, and this shape and structure begins to be compromised as the water internally and externally begins to freeze. This increase in tubulin allows the nematode to maintain the integrity of the cell's structure.

Cluster 3 contains genes responsible for ribonucleotide binding, an annotation shared with clusters found in both the heat and desiccation stress responses. Lastly, Cluster 5 is comprised of genes responsible for protein kinases, a series of enzymes that add phosphate groups to specific proteins. This is done to regulate many different biochemical pathways, so the more specific downstream effects of this cluster cannot be immediately determined.

As was the case under desiccation stress, there is no cluster to be found that contains the genes responsible for antioxidant behavior, even though this behavior is known to be an essential response to freezing stress. This could be due to the genes being clustered into an annotation group due to a more general GO term.

Comparisons between Treatments

Overall, the genes that were exclusive to only one of the three treatments had the same relative numbers as the upregulation patterns as a whole: desiccation stress caused the highest amount of upregulation, and heat caused the lowest amount (Figure 3.5). The largest subsection of overlapping genes was those upregulated in desiccation and freezing stress, but not in heat stress, further indicating the idea that desiccation and freezing stress responses are more closely related to each other than either are to the heat response. This subgroup is made up of 742 genes, or 23.7% of the total 3,123 genes upregulated in at least one treatment. The overlap of genes upregulated in the heat and desiccation

treatments but not the freezing treatment was made up of 109 genes, whereas only 44 genes were upregulated under both heat and freezing stress but not desiccation stress.

Elements of the removal of denatured proteins can be found in all three single stress-exclusive gene lists (Figure 3.5; Appendix Tables 3.4, 3.5, 3.6). Genes involved with peptidase activity are upregulated only when under heat stress, proteasome-related genes are upregulated only while freezing, and genes involved with the chaperoning of unfolded proteins are upregulated under desiccation stress. There is also a cluster of genes that was upregulated in all three stresses that plays a role in the response to unfolded proteins within the cell. This means when faced with the denatured proteins brought on by these three abiotic stresses, a set of generalized genes are transcribed to help ameliorate the problem, but there are also stress-specific genes that are utilized as well.

Similarly, sex differentiation and the development of the gonad and genitalia was also a recurring annotation across the treatments. Of the clusters with the ten highest Enrichment Scores, there was one cluster found upregulated only under heat stress, one found only under freezing stress, and three clusters found to be upregulated only while under desiccation stress. Two clusters were found to be upregulated under both freezing and heat stresses but not desiccation, while another cluster of genes was upregulated under both desiccation and heat stresses but not freezing.

Larval development was also a heavily recurring cluster annotation among the various treatment groups. Each of the following groups had one cluster of upregulated genes within their top ten clusters that focused on larval development: heat stress only, desiccation stress only, freezing and heat stresses but not desiccation, desiccation and

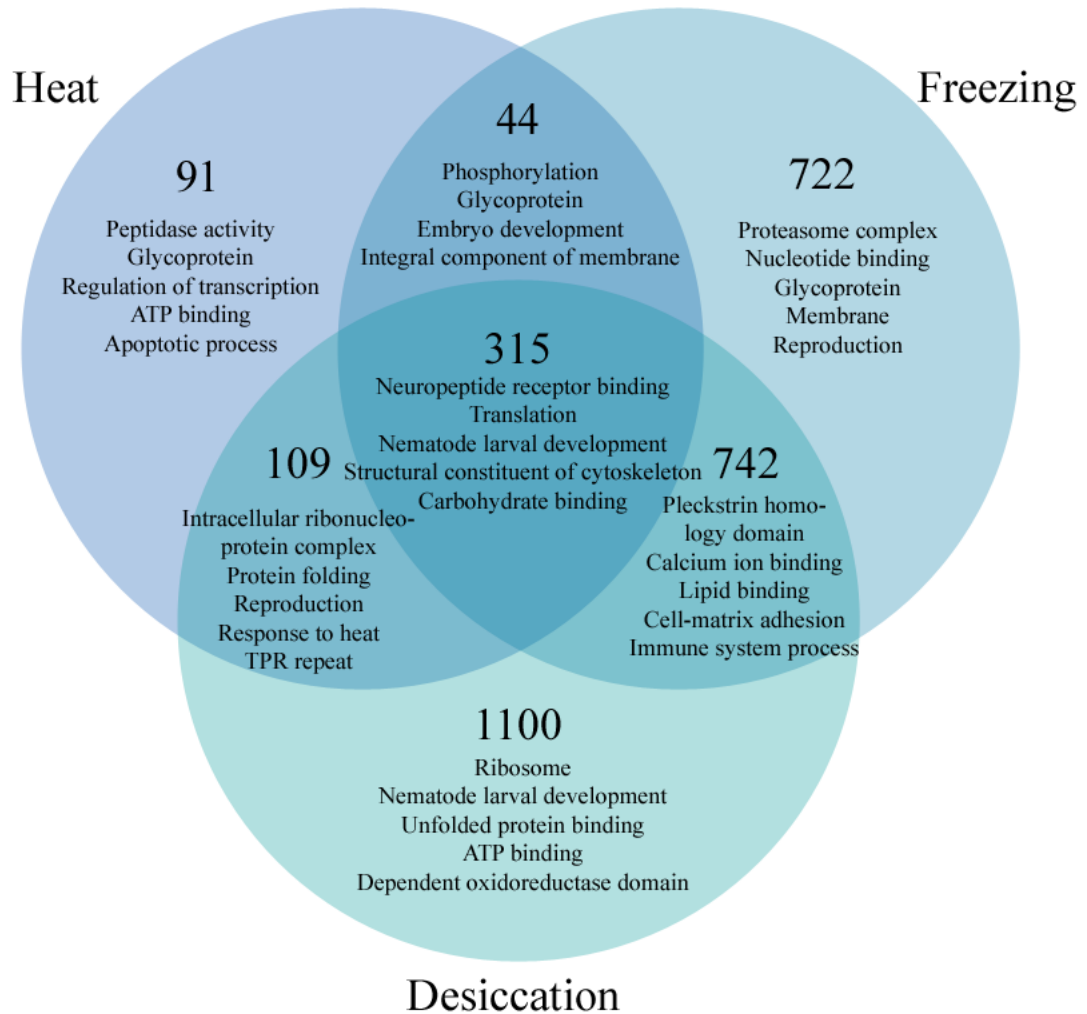


Figure 3.5 Venn diagram of genes upregulated in at least one of the three treatments. The number in each section indicates the number of genes that were found to be upregulated, and the text represents the five DAVID annotation clusters with the highest annotation scores. Each cluster is represented by the Gene Ontology (GO) term within it with the highest annotation score. For clusters with no GO term, the highest annotation of any type was used.

heat stresses but not freezing, and all three stresses. Each of these groups was made up of unique gene sets, but the same three Gene Ontology terms were found represented in each: “larval development”, “post-embryonic development”, and “nematode larval development”. This shows how there is both substantial overlap in the genes used to regulate larval development under these three stresses (as evidenced by the latter of the three groups), and there are also unique sets of genes specific to individual stresses as well (as evidenced by the first two groups).

Another common annotation found in the analysis was ribonucleotide binding. This annotation could be found in the following groups: heat stress only, freezing stress only, desiccation stress only, desiccation and heat stresses but not freezing, and all three stresses. Protein kinases and phosphatases are also found in multiple groups: freezing stress only, freezing and desiccation stresses but not heat, and freezing and heat stress but not desiccation. These two annotations provide further support for the idea that nematodes have unique suites of genes responsible for creating a molecular response under each individual stress, but they also have other sets of genes that produce the same molecular response that are transcribed under a variety of stresses.

Other functional annotations are found within the analysis, either appearing only in one or two groups. For the sake of simplicity, this analysis only compared and contrasted the ten functional annotation clusters with the ten highest DAVID Enrichment Scores for each of the seven possible groups. Full results of the DAVID analysis for each of the seven groups from Figure 3.5 can be found in Appendix Tables 3.7-3.13.

Downregulated Genes

Heat Treatment

Under the heat treatment, 257 genes were downregulated, comprising 1.3% of the predicted protein-coding genes. These genes were sorted into 36 functional annotation categories from DAVID (Appendix Table 3.14). The largest biological theme found within the ten clusters with the highest Enrichment Scores is that of transmembrane transport. Cluster 2 is comprised of genes involved with ABC transporters, which are directly responsible for transporting various compounds across a cell membrane. The genes found in Cluster 1 have annotations directly tied to transmembrane transport activity, and more specifically, ATPase activity. ATPases are a significant part of ABC transporters. Cluster 9 features genes responsible for integral components of the plasma membrane, and while this is a very general annotation, ABC transporters do require integral membrane proteins. This all indicates that under heat stress, transcription of the genes responsible for transmembrane transport is reduced. This could be to either maintain the homeostatic conditions prior to the influence of the heat stress or in order to allow for more efficient higher transcription of more essential stress response genes by reducing the transcription of the less essential genes.

GTPases have various roles within the cell, and they are also represented in the clusters found to be downregulated under heat stress (Cluster 4). One of their functions is to aid in protein transport across the membrane, which coincides with the previously mentioned clusters. GTPases also aid in the transport of vesicles within a cell, and the vesicles is another annotation found within these clusters (Cluster 7). As with transmembrane transport, the downregulation of these genes could be a way to prioritize transcribing more essential stress response genes.

Other annotations found within the ten clusters with the highest Enrichment Scores include genes involved with vacuoles, ribonucleotide binding, aspartic peptidase, positive selection, and the phosphatidylinositol signaling system.

Desiccation Treatment

Desiccation stress caused the downregulation of 2,494 of the predicted 20,402 genes (12.2%). This is a substantially larger amount of genes than were downregulated under heat stress, and the 163 DAVID functional annotation clusters they were sorted into is also much larger (Appendix Table 3.15). Of those clusters, the ten with the highest Enrichment Scores can be divided into three broader biological properties: membrane transport, neural signaling, and muscle.

Two clusters contain genes that are involved with transmembrane transport. Cluster 1 is comprised of genes responsible for intrinsic components of the plasma membrane, while the genes in Cluster 7 are responsible for ABC transporters. These annotations were also found to be downregulated under heat stress, indicating that transmembrane transport may be decreased purposefully as a general stress response across a variety of stresses.

The second broad category that these clusters fall into is that of neural signaling. Five clusters fall into this category. Cluster 2 genes are involved with the neuronal cell body, and Cluster 5 genes are involved with the morphogenesis of the neuron. The remaining three clusters have genes involved in neural signaling. The genes in Cluster 6 are involved with the synaptic membrane, while Clusters 4 and 9 contain genes responsible for neurotransmitter, and more specifically, acetylcholine, gated ion channels. This indicates that the typical neural activity of the nematode is deemed expendable

while under stressful conditions, as these genes are likely downregulated in order to allow the required enzymes needed for transcription to focus on transcribing genes more essential for the stress response.

This could also be the case for the muscular system of the nematode, as three remaining clusters with high Enrichment Scores are all comprised of genes involved with the musculature. Cluster 8 genes are responsible for the development, differentiation, and assembly of various parts of muscle tissue. More specifically, Cluster 3 is comprised of genes responsible for the M and A bands, while the genes of Cluster 10 are involved with the myosin complex that causes motor activity.

Freezing Treatment

The 1660 genes that were downregulated under freezing stress were clustered into 127 functional annotation groups by DAVID (Appendix Table 3.16). These genes make up 8.1% of the total number of predicted genes. Looking at the ten clusters with the highest Enrichment Scores, similar to the results from the desiccation experiment, multiple clusters of the downregulated genes under freezing stress are related to the muscle and motility of the nematode. Clusters 5 and 6 contain genes responsible for myosin and muscle tissue, and Cluster 8 contains genes responsible for A band and M band formation. Similar to while under desiccation stress, this indicates that the needs for muscle tissue and physical movement is not a priority under freezing stress. In order to allow for higher transcription levels of more necessary genes, these ones are downregulated.

Clusters 1 and 10 can be grouped together as well, as the genes that comprise each of them are involved with flavin adenine dinucleotide. The remaining five clusters are

unique to one another, and the fact that they are downregulated under freezing stress indicates that these aspects of the nematode's biology are not important to the response to freezing stress. Clusters 2, 3, 4, 7, and 9 contain genes responsible for ABC transporters, the endoplasmic reticulum membrane, kinases, catabolism, and the regulation of development, respectively.

Comparisons between Treatments

When looking at the Venn diagram of downregulation (Figure 3.6), many things initially stand out regarding the number of genes that fall into the various groups. Firstly, the numbers for each group is fairly similar to the respective number found in the analysis of the upregulated genes. Consequentially, the numbers of downregulated genes found in each group is also similar in relation to the other groups of downregulated genes. This means that differential expression of genes across the transcription profiles for each stress is constant, whether looking at the upregulated or downregulated genes.

The numbers of genes found downregulated in only one of the stresses was proportional to the numbers of genes found downregulated under those stresses in total. This means, there were more genes downregulated under desiccation stress only than there were downregulated in either heat or freezing stress only. Of these three groups, the number of genes found downregulated under heat stress only was the smallest. Of the groups in which genes were shared between two treatments, the number of genes found to be downregulated under both desiccation and freezing stresses but not heat stress was the highest with 986 genes. Only 19 genes were downregulated in both freezing and heat stresses but not desiccation, whereas 63 genes were downregulated under both desiccation and heat stress but not freezing. All seven groups also fell into the same order

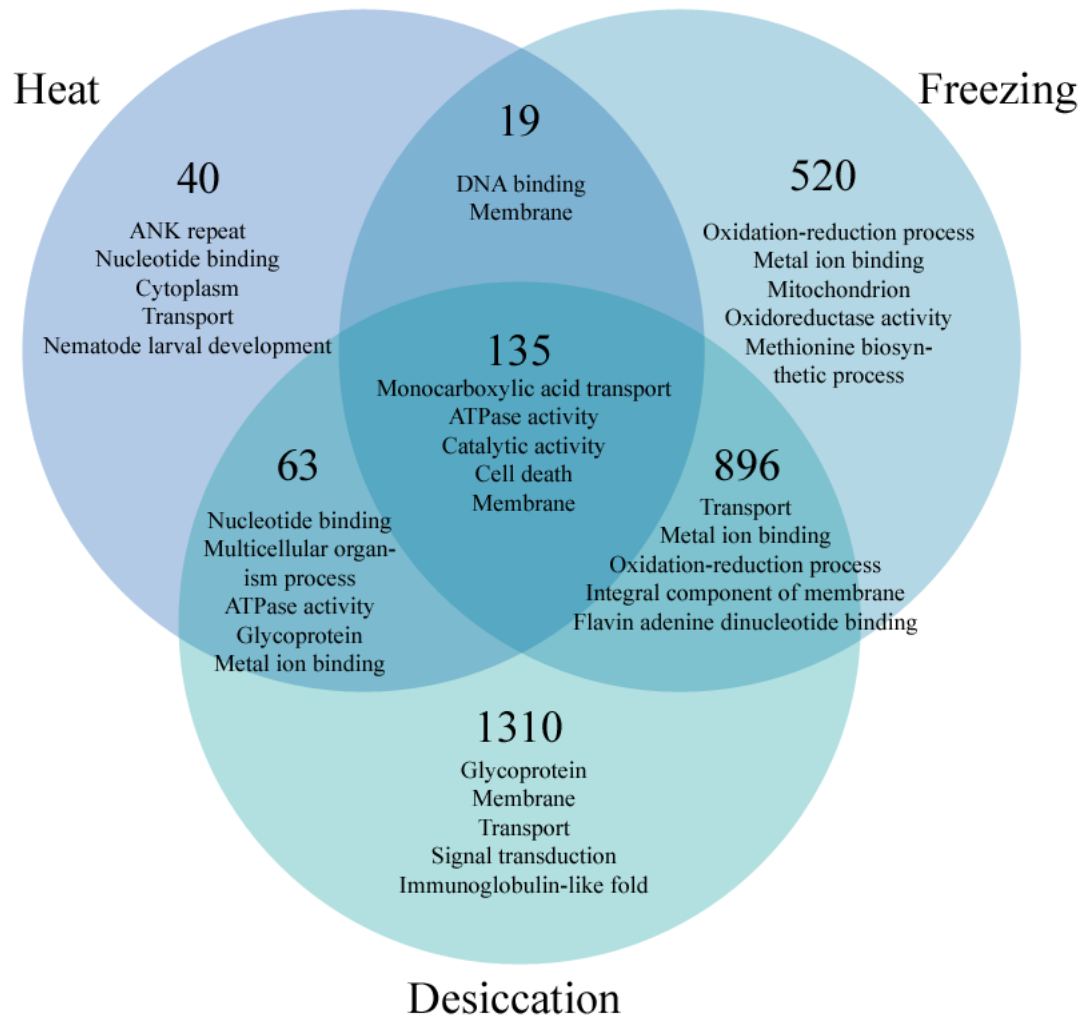


Figure 3.6 Venn diagram of genes downregulated in at least one of the three treatments. The number in each section indicates the number of genes that were found to be downregulated, and the text represents the five DAVID annotation clusters with the highest annotation scores. Each cluster is represented by the Gene Ontology (GO) term within it with the highest annotation score. For clusters with no GO term, the highest annotation of any type was used.

when it came to the number of DAVID annotation clusters found within each (Appendix Tables 3.17-3.23).

In comparing the DAVID function annotation results, only the ten clusters with the highest Enrichment Scores were analyzed. No annotation was found across all seven groups, but one annotation was found among six. Genes involved with integral components of the membrane were found clustered together in all of the comparison groups but one. This annotation cluster was not found among the top ten of the genes downregulated only under freezing stress. The widespread nature of this annotation not only indicates that its downregulation may be important to nematode survival under all of these stresses, but it also shows how the downregulation of some genes is stress-specific and for others is shared.

There were multiple clusters involved with larval development as well. These clusters of unique genes were found to be downregulated under only heat stress, under heat and freezing stress but not desiccation, and under heat and desiccation stress but not freezing. Interestingly, a cluster of downregulated genes under only freezing stress are involved in the negative regulation of larval and vulva development. A few other clusters were annotated to be involved with development on some level. Oocyte development genes were found to be downregulated under freezing stress only, whereas a cluster of downregulated genes under desiccation and heat stresses but not freezing were annotated to be involved with germ cell development, which is slightly more broad. Also found in this group is a cluster of downregulated genes involved in genitalia and gonad development.

Another common annotation found are genes involved with phosphorylation and kinase activity. Annotations in this vein were found to be downregulated under only freezing stress, only heat stress, under both desiccation and heat stresses but not freezing, and under all three stresses. Genes that were involved with ribonucleotide binding were found to be clustered together under three groups. One cluster of genes was downregulated under only heat stress, another cluster was downregulated in both heat and desiccation but not freezing, and a third cluster was downregulated under all three stresses. Lastly, genes involved with neural signaling were also found to be downregulated in both the desiccation stress only group and the group of all three stresses, but genes responsible for the negative regulation of signal transduction were also downregulated.

The remaining clusters of downregulated genes do not fall into any easily comparable categories, nor do they have any easily recognizable explanation due to the stresses. The downregulation of these specific annotations may not be highly relevant to the nematode's survival under these stresses. Instead, these genes may simply be downregulated in order to free up the molecular components responsible for transcriptions so that they can instead focus on transcribing the more important genes for survival.

Positive Selection

Of the upregulated genes, 169 of them were found to be positively selected, meaning they had a dN/dS ratio greater than 1 (Figure 3.10). When looking at which of the stresses the positively selected genes were upregulated in, the counts are similar, relative to one another, to those found in the total list of upregulated genes (Figure 3.11a).

Histogram of dN/dS Ratios for Upregulated Genes

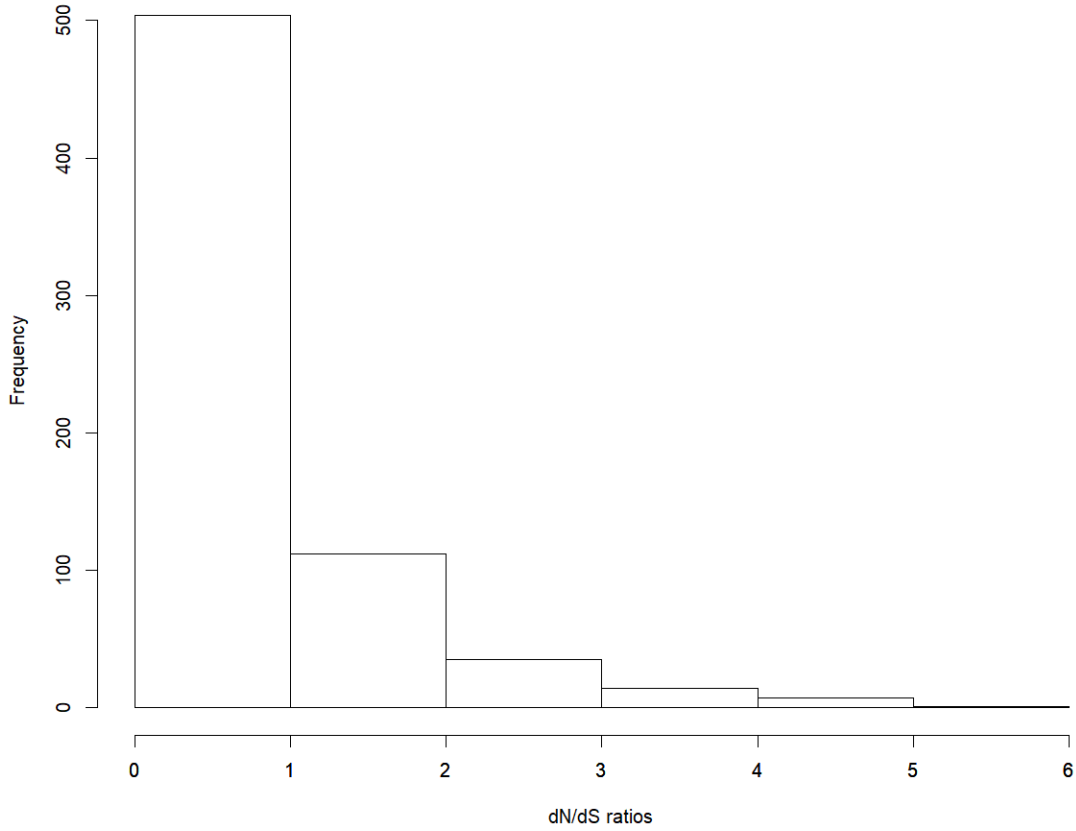


Figure 3.10 Histogram of the dN/dS ratios for the upregulated genes. Each gene had reciprocal orthologs across the seven species used. A dN/dS ratio greater than 1 indicates positive selection.

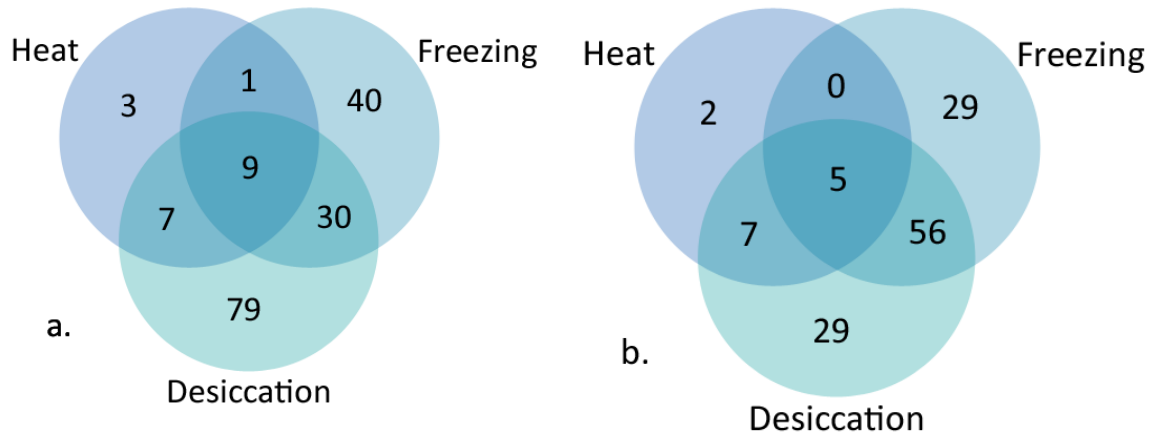


Figure 3.11 Venn diagrams of the positively selected genes. The counts represent the number of genes found to be (a) upregulated or (b) downregulated under at least one of the three treatments.

Both diagrams include the same total number of genes (169) in similar ratios.

The highest number of positively selected upregulated genes was found upregulated only under desiccation stress with 79. Under only freezing stress, 40 of the upregulated genes were also positively selected, and 30 genes were found to be positively selected and upregulated under both freezing and desiccation stresses. These three groups, the groups only involved with desiccation and freezing stresses, makes up 149 of the 169 positively selected upregulated genes. This clearly shows these two stresses, of the three, are the leaders in placing selective pressures on genes. Relative to the total numbers of upregulated genes under these stresses, these numbers are fairly low, indicating that while necessary for survival, these genes are predominantly not under strong selective pressures.

Of the genes found to be downregulated in at least one of the stresses, 169 genes were found to be positively selected (Figure 3.12). This is the same number found to be both upregulated and positively selected. Just as in the upregulation analysis, the three highest counts were found in the lists of genes downregulated in desiccation stress only (71), freezing stress only (29), and both desiccation and freezing stress (56), showing that these two stresses are placing more selective pressure on the genes than heat stress is (Figure 3.11b). Again, these numbers are all still drastically lower than the total amounts of genes downregulated in the overall study, further supporting claims from the upregulation analysis.

Conclusions

The large-scale analysis across all genes and all replicates shows that, genomically, the responses to freezing stress and desiccation stress are more similar to one another than either are to the response to heat stress or the control (Figure 3.1). This

Histogram of dN/dS Ratios for Downregulated Genes

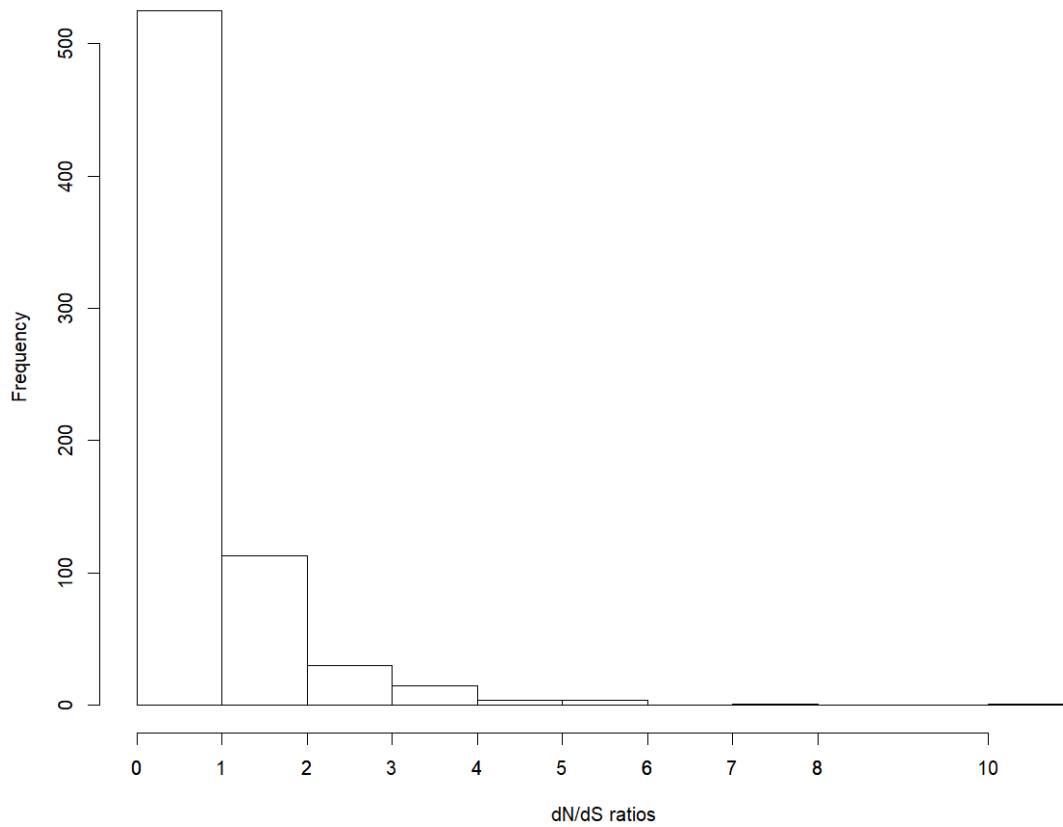


Figure 3.12 Histogram of the dN/dS ratios for the downregulated genes. Each gene had reciprocal orthologs across the seven species used. A dN/dS ratio greater than 1 indicates positive selection.

further supported not only by the known molecular responses to each stress, but also by the numbers of the differentially expressed genes under each stress. Whether upregulated or downregulated, the relative number of genes found to be differentially expressed under both treatments was much higher than either stress response had with the heat stress response (Figures 3.5 and 3.6). Also, in terms of the numbers of genes differentially expressed, the heat response is much smaller than that of freezing or desiccation, with much fewer genes being upregulated or downregulated under heat stress than the other two.

The DAVID functional annotation analysis also shed light on the ways in which the responses are organized on a genome-wide level. Each stress has a known set of physiological products that need to be transcribed in order to survive the stress, and there are instances of commonalities across the three stress responses. Looking at the upregulated genes, while some of the genes responsible for these known overlapping physiological responses can be found upregulated under multiple stresses, there are still certain genes that produce these same responses that are unique to each stress. This means that the transcription patterns are not defined solely by the molecular responses needed, but they are not also fully defined by the type of abiotic stress to which they must respond. The transcription patterns follow a hybrid of these two ideas, where in order to produce the necessary products, genes are both shared and kept unique across the three stresses. This provides an optimization of the genome, where genome size is reduced by containing genes that are parts of multiple stress responses, and by also including stress-specific genes that ensure survival under each stress.

References

- Adhikari, B. N., Wall, D. H., & Adams, B. J. (2009). Desiccation survival in an Antarctic nematode: molecular analysis using expressed sequenced tags. *BMC Genomics*, *10*, 69. doi:10.1186/1471-2164-10-69
- Alsheikh, M. K., Svensson, J. T., & Randall, S. K. (2005). Phosphorylation regulated ion-binding is a property shared by the acidic subclass dehydrins. *Plant, Cell and Environment*, *28*, 1114–1122.
- Beadle, G. W., & Tatum, E. L. (1941). Genetic control of biochemical reactions in *neurospora*. *PNAS*, *27*, 499-506.
- Behm, C. A. (1997). The Role of Trehalose in the Physiology of Nematodes. *International Journal for Parasitology*, *27*(2), 215-229.
- Bird, A. F. (1972). Influence of Temperature on Embryogenesis in *Meloidogyne javanica*. *Journal of Nematology*, *4*(3), 206-213.
- Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D., & Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*, *27*(4), 578-579. doi:10.1093/bioinformatics/btq683
- Boetzer, M., & Pirovano, W. (2012). Toward almost closed genomes with GapFiller. *Genome Biology*, *13*(R56), 1-9.
- Brenner, S. (1974). The Genetics of *Caenorhabditis elegans*. *Genetics*, *77*, 71-94.
- Chen, Y., Lun, A. T. L., & Smyth, G. K. (2014). Differential expression analysis of complex RNA-seq experiments using edgeR. In S. Datta & D. S. Nettleton (Eds.), *Statistical Analysis of Next Generation Sequence Data*. Springer, New York.
- Crowe, J. H., Crowe, L. M., & Chapman, D. (1984). Preservation of membranes in anhydrobiotic organisms - the role of trehalose. *Science*, *223*(4637), 701-703.
- Demeure, Y., Freckman, D. W., & Van Gundy, S. D. (1979). Anhydrobiotic Coiling of Nematodes in Soil. *Journal of Nematology*, *11*(2), 189-195.
- Eckl, J., Sima, S., Marcus, K., Lindemann, C., & Richter, K. (2017). Hsp90-downregulation influences the heat-shock response, innate immune response and onset of oocyte development in nematodes. *PLoS One*, *12*(10), e0186386. doi:10.1371/journal.pone.0186386
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*, *32*(5), 1792-1797. doi:10.1093/nar/gkh340

- Goyal, K., Walton, L. J., & Tunnacliffe, A. (2005). LEA proteins prevent protein aggregation due to water stress. *Biochemical Journal*, 388, 151-157.
- Hara, M., Fujinaga, M., & Kuboi, T. (2004). Radical scavenging activity and oxidative modification of citrus dehydrin. *Plant Physiol Biochem*, 42(7-8), 657-662. doi:10.1016/j.plaphy.2004.06.004
- Hara, M., Fujinaga, M., & Kuboi, T. (2005). Metal binding by citrus dehydrin with histidine-rich domains. *J Exp Bot*, 56(420), 2695-2703. doi:10.1093/jxb/eri262
- Higa, L. M., & Womersley, C. Z. (1993). New insights into the anhydrobiotic phenomenon: The effects of trehalose content and differential rates of evaporative water loss on the survival of *Aphelenchus avenae*. *Journal of Experimental Zoology*, 267(2), 120-129.
- Holmstrup, M., Bayley, M., & Ramløv, H. (2002). Supercool or dehydrate? An experimental analysis of overwintering strategies in small permeable arctic invertebrates. *PNAS*, 99(8), 5716-5720.
- Honda, Y., Tanaka, M., & Honda, S. (2010). Trehalose extends longevity in the nematode *Caenorhabditis elegans*. *Aging Cell*, 9, 558-569.
- Hottiger, T., Boller, T., & Wiemken, A. (1987). Rapid changes of heat and desiccation tolerance correlated with changes of trehalose content in *Saccharomyces cerevisiae* cells subjected to temperature shifts. *FEBS Letters*, 220(1), 113-115.
- Hottiger, T., de Virgilio, C., Hall, M. N., Boller, T., & Wiemken, A. (1994). The role of trehalose synthesis for the acquisition of thermotolerance in yeast. *European Journal of Biochemistry*, 219, 187-193.
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1), 1-13. doi:10.1093/nar/gkn923
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protocols*, 4(1), 44-57.
- Jagdale, G. B., & Grewal, P. S. (2003). Acclimation of entomopathogenic nematodes to novel temperatures: trehalose accumulation and the acquisition of thermotolerance. *International Journal for Parasitology*, 33(2), 145-152. doi:10.1016/s0020-7519(02)00257-6
- Jagdale, G. B., Grewal, P. S., & Salminen, S. O. (2005). Both Heat-Shock and Cold-Shock Influence Trehalose Metabolism in an Entomopathogenic Nematode. *The Journal of Parasitology*, 95(5), 988-994.

- Joo, H. J., Park, S., Kim, K. Y., Kim, M. Y., Kim, H., Park, D., & Paik, Y. K. (2016). HSF-1 is involved in regulation of ascaroside pheromone biosynthesis by heat stress in *Caenorhabditis elegans*. *Biochem J*, *473*(6), 789-796. doi:10.1042/BJ20150938
- Lillie, S. H., & Pringle, J. R. (1980). Reserve Carbohydrate Metabolism in *Saccharomyces cerevisiae*: Responses to Nutrient Limitation. *Journal of Bacteriology*, *143*(3), 1384-1394.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. doi:10.1186/s13059-014-0550-8
- Lun, A. T. L., Chen, Y., & Smyth, G. K. (2016). It's DE-licious: a recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. In *Methods in Molecular Biology* (Vol. 1418, pp. 391-416).
- Margesin, R., Neuner, G., & Storey, K. B. (2007). Cold-loving microbes, plants, and animals--fundamental and applied aspects. *Naturwissenschaften*, *94*(2), 77-99. doi:10.1007/s00114-006-0162-6
- Martinez, J., Perez-Serrano, J., Bernadina, W. E., & Rodriguez-Caabeiro, F. (2001). Stress response to cold in *Trichinella* species. *Cryobiology*, *43*(4), 293-302. doi:10.1006/cryo.2001.2363
- Mazur, P. (1984). Freezing of living cells: mechanisms and implications. *American Journal of Physiology - Cell Physiology*, *247*, C125-142.
- Muldrew, K., & McGann, L. E. (1994). The Osmotic Rupture Hypothesis of Intracellular Freezing Injury. *Biophysical Journal*, *66*, 532-541.
- Potts, M. (1994). Desiccation Tolerance of Prokaryotes. *Microbiological Reviews*, *58*(4), 755-805.
- Ramløv, H. (2000). Aspects of natural cold tolerance in ectothermic animals. *Human Reproduction*, *15*, 26-46.
- Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(R25), 1-9.
- Singer, M. A., & Lindquist, S. (1998). Multiple Effects of Trehalose on Protein Folding In Vitro and In Vivo. *Molecular Cell*, *1*, 639-648.
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

- Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33(Web Server issue), W465-467. doi:10.1093/nar/gki458
- Torrini, G., Mazza, G., Strangi, A., Barabaschi, D., Landi, S., Mori, E., . . . Roversi, P. F. (2016). *Oscheius tipulae* in Italy: Evidence of an Alien Isolate in the Integral Natural Reserve of Montecristo Island (Tuscany). *Journal of Nematology*, 48(1), 8-13.
- Tunnacliffe, A., & Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften*, 94(10), 791-812. doi:10.1007/s00114-007-0254-y
- Wagner, G. P. (1996). Homologues, Natural Kinds and the Evolution of Modularity. *American Zoologist*, 36, 36-43.
- Wang, Z., Liao, B. Y., & Zhang, J. (2010). Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Sciences*, 107(42), 18034-18039. doi:10.1073/pnas.1004666107
- Wharton, D. A. (2003). Freezing survival and cryoprotective dehydration as cold tolerance mechanisms in the Antarctic nematode *Panagrolaimus davidi*. *Journal of Experimental Biology*, 206(2), 215-221. doi:10.1242/jeb.00083
- Wharton, D. A. (2011). Cold Tolerance. In R. N. Perry & D. A. Wharton (Eds.), *Molecular and Physiological Basis of Nematode Survival*: CAB International.
- Wharton, D. A., & Worland, M. R. (1998). Ice Nucleation Activity in the Freezing-Tolerant Antarctic Nematode *Panagrolaimus davidi*. *Cryobiology*, 36, 279-286.
- Wong, T. K., & Mai, W. F. (1973). Effect of Temperature on Growth, Development and Reproduction of *Meloidogyne hapla* in Lettuce. *Journal of Nematology*, 5(2), 139-142.
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*, 24(8), 1586-1591. doi:10.1093/molbev/msm088
- Zhang, C., & Guy, C. L. (2006). In vitro evidence of Hsc70 functioning as a molecular chaperone during cold stress. *Plant Physiol Biochem*, 44(11-12), 844-850. doi:10.1016/j.plaphy.2006.09.012

CHAPTER IV
PRESENCE AND LOCATION OF DNA METHYLATION IN THE GENOMES
OF THE SOIL NEMATODES *OSCHEIUS TIPULAE* AND *CAENORHABDITIS*
ELEGANS

Abstract

DNA methylation is just one of the many mechanisms organisms use to alter phenotypic expression. This is done through the adding of methyl groups to DNA, which in turn alters protein-DNA interactions and influences gene transcription rates. Traditionally, DNA methylation was measured and studied on small-scale genetic levels, but advances in high-throughput sequencing have allowed for larger-scale analyses that measure genome-wide methylation levels. It has long been assumed that nematodes do not utilize DNA methylation as studies indicated methylated cytosines were not found within the model organism *Caenorhabditis elegans*. Recently, this has begun to be questioned as newer studies have found evidence of methylation within the Nematoda phylum. In this chapter, genome-wide DNA methylation was assessed in the soil nematodes *Oscheius tipulae* and *C. elegans* using whole-genome bisulfite sequencing (WGBS) and MethylCap-Seq analyses. Going against traditional assumptions, results indicated that cytosine methylation does exist in both of the species, with levels in *C. elegans* being higher than those in *O. tipulae*. In *O. tipulae*, methylated cytosines were primarily found within genes rather than intergenic regions, following trends found in the

genomes of other invertebrates, but the *C. elegans* genome found differing patterns of methylation, leaving its role in the genome uncertain. Utilizing data from Chapter III of this dissertation, it was also concluded that DNA cytosine methylation likely does not play a role in the genome-wide response to abiotic stresses in *O. tipulae*. The results from this chapter question historical beliefs in regard to *C. elegans* genome methylation and present new questions in the role of DNA cytosine methylation in the Nematoda phylum.

Introduction

The field of epigenetics looks toward the modifications of gene expression as a way to explain the phenotypic changes within an organism. One of the ways in which this is done in organisms is through DNA methylation, a process where methyl groups are attached to molecules of DNA, altering their chemical properties. In turn, this modifies the protein-DNA interactions, altering whether or not the locations close to that methyl group can be transcribed by the transcription enzyme suite. Historically, it has been believed that DNA methylation is responsible for the preventing of DNA transcription (Razin & Cedar, 1991). More recent work has not only supported this traditionally held belief, but it has also shown that the role of methylation is not that simple. It has also been shown that DNA methylation can be correlated with the increase in transcription rates as well (Flores et al., 2012; Wan et al., 2015).

The predominant method of DNA methylation is through the methylation of the carbon in the 5 position of cytosine nucleotides, forming 5-methylcytosine. Within the sequence of DNA, the three most commonly reported methylation contexts in which 5-methylcytosine can be found are: CG, CHG, and CHH, with each sequence written in the 5' to 3' orientation. The CG sequence context is frequently discussed as CpG. In the

CHG and CHH sequences, the H refers to any nucleotide that is not a guanine. Cytosine is not the only nucleotide onto which methyl groups can be located. Studies have shown adenine is also capable of being methylated (Greer et al., 2015; Wu et al., 2016; G. Zhang et al., 2015), but methyladenines have not received the same amount of research as methylcytosines.

As reviewed by Suzuki & Bird, in order to regulate gene transcription, invertebrates tend to show patterns of a mosaic DNA methylation (2008). This means methylation patterns are found in stable domains scattered throughout the genome where they are utilized, contrasting the general patterns of global methylation found in mammals and other vertebrates. These methylated domains tend to fall within gene bodies, which is a feature that highly conserved across both animals and plants (Feng et al., 2010; Zemach, McDaniel, Silva, & Zilberman, 2010). More specifically, methylation sites can be found more in the exons of the gene than in introns (Feng et al., 2010), and exons that are involved in transcription are found to be more highly methylated than exons that are not involved transcription (Flores et al., 2012). Promoter methylation, a feature associated strongly associated with the transcription levels of genes in mammals (Boyes & Bird, 1992; Jackson-Grusby et al., 2001), is not believed to play a role in the transcriptional activities in invertebrates (Keller, Han, & Yi, 2016).

Gene body methylation is positively correlated with higher transcription levels (Aran, Toperoff, Rosenberg, & Hellman, 2011; Ball et al., 2009; Bonasio et al., 2012; Lister et al., 2009; Wang et al., 2013). Methylation within active gene bodies is also higher than both the methylation levels of inactive gene bodies and the flanking sequences surrounding the gene (Aran et al., 2011). It is also believed that gene body

methylation has a role to play in alternative gene splicing. Changes in methylation patterns correspond to changes in splicing (Bonasio et al., 2012), and higher methylation levels can correlate with higher alternative splicing (Flores et al., 2012). Functionally, in insects, genes with methylated gene bodies produce protein sequences that are more highly conserved than those that do not contain gene body methylation. Genes with conserved methylated gene bodies are also more likely to be responsible for the housekeeping processes of transcription and translation, while genes that have lost their gene body methylation are more likely to be involved with cellular signaling and reproductive processes (Sarda, Zeng, Hunt, & Yi, 2012).

DNA methylation was believed to not exist in nematodes due to it not being found within the genome of the model organism *Caenorhabditis elegans*. This has caused much of the invertebrate methylation work to be performed on insects. The discovery of DNA methylation within the genome of the parasitic nematode *Trichinella spiralis* proved that the Nematoda phylum was not void of DNA methylation (Gao et al., 2012). Outside of cytosine methylation, it has been shown that adenine methylation exists within the *C. elegans* genome (Greer et al., 2015). With these findings starting to cast doubt on the traditional idea that nematodes as a phylum lack DNA methylation, there is room for further work to be done to confirm what may be true.

A secondary indicator of DNA methylation within the genome is the presence of methyltransferase enzymes. DNA (cytosine-5-)-methyltransferase (DNMT) is responsible for the addition of methyl groups onto the cytosines in the CG context within a genome, and the three most commonly studied DNMTs based on their relevance in humans are DNMT1, DNMT3A, and DNMT3B. These genes are not found within *C. elegans*,

another factor leading to the belief cytosine methylation does not exist in the organism, but orthologs have been identified within the *T. spiralis* genome (Gao et al., 2012). DNA N6-methyltransferase is a second family of enzymes responsible for the methylation of adenine, and this gene family has been found within the *C. elegans* genome.

The purpose of this chapter is to document the pattern of DNA methylation, particularly cytosine methylation, in the genome of *Oscheius tipulae*, a free-living soil nematode found within the same taxonomic family as *C. elegans*. In order to allow for comparison, the same analyses were performed on the *C. elegans* genome as well. Two high throughput sequencing analyses were utilized to do this: MethylCap-Seq and Whole Genome Bisulfite Sequencing (WGBS). MethylCap-Seq is an analysis where after random genome-wide DNA fragmentation, methylated pieces are extracted and sequenced. In the WGBS analysis, DNA is fragmented before unmethylated cytosines are converted to uracil, sequenced, and mapped back to the original genome. DNA methylation presence was also indirectly tested by the identification of methyltransferase genes within the *O. tipulae* genome was also determined using known methyltransferase from a variety of species. Lastly, I assessed the role that DNA methylation plays in the genomic regulation of abiotic stress response by using the assembled genome, transcriptome, and differentially expressed gene lists obtained from Chapters II and III of this dissertation. These analyses can provide useful detail regarding amounts, locations, and effects of methylcytosines in nematode genomes, allowing for a more comprehensive understanding of the role DNA methylation may have in the nematode phylum.

Materials and Methods

Nematode Growth and DNA Extraction

Individual populations of the N2 strain of *Caenorhabditis elegans* and the KJO strain of *Oscheius tipulae* were grown on 60x15 mm non-vented dishes of Nematode Growth Media and fed the OP50 strain of *Escherichia coli* at 18°C (Brenner, 1974). Populations grew until the bacteria had been fully consumed, and the nematodes from multiple plates were pooled by washing them off and into a 15 mL conical tube with M9 buffer solution, ensuring that the two species were kept separate. The samples were then incubated at 6°C for 1-7 days in order to allow for any remaining bacteria in the digestive tract to be digested. This process was repeated once for each species, ensuring there were two samples established for DNA extraction for each.

Genomic DNA was then extracted from each of the four samples using the PureLink Genomic DNA Mini Kit (Invitrogen). Each sample was visualized via gel electrophoresis on a 1% agarose gel, and concentrations were measured via the Qubit 3.0 Fluorometer (Thermo Fisher Scientific) in order to ensure they reached the minimum concentration threshold for sequencing (all samples were >30 ng/uL). Both *O. tipulae* samples and both *C. elegans* samples were then sent to the University of Missouri DNA Core Facility for library preparation and sequencing.

Library Preparation and Sequencing

At the University of Missouri Core DNA Facility, one *O. tipulae* sample and one *C. elegans* sample underwent bisulfite conversion via the EpiMark® Bisulfite Conversion kit (New England BioLabs). Under this process, the sodium bisulfite treatment converts all unmethylated cytosines to uracil while retaining the bases for methylated cytosines. Each sample then underwent whole-genome bisulfite sequencing (WGBS) library preparation via the NEBNext Ultra DNA Library Prep Kit for Illumina

(New England BioLabs). The remaining *O. tipulae* and *C. elegans* samples underwent methylation enrichment treatment, or MethylCap-Seq, via the EpiMark® Methylated DNA Enrichment Kit (New England BioLabs). This process fragments the DNA and uses the methyl-CpG binding properties of the human protein MBD2 to extract DNA fragments containing methylated cytosines under the CG context. These extracted fragments then underwent methylated enriched library preparation via the NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioLabs). All four samples were then sequenced in a single lane of a HiSeq 2500 run producing single end reads that were 50 bp in length.

Whole-Genome Bisulfite Sequencing Analysis

Whole Genome Bisulfite Sequencing (WGBS) analysis was performed using three bioinformatics programs in order to ensure results were consistently obtained for both the KJO *O. tipulae* analysis as well as the *C. elegans* analysis. In these analyses, the WGBS reads from the *O. tipulae* WGBS sequencing were mapped to the genome scaffolds assembled in Chapter II of this dissertation. The *C. elegans* WGBS reads were mapped to the archived WS201 *C. elegans* chromosomes obtained from WormBase.

The first program utilized was Bismark v0.17.0 (Krueger & Andrews, 2011), and the default parameters were used. A second WGBS analysis was run in BS-Seeker2 v2.1.0 (Guo et al., 2013). The WGBS reads for each species were filtered for quality using the programming software before being aligned to the corresponding genome. Parameters at each stage of the analysis were set to their default. The third WGBS mapping analysis was performed using CLC Genomics Workbench 10 (CLC Bio, Cambridge, MA, USA). The read mapping was directional, and the parameters were set

to the CLC standard settings: match and mismatch scores of 1 and 2, insertion and deletion costs each set to 3, and length and similarity fractions were set to 0.8. The reads were mapped randomly to the assembled scaffolds. When assessing for methylation levels, duplicate and non-specific matches were ignored, and methylation levels were confirmed by the reads.

MethylCap-Seq Analysis

The reads from the MethylCap-Seq sequencing runs were mapped to the corresponding genome using Bowtie2 under the default parameters (Langmead & Salzberg, 2012). Duplicate reads were then removed and mapping peaks, representing hypermethylated regions, were detected using a p-value threshold of 0.001 in MACS v1.4.2 (Y. Zhang et al., 2008). This process was done for both *O. tipulae* and *C. elegans*, with estimated genome sizes of 60 Mb and 100 Mb used for the estimated genome size parameter for each analysis, respectively.

In order to determine the location of hypermethylated regions within the nematode species' genes, the hypermethylated peak data obtained from the MACS software were then compared to the annotated genome files for the corresponding genome. The locations of the peaks were defined by the peak summit value obtained from MACS. The annotated genome for *O. tipulae* was obtained from the Augustus program (Stanke & Morgenstern, 2005) during the genome's initial gene prediction, and the *C. elegans* annotated genome was the archived WS201 annotated file obtained from WormBase. Due to the variances in annotation within these two files, the analyses differed.

In the *O. tipulae* analysis, all peak summit points were initially separated into two groups: those that fell within the gene body and those that fell within the intergenic regions. Next, the peaks located within the gene body were then labeled more specifically, as either being located within the coding sequence (CDS), an intron, or the 5' or 3' untranslated regions (UTRs). If a peak summit fell within any exon or the start or stop codon, it was deemed part of the CDS. Transcription start sites (tss) were deemed to be part of the 5' UTR, and transcription termination sites (tss) were grouped with the 3' UTR. Peaks located within the intergenic region were then compared to the locations of the tandem repeats that were obtained from Phobos v3.3.11 (Mayer, 2006-2010).

For the *C. elegans* analysis, the archived annotation of the WS201 genome was obtained from WormBase, and annotations of the mitochondrial DNA sequence were removed. Only annotations from the source "Coding_transcript" were used in order to eliminate peaks landing within multiple annotations based on different sources. Alternate transcripts were also removed to prevent peaks from being counted more than once. Peaks that fell within the following annotations were counted and deemed to be located within the gene body: CDS, intron, 3'-UTR, and 5'-UTR. Peaks that did not land within these gene parts were determined to be located within the intergenic region.

In order to determine the methylated cytosine composition of these peaks, the location of each CG cytosine predicted to be methylated by the Bismark WGBS analysis was extracted, and its position within the genome was compared to the coordinates of each MethylCap-Seq peak. Methylated cytosines located within the range of the peak were counted, and the mean methylated CG per peak was calculated. This was performed on both the *O. tipulae* and *C. elegans* data.

Methyltransferase Presence

As an indirect measure of DNA methylation presence in *O. tipulae*, the predicted gene amino acid sequences were analyzed to determine whether DNA methyltransferase genes were present. In order to do this, two reciprocal Blastp runs were performed using the BLAST executables from NCBI in order to find similarity of protein sequences. The first run blasted DNA (cytosine-5-)-methyltransferase (DNMT) genes obtained from NCBI against the amino acid sequences of the predicted *O. tipulae* genes. The *O. tipulae* genes that were found as a result were then reciprocally blasted against the DNMT genes. The second reciprocal Blastp setup was between the same predicted *O. tipulae* amino acid gene sequences and a list of N6-methyltransferases obtained from NCBI. Lists of genes used for both analyses can be found in Appendix Table 4.1. Each Blastp used an E-value cutoff of 0.001, and the remaining parameters were left at their defaults.

Role of Methylation in Abiotic Stress Response

This data was also able to be utilized to determine if DNA methylation has a role in the genomic regulation of heat, desiccation, and freezing stresses in *O. tipulae*. In order to do this, the list of *O. tipulae* genes found to have a hypermethylated peak summit within its CDS was compared to the lists of genes that were upregulated under the heat, freezing, and desiccation stresses as described and obtained in Chapter III. Only peaks located within the CDS were used because methylated regions within the gene body, particularly exons, are believed to be the most influential in increasing transcription levels of that gene in invertebrates.

The upregulated genes found to contain hypermethylated peak summits were then annotated using DAVID 6.8 (Huang, Sherman, & Lempicki, 2009a, 2009b). Refseq

accession numbers obtained from the initial annotation Blast in Chapter II were used for the analysis. Functional annotation clustering was performed using a stringency level of “high” and the following annotation categories: level 4 Gene Ontology terms, direct Gene Ontology terms, KEGG pathway terms, and terms from the InterPro database.

Results and Discussion

Whole-Genome Bisulfite Sequencing

The Bismark analysis for the *O. tipulae* genome had a mapping efficiency of 90.8%, uniquely mapping 49,750,304 of the 54,765,892 reads to the scaffolds. It found that 5.4% of the cytosines found in the CpG context were methylated, 5.6% of the cytosines in the CHG context were methylated, and 5.7% of the cytosines in the CHH context were methylated. In these contexts, ‘H’ represents any nucleotide other than guanine (G). After filtering, 83.36% of the reads remained in the BS-Seeker2 analysis. Of these, 95.62% successfully mapped to the genome. The percentages of cytosines methylated in the CpG, CHG, and CHH contexts were 6.253%, 6.400%, and 6.415%, respectively. This is a slight increase than the levels assessed from the Bismark analysis. In the CLC Genomics Workbench analysis, approximately 22.9% of the reads were not included in the analysis as they were either found to be duplicate or non-specific. The proportion of the cytosines methylated in the CpG, CHG, and CHH contexts were 5.41%, 5.60%, and 5.68%, respectively. These values are nearly exactly the values obtained using Bismark, and less than those using BS-Seeker2.

The *C. elegans* Bismark analysis only had an 86.2% mapping efficiency as 39,166,626 of the 45,416,203 reads successfully mapped uniquely. This is lower than that of the *O. tipulae* analysis. Within the genome, it was found that 8.9% of the cytosines in a

CpG context were methylated. Under the CHG and CHH contexts, cytosines were methylated 9.2% and 9.4% of the time, respectively. Similar to the *O. tipulae* analysis, methylation levels were estimated to be higher by the BS-Seeker2 program. Cytosines in the CpG, CHG, and CHH contexts were methylated 9.614%, 9.787%, and 9.855% of the time, respectively. These results come after the 12.85% of the initial WGBS reads were filtered out due to poor quality. Of the remaining reads, approximately 92.73% successfully mapped to the genome. In the CLC Genomics analysis for *C. elegans*, approximately 77.7% of the WGBS reads were included in the analysis, as the other 22.3% reads was excluded due to being duplicates or non-specific in nature. The methylation levels called for the cytosines in the CpG, CHG, and CHH contexts were 9.02%, 9.24%, and 9.34%, respectively. These numbers are only slightly larger than the Bismark values and smaller than the BS-Seeker2 values.

These *C. elegans* values go against the historical conclusions claiming that there is no DNA methylation found in the species. Due to the species' role as a model organism, this conclusion has also been extrapolated to other soil nematodes, especially those that are similar taxonomically. The *O. tipulae* data, while lower in number than that in *C. elegans* shows evidence that not only are nematode genomes highly variable, but their epigenomes are as well.

MethylCap-Seq

The MACS analysis for the assembled *O. tipulae* genome found 1,256 peaks ($p < 0.001$), each representing a specific hypermethylated region. The average peak length was 949 bp, and the total additive length of the peak ranges was 1,191,805 bp. This is approximately 2% of the total length of the assembled *O. tipulae* genome. Each peak

contained an average of 112 methylated CGs. This result was obtained using ~11% of the reads as the remaining 89% were found to map redundantly and were removed from the analysis.

Of the 1,256 peaks in the *O. tipulae* genome, 14.3% of the peak summits were found within the intergenic regions, and approximately 85.7% of the hypermethylated CG regions were found within the gene (Table 4.2). Only 74.1% of the base pair content in the genome is found within genes, meaning this can support the known tendency for DNA methylation in invertebrates to fall more likely within the gene body than not (Zemach et al., 2010). The CDS makes up 50.8% of the base pair content of genes, but of the 1,076 peak summits within the gene body, 78.9% fell within the CDS. This follows what was expected based on the work of Feng et al. (2010). Peaks found within CDS comprised of multiple exons tended to fall more likely toward the 3' end of the gene, with 182 falling within terminal exons and 100 falling within initial exons.

Behind the CDS, introns had the next highest peak count with 170. Of the remaining 57 peaks, 6 fell within gaps of the annotation. In all cases, the gaps fell within one of the UTRs, with annotated UTR regions falling both upstream and downstream of the gap. In these cases, the peaks were deemed to be a part of the corresponding UTR. The final totals for the 5' UTR and 3' UTR were 26 and 31, respectively.

Of the 180 hypermethylated peaks located within the intergenic regions, 107 (59.4%) were located within the tandem repeat regions. Since tandem repeats only make up 10.9% of the intergenic regions in regards to base pair content, this leads to the conclusion that tandem repeats are methylated at higher rates than the intergenic regions in *O. tipulae*. Within these tandem repeats, more peaks were located within

Table 4.2 Annotation of hypermethylated region peak locations within the *O. tipulae* genome compared to total bp content in genome. Methylation peaks occur within gene bodies and tandem repeats at higher percentages than either total base pair (bp) percentage within the genome. Within the gene body, hypermethylated peaks are located within coding sequences (CDS) at higher rates and within introns at lower rates than either of their genomic makeup within the genome.

Genomic feature	Genome content (bp)	Percentage of total content	Hypermethylated peaks (#)	Percentage of total peaks
Total genome	60,646,666	100%	1,256	100%
Gene body	44,957,717	74.1%	1,076	85.7%
(CDS)	(22,842,686)	(37.7%)	(849)	(67.6%)
(introns)	(17,734,451)	(29.2%)	(170)	(11.1%)
Tandem repeats	1,712,313	2.8%	107	8.5%
(microsatellites [2-10 bp])	(1,642,662)	(2.7%)	(75)	(6.0%)
(minisatellites [11-100 bp])	(69,651)	(0.1%)	(32)	(2.5%)

microsatellites than within minisatellites, but that is likely due to the total amount of microsatellite content found within the genome (1,642,662 bp) is much larger than the amount of minisatellite content (69,651).

The *C. elegans* analysis found 2,774 peaks ($p < 0.001$) across all six of its chromosomes with the average peak length to be 1,464 bp and total additive length of 4,059,813 bp. This means approximately 6.7% of the total genome is represented within one of these peaks, and, on average, each peak contained 47 methylated CGs. Similar to the *O. tipulae* analysis, a large majority (~86%) of the reads were excluded from the analysis due to redundant mapping. These numbers of hypermethylated regions fall in line with the WGBS analysis, supporting the idea that cytosine methylation can be found in the species, with those in *C. elegans* occurring more frequently than in *O. tipulae*.

The locations of the hypermethylated peaks in *C. elegans* differ from those found in the *O. tipulae* genome (Table 4.3). Of the 2,774 peaks, slightly more than half (51.2%) were found within the gene body. This is much less than the 85.7% found within the gene body in *O. tipulae*. The genome also contains more gene content than intergenic content (65.2% vs 34.8%, respectively), so this looks to be differentially methylated. A large majority of the peaks found within the gene body were located in introns, which is different from the methylation patterns found in *O. tipulae* and the patterns found across invertebrates, where CDS are more highly methylated than introns. Introns make up 54.5% of the gene content, but 79.5% of the hypermethylated peak summits within genes are found within introns, showing this high percentage is not due to genetic content alone. Only 19% of the peaks within the gene body were found in the CDS. The UTRs were

Table 4.3 Annotation of hypermethylated region peak locations within the *C. elegans* genome compared to total bp content in genome. Hypermethylated peaks are located within the gene body and intergenic regions at a nearly equal amount. Peaks within the gene body are predominantly located in introns, rather than in the coding sequences (CDS). Both of these facts go against what was found in *O. tipulae* (Table 4.2) and the general consensus of methylation levels within invertebrates.

Genomic feature	Genome content (bp)	Percentage of total content	Hypermethylated peaks (#)	Percentage of total peaks
Total genome	100,272,208	100%	2,774	100%
Gene body	65,385,566	65.2%	1,421	51.2%
(CDS)	(25,963,224)	(25.9%)	(271)	(9.8%)
(introns)	(35,666,495)	(35.6%)	(1,130)	(40.7%)
Intergenic region	34,886,642	34.8%	1,353	48.8%

equally methylated, but in very low numbers (<1% of the peaks found within the gene body).

This data indicates that while cytosine methylation is present within the genome of *C. elegans*, it may not have the same effects on gene transcription that it does in other invertebrates. Whereas higher transcription in invertebrates is paired with higher methylation levels within the CDS, it is unclear as to what high intron and low CDS methylation levels mean for gene transcription in *C. elegans*.

Methyltransferase Presence

In order to determine whether *O. tipulae* is capable of methylating DNA, the predicted genes from the assembled genome (Chapter II of this dissertation) were compared to known DNA methyltransferase genes in two reciprocal Blastp analyses, one for DNMT genes and one for N6-methyltransferase genes. The DNMT Blastps indicate there are three DNMT orthologs within the *O. tipulae* genome (Appendix Tables 4.4 and 4.5). These orthologs include one possible DNMT1 gene (g13819), one possible DNMT3B gene (g18780), and one gene that's reciprocal best match was a methyltransferase 2 gene from *Drosophila melanogaster* (g7662). This hit might also be a second possible DNMT1 ortholog as it obtained a hit to a large amount of DNMT1 genes as well. The identification of these enzyme genes supports the idea that cytosine methylation is found within the *O. tipulae* genome, despite the fact that there was no DNMT3A ortholog found.

The N6-methyltransferase Blastp analysis showed similar results. Of the nine methyltransferase genes included, eight mapped to a total of three orthologs within the *O. tipulae* genes, and those three each mapped back to the original methyltransferase gene

(Appendix Tables 4.6 and 4.7). The identification of these three possible N6-methyltransferase genes suggests that the *O. tipulae* genome is capable of adenine methylation, but further work still needs to be done in order to confirm the presence of and examine the role of adenine methylation in *O. tipulae*.

Role of Methylation on Abiotic Stress Response

Across all three stress treatments, a total of 3,123 genes were found to be upregulated under at least one of heat, freezing, or desiccation stresses (Chapter III). Only 121, or approximately 4%, of these have at least one methylation peak within their CDS (Table 4.8). In order to determine if any specific stress was more influenced by DNA methylation, the list of 121 genes was broken down into one of seven groups based on whether they were upregulated in either stress individually or any combination of the three. The data shows that neither stress is more heavily represented within the methylated lists than in the overall total list. This data also indicates that genes involved with multiple of the stress responses are also not more heavily methylated than those only involved in one. Across all comparisons, and with only ~4% of the upregulated genes containing methylated peaks, it is likely that DNA methylation is not a direct driving factor in the responses to these stresses.

Evolutionarily conserved hypermethylated genes have been shown to be involved with the housekeeping genes of transcription and translation (Sarda et al., 2012), meaning while methylation may not have a direct effect on many genes, it may play a larger downstream role if those genes are involved in the transcription and translation of multiple other genes. The DAVID functional annotation analysis of the upregulated and hypermethylated genes does not support this idea (Appendix Table 4.9). The majority of

Table 4.8 Presence of hypermethylated-CG peaks within genes upregulated due to abiotic stress. Upregulated gene counts were obtained from Chapter III of this dissertation. Upregulated peaks were obtained using MACS v1.4.2. Relative to the total upregulated gene numbers, the low numbers containing hypermethylated peaks show that DNA cytosine methylation does not play a large role in the differential expression of genes responsible for the stress responses for heat, freezing, and desiccation stresses.

Treatment cohort	Total number of upregulated genes	Upregulated genes containing at least one hypermethylated peak
Heat only	91	6
Freezing only	722	29
Desiccation only	1100	32
Heat and freezing	44	1
Heat and desiccation	109	9
Freezing and desiccation	742	26
Heat, freezing, and desiccation	315	18
Total	3,123	121

clusters are not involved with the regulation of gene expression, and those that are (Cluster 8) contain very few genes (4). In the large picture of genome-wide differential gene expression, this is likely not going to have a large effect.

Conclusions

In this chapter, genome-wide cytosine methylation levels were obtained for two species of soil nematode: *O. tipulae* and the model organism *C. elegans*. It has historically been believed that nematodes do not contain DNA cytosine methylation, but the results of the WGBS analysis contradict this, with approximately 5% and 9% of cytosines shown to be methylated in *O. tipulae* and *C. elegans*, respectively. The MethylCap-Seq analysis, which only sequenced those regions of the genome containing methylated cytosines in the CG context, supported this, as many hypermethylated regions were located within each genome.

The *O. tipulae* analysis fell in line with methylation patterns known for other invertebrate species. The peaks of the hypermethylated peaks predominantly fell within the gene body, and of those, peaks were differentially located within CDS more than introns. This data was supported by a reciprocal Blastp analysis that identified three possible DNA (5-cytosine-)-methyltransferase genes within the *O. tipulae* genome. Three N6-methyltransferase genes were also identified, indicating the possibility of adenine methylation within the genome, a feature known to exist within *C. elegans*. Further analysis would need to be performed to confirm this.

The *C. elegans* methylation pattern did not follow this pattern as hypermethylated peaks were found nearly evenly within intergenic regions and gene bodies. And of those peaks within gene bodies, peaks were found at a much higher level in introns than the

CDS. This does not follow the typical methylation pattern known for invertebrates and raises questions as to the role of cytosine methylation within the species, especially because it has historically been believed that cytosine methylation does not exist in *O. tipulae*.

The role of cytosine methylation in the genomic regulation of abiotic stress response in *O. tipulae* looks to be minimal or nonexistent. While thousands of genes were found to be upregulated in at least one of heat, freezing, or desiccation stress, very few of these genes were found to have hypermethylated peaks located within their gene body. As the increase of gene transcription is correlated with increased methylation levels in invertebrates, it would be expected that these upregulated genes would contain hypermethylated regions, but this was not the case. Because this is not the case in *O. tipulae*, either cytosine methylation plays a different role in *O. tipulae* gene expression or the genes responsible for abiotic stress response are regulated by a different molecular mechanism.

References

- Aran, D., Toperoff, G., Rosenberg, M., & Hellman, A. (2011). Replication timing-related and gene body-specific methylation of active human genes. *Hum Mol Genet*, 20(4), 670-680. doi:10.1093/hmg/ddq513
- Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., . . . Church, G. M. (2009). Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol*, 27(4), 361-368. doi:10.1038/nbt.1533
- Bonasio, R., Li, Q., Lian, J., Mutti, N. S., Jin, L., Zhao, H., . . . Reinberg, D. (2012). Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr Biol*, 22(19), 1755-1764. doi:10.1016/j.cub.2012.07.042
- Boyes, J., & Bird, A. (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *The EMBO Journal*, 11(1), 327-333.
- Brenner, S. (1974). The Genetics of *Caenorhabditis elegans*. *Genetics*, 77, 71-94.
- Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., . . . Jacobsen, S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A*, 107(19), 8689-8694. doi:10.1073/pnas.1002720107
- Flores, K., Wolschin, F., Corneveaux, J. J., Allen, A. N., Huentelman, M. J., & Amdam, G. V. (2012). Genome-wide association between DNA methylation and alternative splicing in an invertebrate. *BMC Genomics*, 13(480), 1-9.
- Gao, F., Liu, X., Wu, X.-P., Wang, X.-L., Gong, D., Lu, H., . . . Liu, M. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. *Genome Biology*, 13(R100), 1-13.
- Greer, E. L., Blanco, M. A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corrales, D., . . . Shi, Y. (2015). DNA Methylation on N6-Adenine in *C. elegans*. *Cell*, 161(4), 868-878. doi:10.1016/j.cell.2015.04.005
- Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M. Q., . . . Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, 14(774), 108.
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1), 1-13. doi:10.1093/nar/gkn923
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protocols*, 4(1), 44-57.

- Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., . . . Jaenisch, R. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genetics*, 27, 31-39.
- Keller, T. E., Han, P., & Yi, S. V. (2016). Evolutionary Transition of Promoter and Gene Body DNA Methylation across Invertebrate-Vertebrate Boundary. *Mol Biol Evol*, 33(4), 1019-1028. doi:10.1093/molbev/msv345
- Krueger, F., & Andrews, S. R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, 27(11), 1571-1572. doi:10.1093/bioinformatics/btr167
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), 357-359. doi:10.1038/nmeth.1923
- Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., . . . Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, 462(7271), 315-322. doi:10.1038/nature08514
- Mayer, C. (2006-2010). Phobos 3.3.11. Retrieved from http://www.rub.de/ecoevo/cm/cm_phobos.htm
- Razin, A., & Cedar, H. (1991). DNA Methylation and Gene Expression. *Microbiological Reviews*, 55(3), 451-458.
- Sarda, S., Zeng, J., Hunt, B. G., & Yi, S. V. (2012). The evolution of invertebrate gene body methylation. *Mol Biol Evol*, 29(8), 1907-1916. doi:10.1093/molbev/mss062
- Stanke, M., & Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Research*, 33(Web Server issue), W465-467. doi:10.1093/nar/gki458
- Suzuki, M. M., & Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet*, 9(6), 465-476. doi:10.1038/nrg2341
- Wan, J., Oliver, V. F., Wang, G., Zhu, H., Zack, D. J., Merbs, S. L., & Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics*, 16, 49. doi:10.1186/s12864-015-1271-4
- Wang, X., Wheeler, D., Avery, A., Rago, A., Choi, J. H., Colbourne, J. K., . . . Werren, J. H. (2013). Function and evolution of DNA methylation in *Nasonia vitripennis*. *PLoS Genet*, 9(10), e1003872. doi:10.1371/journal.pgen.1003872
- Wu, T. P., Wang, T., Seetin, M. G., Lai, Y., Zhu, S., Lin, K., . . . Xiao, A. Z. (2016). DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature*, 532(7599), 329-333. doi:10.1038/nature17640
- Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation. *Science*, 328, 916-919.

- Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., . . . Chen, D. (2015). N6-methyladenine DNA modification in *Drosophila*. *Cell*, *161*(4), 893-906. doi:10.1016/j.cell.2015.04.018
- Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., . . . Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol*, *9*(9), R137. doi:10.1186/gb-2008-9-9-r137

CHAPTER V

EPILOGUE

The fields of genetics and ecology are commonly seen as two independent aspects within the subject of biology, but my research interests lie at the intersection of the two. The molecular makeup of an organism's genome has large-scale effects on the way it reacts, responds, and evolves with its environment, and it was through this lens that I was introduced to the Nematode phylum. Incredible diversity across the phylum in both ecological niches and genomic diversity opens the door for a variety of questions integrating the two fields, one of which being how the responses to abiotic stresses are regulated on a genome-wide level. Exciting and fast-paced advancements in high-throughput, next generation sequencing and bioinformatics have also allowed for these questions to be answered.

In this dissertation, I was able to use genomics methods to answer ecological questions. In this epilogue, I will detail the conclusions I have made in regards to the ways the genome of *Oscheius tipulae* is able to regulate its responses to abiotic stress by touching on three major focus points: the genome, the transcriptome, and the methylome. As a whole, this project begins to establish *O. tipulae* as a satellite model organism within the Nematoda phylum and provides an in-depth analysis of its "-omics", one of the first performed on a nematode species.

The *Oscheius tipulae* Genome

Genome sequencing is highly prevalent in the Nematoda phylum. The model organism *Caenorhabditis elegans* has one of the most detailed and frequently studied genomes across all animal taxa, but the fact that nematode genomes are incredibly genetically diverse means there is a need to establish satellite model organisms for nematode genomic studies. The research provided in Chapter II contributes to this process by providing a draft sequence and annotation of the KJO strain of *O. tipulae*, a strain obtained from the Konza Prairie outside of Manhattan, Kansas, USA. This work also helps build upon the recently sequenced genome of the CEW1 *O. tipulae* strain.

Results showed that the genome is shorter than previously estimated, with a length of approximately 60 Mb, but the approximate 20,000 predicted protein-coding genes falls in line with early estimates. While this amount of protein-coding genes is slightly higher, these statistics, along with the presence of large-scale synteny between the two, provide support for both this and the CEW1 *O. tipulae* genome. The individual gene annotation proved that despite being located within the same taxonomic family, *O. tipulae* and *C. elegans* genomes are vastly different. Approximately one-third of the *O. tipulae* genes have no ortholog within the *C. elegans* genome, and this variation was also supported by the lack of large-scale synteny between the two genomes. These data both further prove the vast genetic diversity found within nematodes and begin to solidify *O. tipulae* as a satellite model organism in nematode genomic and genetic studies.

Genomic Regulation of Abiotic Stress in *O. tipulae*

Due to its ability to survive in a multitude of ecosystems, *O. tipulae* individuals face a variety of abiotic stresses, and most of the studies looking into stress response have

been on the level of individual genes or individual molecular compounds. The work done in Chapter III utilized advancements in high-throughput next generation sequencing to answer the questions of stress response on a genome-wide level. The results of this RNA-Seq study show that while nematodes try to reduce the complexity of their genomes by sharing generalized stress response genes for similar molecular products, they also utilize stress-specific genes for both similar and unique molecular functions necessary for stress response. It also showed that when it comes to large-scale transcription patterns, desiccation and freezing responses are more commonly related to one another than they are to the heat response or the control, matching the fact that these two stresses pose the most similar biological threats to the organism.

This study used quickly growing genomics methods to answer an ecological question, providing one of the first large-scale analyses into genome-wide transcription patterns in nematodes. These data also build upon the results obtained in Chapter II in order to further strengthen the role of *O. tipulae* as a satellite model organism within the nematode phylum.

The Presence and Role of DNA Methylation

DNA methylation was assumed to be absent within nematodes due to early studies indicating it was absent within the *C. elegans* genome, but recent studies on various nematode species have begun to doubt this previously held belief. The data from Chapter IV continue this trend and cast further doubt on this belief by not only showing the existence of DNA cytosine methylation in *O. tipulae*, but also in *C. elegans*. Whole-genome bisulfite sequencing and MethylCap-Seq analyses showed that cytosine

methylation is present in each genome, but they vary in frequency and location. In *C. elegans*, it was shown that hypermethylated regions tend to fall predominantly within intergenic regions and introns, where in the *O. tipulae* genome, hypermethylated regions mainly fell within transcribed exons. Both (5-cytosine-)-DNA methyltransferase and N6-methyltransferase gene orthologs were also identified within the *O. tipulae* genome, further supporting the idea that cytosine methylation is present and indicating that adenine methylation may also be present as well.

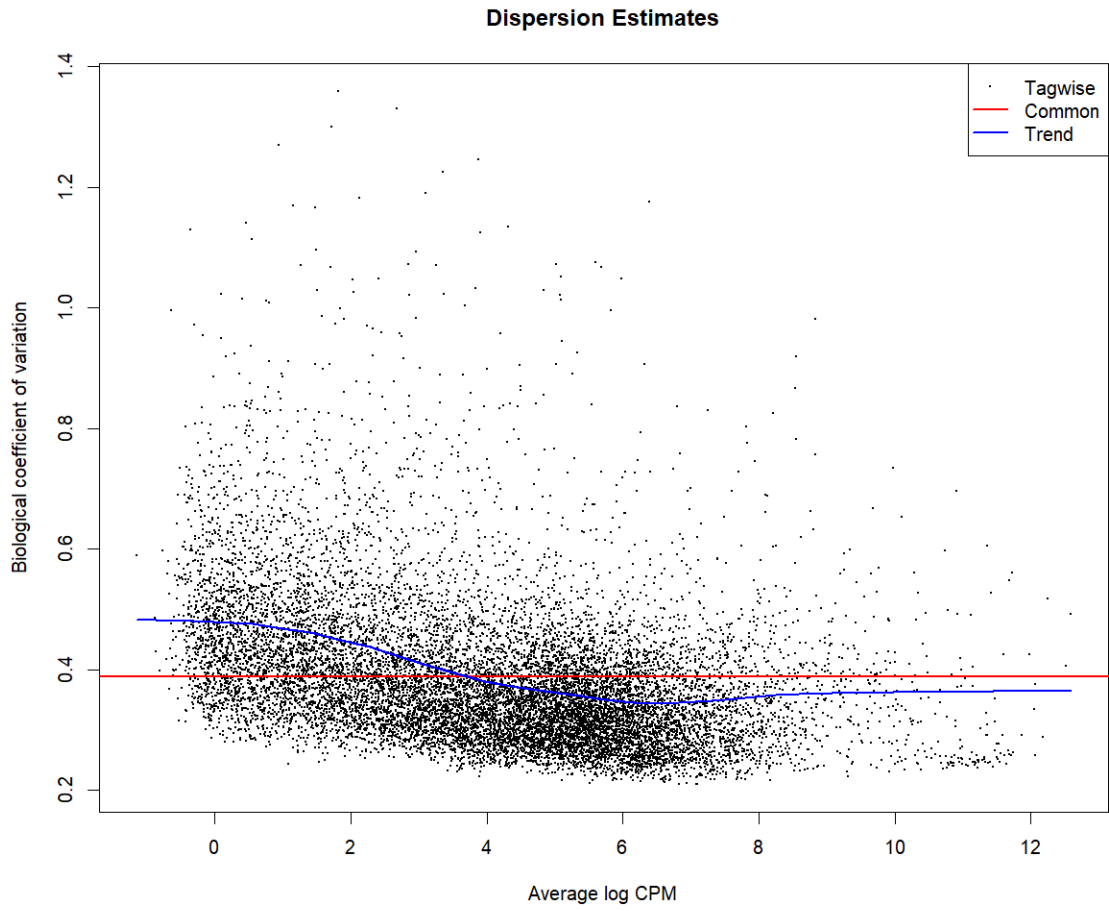
These data contradict the previously held belief that *C. elegans* does not contain methylated cytosines within its genome, while also providing *O. tipulae* as another example of a nematode species that does. The *O. tipulae* methylation patterns also follow known trends within invertebrates, indicating these methylated genes may be utilized at higher rates when necessary. Regarding abiotic stress response, this is not the case, as very few of the genes upregulated under at least one stress were found to also include hypermethylated regions. The *C. elegans* results do not follow the known invertebrate methylation scheme, meaning the role of the methylated cytosines within the genome of *C. elegans* is still unknown, and further work needs to be done in order to clarify this.

Conclusion

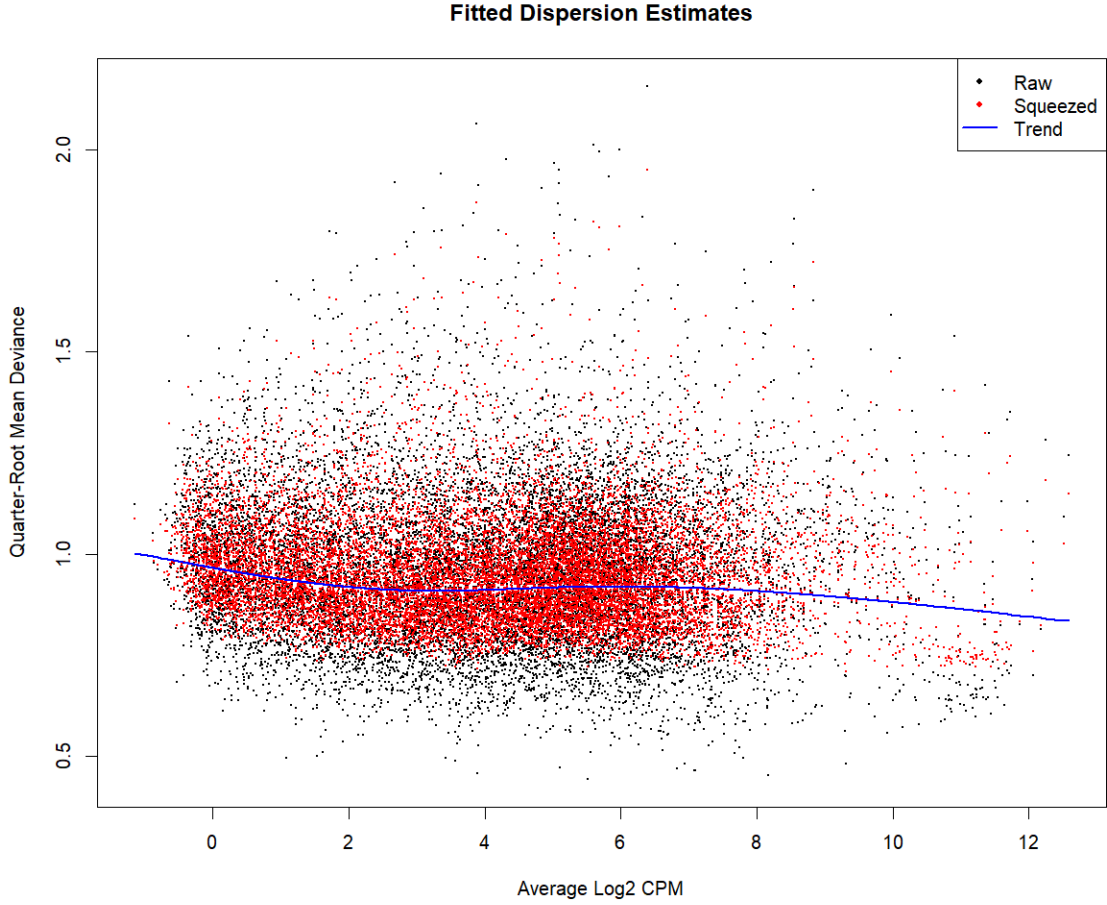
In this dissertation, the genome of the KJO strain of *O. tipulae* was sequenced, annotated, and compared to its closest model organism and another strain within the species. Genome-wide transcription patterns were assessed in order to further understand the ways in which the genome controls abiotic stress response in soil nematodes, and DNA methylation levels were located, measured, and used to determine what kind of role

methylated cytosines play in the genomes of *O. tipulae* and *C. elegans*. My work integrated the fields of ecology and genomics and used rapidly-growing next generation sequencing technologies and bioinformatics programs to do so. My hopes are that my future work will continue to remain integrative and evolve with research landscape around it just as this project did.

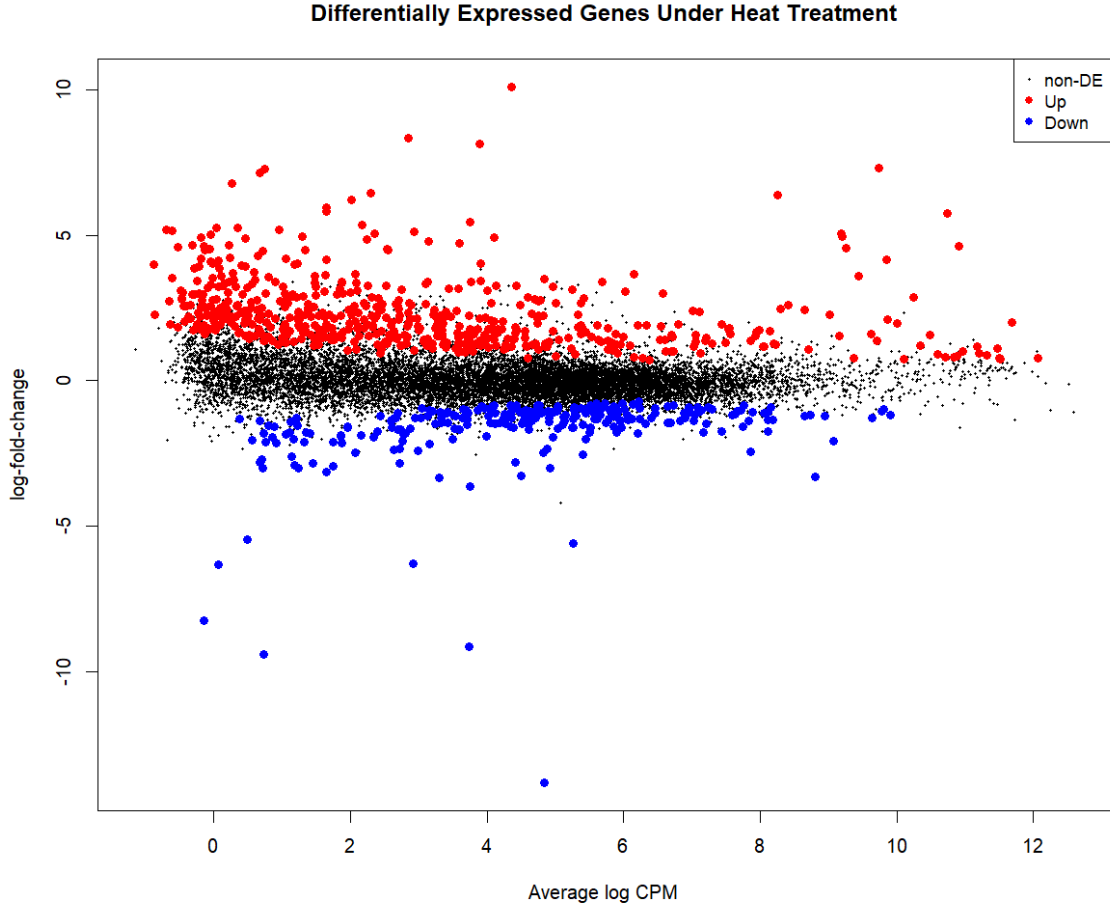
APPENDICES



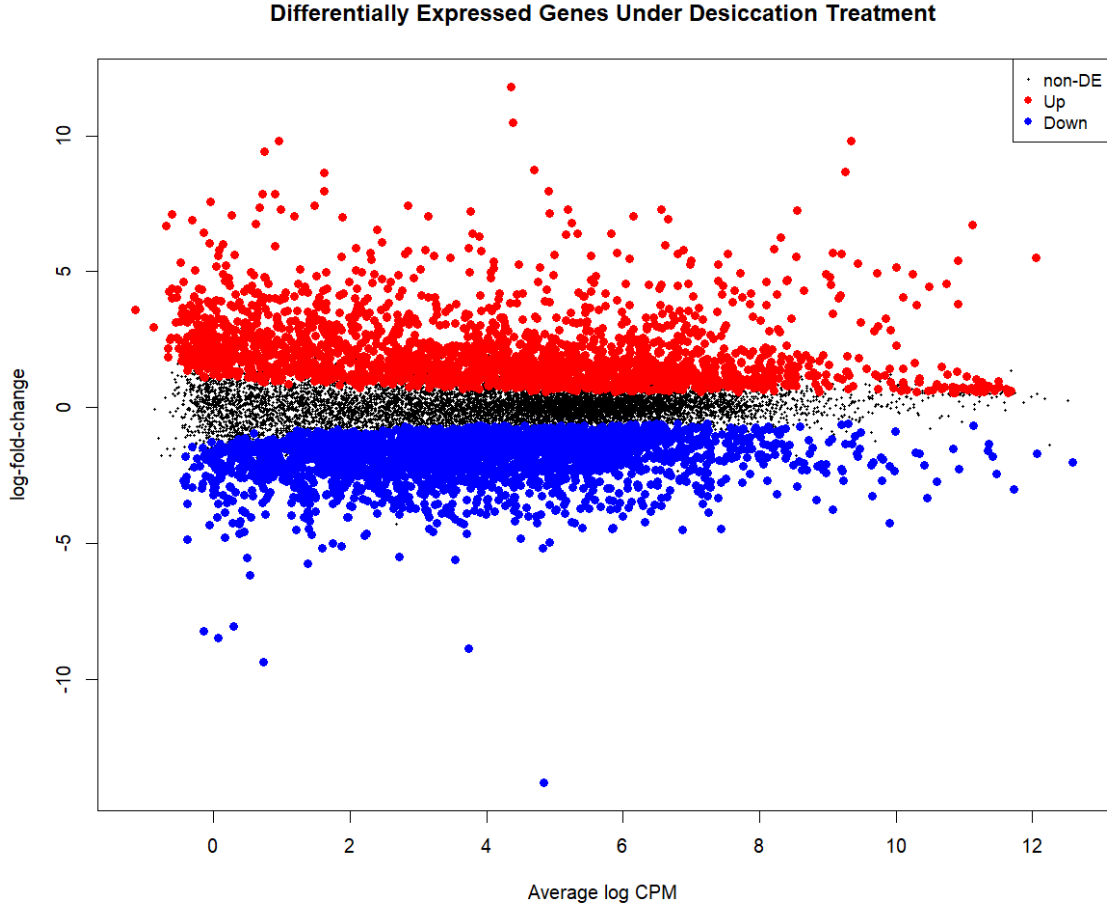
Appendix Figure 3.3 RNA-Seq dispersion estimates. Each data point represents one gene.



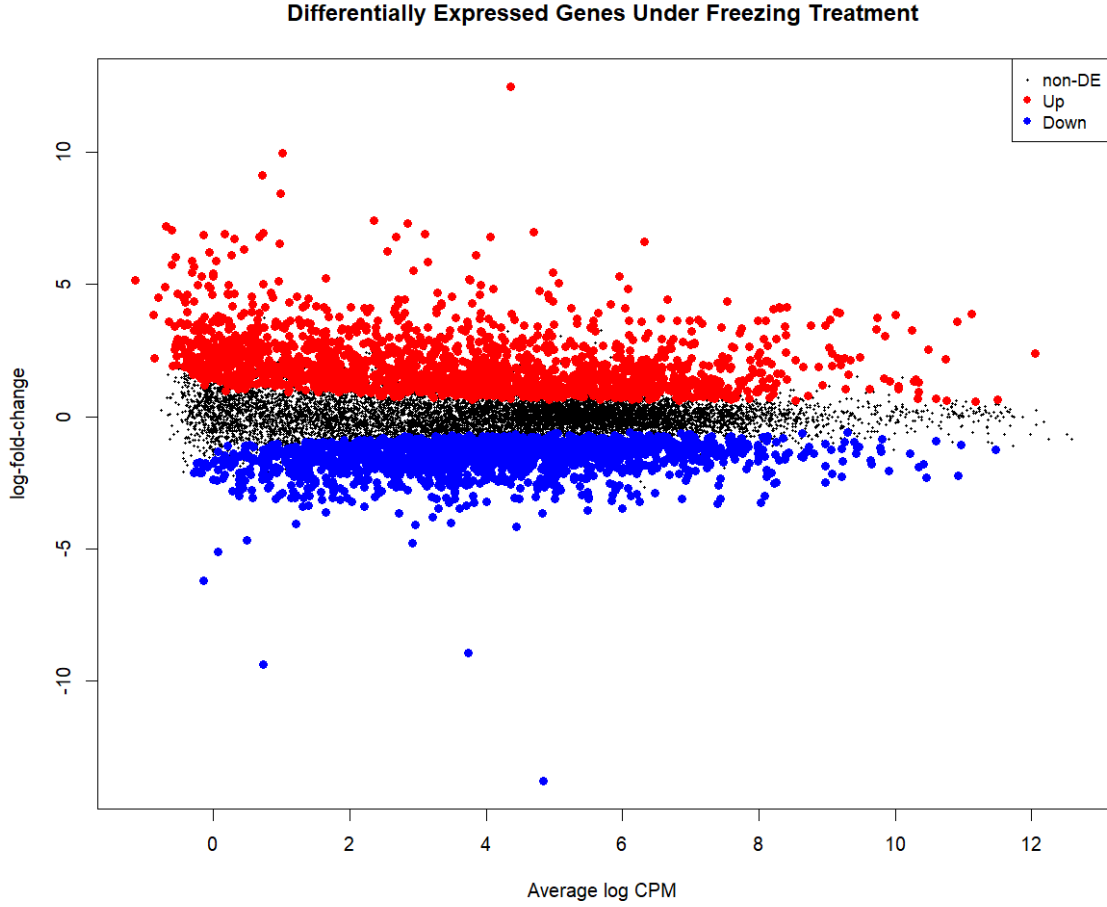
Appendix Figure 3.4 RNA-Seq fitted dispersions. Each data point represents a unique gene either before (black) or after (red) normalization using the edgeR package in R.



Appendix Figure 3.7 Differentially expressed genes under the heat treatment. Average log counts per million plotted against log-fold change for each gene from the heat treatment RNA-Seq analysis in order to assess differential expression. Red indicates significant upregulation, and blue indicates significant downregulation. Significance was determined by $P < 0.05$.



Appendix Figure 3.8 Differentially expressed genes under the desiccation treatment. Average log counts per million plotted against log-fold change for each gene from the desiccation treatment RNA-Seq analysis in order to assess differential expression. Red indicates significant upregulation, and blue indicates significant downregulation. Significance was determined by $P < 0.05$.



Appendix Figure 3.9 Differentially expressed genes under the freezing treatment. Average log counts per million plotted against log-fold change for each gene from the freezing treatment RNA-Seq analysis in order to assess differential expression. Red indicates significant upregulation, and blue indicates significant downregulation. Significance was determined by $P < 0.05$.

Appendix Table 3.2 DAVID functional annotation results for all genes upregulated under at least one abiotic stress. The three stresses were heat, freezing, and desiccation stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 12.484468067025665				
GOTERM_BP_4	larval development	386	3.28E-14	1.3919	5.34E-11
GOTERM_BP_4	post-embryonic development	387	7.66E-14	1.383952	1.25E-10
GOTERM_BP_DIRECT	nematode larval development	336	1.40E-11	1.383343	2.31E-08
Annotation Cluster 2	Enrichment Score: 9.120018765946394				
GOTERM_CC_4	ribosomal subunit	47	2.25E-12	3.046376	3.15E-09
KEGG_PATHWAY	Ribosome	47	7.56E-11	2.691489	8.54E-08
GOTERM_CC_4	ribosome	50	4.32E-09	2.412614	6.06E-06
GOTERM_CC_DIRECT	ribosome	46	8.44E-09	2.47099	1.17E-05
GOTERM_MF_DIRECT	structural constituent of ribosome	52	4.05E-08	2.197666	6.19E-05
Annotation Cluster 3	Enrichment Score: 6.6811937985733145				
GOTERM_CC_4	proteasome complex	19	1.04E-07	4.12557	1.46E-04
GOTERM_CC_DIRECT	proteasome complex	18	1.39E-07	4.243657	1.92E-04
KEGG_PATHWAY	Proteasome	18	6.25E-07	3.824748	7.07E-04
Annotation Cluster 4	Enrichment Score: 6.056111359554534				
GOTERM_BP_4	sex differentiation	185	8.37E-08	1.440536	1.36E-04
GOTERM_BP_4	animal organ development	213	1.70E-07	1.385202	2.77E-04
GOTERM_BP_4	reproductive structure development	169	1.37E-06	1.410997	0.002228
GOTERM_BP_4	reproductive system development	169	1.37E-06	1.410997	0.002228
GOTERM_BP_4	development of primary sexual characteristics	165	4.15E-06	1.393191	0.006752
GOTERM_BP_4	gonad development	165	4.15E-06	1.393191	0.006752

Appendix Table 3.2 cont.

Annotation Cluster 5		Enrichment Score: 4.947380928641273			
GOTERM_MF_4	ribonucleotide binding	207	2.99E-06	1.328659	0.003954
GOTERM_MF_4	purine nucleoside binding	202	5.96E-06	1.322863	0.007889
GOTERM_MF_4	purine ribonucleoside triphosphate binding	202	5.96E-06	1.322863	0.007889
GOTERM_MF_4	purine ribonucleotide binding	203	6.00E-06	1.321256	0.007937
GOTERM_MF_4	purine nucleotide binding	203	6.15E-06	1.320099	0.008143
GOTERM_MF_4	ribonucleoside binding	202	7.12E-06	1.319373	0.00942
GOTERM_MF_DIRECT	ATP binding	168	3.10E-05	1.337818	0.047458
GOTERM_MF_4	nucleotide binding	232	3.05E-04	1.219333	0.402818
Annotation Cluster 6		Enrichment Score: 4.155874043966426			
GOTERM_BP_4	male sex differentiation	26	3.16E-05	2.424842	0.051422
GOTERM_BP_4	nematode male tail tip morphogenesis	23	8.45E-05	2.451489	0.137356
GOTERM_BP_DIRECT	nematode male tail tip morphogenesis	23	8.53E-05	2.450953	0.140756
GOTERM_BP_4	male anatomical structure morphogenesis	23	1.04E-04	2.42006	0.169637
Annotation Cluster 7		Enrichment Score: 3.8680091552511175			
INTERPRO	Chaperonin TCP-1, conserved site	7	5.32E-05	7.580192	0.09115
INTERPRO	TCP-1-like chaperonin intermediate domain	7	5.32E-05	7.580192	0.09115
INTERPRO	Chaperone tailless complex polypeptide 1 (TCP-1)	7	1.44E-04	6.737948	0.246437
INTERPRO	GroEL-like equatorial domain	7	1.44E-04	6.737948	0.246437
INTERPRO	Chaperonin Cpn60/TCP-1	7	3.25E-04	6.064153	0.554968
INTERPRO	GroEL-like apical domain	7	3.25E-04	6.064153	0.554968
Annotation Cluster 8		Enrichment Score: 3.4305223280146038			
GOTERM_BP_4	positive regulation of gene expression	55	1.01E-04	1.690613	0.164906
GOTERM_BP_4	positive regulation of cellular biosynthetic process	49	3.50E-04	1.668676	0.568073
GOTERM_BP_4	positive regulation of biosynthetic process	49	3.50E-04	1.668676	0.568073
GOTERM_BP_4	positive regulation of macromolecule biosynthetic process	48	4.03E-04	1.669253	0.652659
GOTERM_BP_4	positive regulation of nitrogen compound metabolic process	49	4.70E-04	1.648159	0.76114

Appendix Table 3.2 cont.

GOTERM_BP_4	positive regulation of RNA metabolic process	44	6.36E-04	1.679605	1.029231
GOTERM_BP_4	positive regulation of nucleobase-containing compound metabolic process	46	6.47E-04	1.65583	1.047816
Annotation Cluster 9	Enrichment Score: 3.3943989809655073				
INTERPRO	Proteasome B-type subunit	6	1.10E-04	8.663076	0.188751
GOTERM_CC_4	proteasome core complex	8	4.32E-04	4.963092	0.604435
INTERPRO	Proteasome, subunit alpha/beta	8	4.41E-04	4.950329	0.753501
GOTERM_CC_DIRECT	proteasome core complex	8	5.00E-04	4.849893	0.689497
GOTERM_MF_DIRECT	threonine-type endopeptidase activity	8	0.001014	4.322882	1.539564
Annotation Cluster 10	Enrichment Score: 2.957340774667185				
GOTERM_BP_4	genitalia development	122	5.81E-04	1.333254	0.941071
GOTERM_BP_4	hermaphrodite genitalia development	119	0.00122	1.312704	1.966144
GOTERM_BP_DIRECT	hermaphrodite genitalia development	117	0.001893	1.30085	3.081079

Appendix Table 3.3 DAVID functional annotation results for all genes downregulated under at least one abiotic stress. The three stresses were heat, freezing, and desiccation stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 7.225121730568072				
GOTERM_CC_DIRECT	neuronal cell body	43	1.82E-08	2.490345	2.45E-05
GOTERM_CC_4	neuronal cell body	47	7.47E-08	2.277292	1.03E-04
GOTERM_CC_4	somatodendritic compartment	53	1.56E-07	2.108068	2.15E-04
Annotation Cluster 2	Enrichment Score: 7.199224075356552				
GOTERM_CC_4	intrinsic component of plasma membrane	147	4.58E-09	1.588146	6.33E-06
GOTERM_CC_4	integral component of plasma membrane	146	5.19E-09	1.588073	7.17E-06
GOTERM_CC_DIRECT	integral component of plasma membrane	118	1.06E-05	1.471276	0.014309
Annotation Cluster 3	Enrichment Score: 5.713086678249685				
GOTERM_CC_4	M band	15	1.55E-06	4.282901	0.002138
GOTERM_CC_DIRECT	M band	15	1.56E-06	4.281573	0.002097
GOTERM_CC_4	A band	16	3.01E-06	3.876242	0.004163
Annotation Cluster 4	Enrichment Score: 4.41758980255623				
INTERPRO	Neurotransmitter-gated ion-channel transmembrane domain	29	2.76E-06	2.571383	0.004706
INTERPRO	Neurotransmitter-gated ion-channel, conserved site	28	4.36E-06	2.568326	0.007431
INTERPRO	Neurotransmitter-gated ion-channel ligand-binding	31	6.92E-06	2.3787	0.011781
INTERPRO	Neurotransmitter-gated ion-channel	29	2.54E-05	2.314245	0.043256
GOTERM_MF_DIRECT	extracellular ligand-gated ion channel activity	31	3.73E-04	1.922089	0.576322
GOTERM_MF_4	ligand-gated channel activity	35	0.003953	1.613262	5.139723
Annotation Cluster 5	Enrichment Score: 4.2359823216163415				
GOTERM_BP_DIRECT	phosphorylation	84	6.79E-08	1.795596	1.12E-04

Appendix Table 3.3 cont.

GOTERM_MF_DIRECT	kinase activity	81	2.65E-05	1.563809	0.041063
GOTERM_MF_4	kinase activity	107	0.108764	1.126592	78.43188
Annotation Cluster 6		Enrichment Score: 3.3459100830339312			
GOTERM_BP_4	muscle cell differentiation	48	6.63E-06	1.94517	0.010778
GOTERM_BP_4	myofibril assembly	42	2.89E-05	1.940722	0.046964
GOTERM_BP_4	muscle cell development	42	9.26E-05	1.850456	0.1505
GOTERM_BP_4	striated muscle myosin thick filament assembly	37	1.58E-04	1.894515	0.256099
GOTERM_BP_DIRECT	striated muscle myosin thick filament assembly	36	3.76E-04	1.834249	0.621479
GOTERM_BP_4	actin cytoskeleton organization	48	0.002897	1.521953	4.607067
GOTERM_BP_4	cellular component assembly involved in morphogenesis	45	0.003044	1.543044	4.836352
GOTERM_BP_4	organelle assembly	53	0.1841	1.15413	96.34226
Annotation Cluster 7		Enrichment Score: 3.2274848935457627			
INTERPRO	ABC transporter, conserved site	17	2.57E-04	2.713252	0.436525
INTERPRO	ABC transporter-like	18	4.26E-04	2.520049	0.723582
INTERPRO	ABC transporter, transmembrane domain, type 1	14	4.64E-04	2.940057	0.787376
GOTERM_MF_DIRECT	ATPase activity, coupled to transmembrane movement of substances	18	0.002425	2.149433	3.687852
Annotation Cluster 8		Enrichment Score: 2.8424804140174835			
GOTERM_BP_4	reproductive behavior	79	5.09E-04	1.453074	0.823921
GOTERM_BP_4	oviposition	70	0.001095	1.453326	1.766164
GOTERM_BP_DIRECT	oviposition	65	0.005328	1.385468	8.467422
Annotation Cluster 9		Enrichment Score: 2.6681694137930183			
INTERPRO	CO dehydrogenase flavoprotein-like, FAD-binding, subdomain 2	6	0.001236	5.985116	2.085461
INTERPRO	FAD-binding, type 2	6	0.001236	5.985116	2.085461
GOTERM_MF_DIRECT	oxidoreductase activity, acting on CH-OH group of donors	6	0.006476	4.298866	9.567985
Annotation Cluster 10		Enrichment Score: 2.6045260754701656			
INTERPRO	Protein kinase, ATP binding site	54	1.86E-04	1.663816	0.316698

Appendix Table 3.3 cont.

INTERPRO	Protein kinase-like domain	88	2.27E-04	1.459987	0.385854
INTERPRO	Protein kinase, catalytic domain	73	0.002222	1.407129	3.719633
GOTERM_BP_DIRECT	protein phosphorylation	77	0.007976	1.322853	12.42091
GOTERM_MF_DIRECT	protein kinase activity	73	0.126785	1.15374	87.73718

Appendix Table 3.4 DAVID functional annotation results for all genes upregulated under heat stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 5.405301263862093				
GOTERM_CC_4	ribosomal subunit	14	3.94E-07	6.151480168	4.81E-04
GOTERM_CC_DIRECT	intracellular ribonucleoprotein complex	14	2.01E-06	5.345890411	0.002328392
KEGG_PATHWAY	Ribosome	14	2.12E-06	4.941750959	0.002051098
GOTERM_CC_DIRECT	ribosome	14	4.86E-06	4.939873418	0.005626316
GOTERM_CC_4	ribosome	14	1.09E-05	4.579435236	0.013270268
GOTERM_MF_DIRECT	structural constituent of ribosome	15	1.21E-05	4.202802164	0.015241101
GOTERM_CC_4	cytosolic part	13	1.35E-05	4.87528592	0.016522458
Annotation Cluster 2	Enrichment Score: 4.985813777030337				
GOTERM_BP_4	larval development	66	7.98E-06	1.67104311	0.011608291
GOTERM_BP_4	post-embryonic development	66	1.06E-05	1.657208766	0.015434135
GOTERM_BP_DIRECT	nematode larval development	60	1.30E-05	1.72431095	0.01795754
Annotation Cluster 3	Enrichment Score: 4.514772397882341				
GOTERM_CC_4	striated muscle thin filament	6	1.25E-05	18.5931957	0.01523641
GOTERM_CC_DIRECT	striated muscle thin filament	5	2.95E-05	25.34090909	0.03413332
GOTERM_CC_4	myofilament	6	7.76E-05	13.08410068	0.094762162
Annotation Cluster 4	Enrichment Score: 4.068042626554801				
GOTERM_BP_DIRECT	translation	17	1.20E-05	3.744960212	0.016511937
GOTERM_BP_4	peptide metabolic process	22	3.73E-05	2.786296057	0.054294426
GOTERM_BP_4	cellular amide metabolic process	22	1.39E-04	2.540610633	0.202655751
GOTERM_BP_4	organonitrogen compound biosynthetic process	24	8.57E-04	2.117941873	1.23964528
Annotation Cluster 5	Enrichment Score: 2.537528520745865				
GOTERM_BP_DIRECT	endoplasmic reticulum unfolded protein response	7	0.001652632	5.492205952	2.254118427
GOTERM_BP_4	response to topologically incorrect protein	8	0.002003843	4.475780074	2.876081124

Appendix Table 3.4 cont.

GOTERM_BP_4	endoplasmic reticulum unfolded protein response	7	0.002096011	5.238697132	3.006497026
GOTERM_BP_4	response to unfolded protein	7	0.00410067	4.58385999	5.803011139
GOTERM_BP_4	response to endoplasmic reticulum stress	8	0.004102131	3.940216646	5.805021977
GOTERM_BP_4	cellular response to topologically incorrect protein	7	0.00509954	4.38456173	7.168240252
Annotation Cluster 6		Enrichment Score: 2.17697124005364			
GOTERM_MF_4	ribonucleotide binding	36	0.001217224	1.702716122	1.355008759
GOTERM_MF_4	purine nucleoside binding	34	0.003300504	1.640736988	3.635363524
GOTERM_MF_4	purine ribonucleoside triphosphate binding	34	0.003300504	1.640736988	3.635363524
GOTERM_MF_4	ribonucleoside binding	34	0.003443594	1.636407867	3.790213376
GOTERM_MF_4	purine ribonucleotide binding	34	0.003642671	1.630671117	4.005274484
GOTERM_MF_4	purine nucleotide binding	34	0.00369396	1.629243209	4.06061034
GOTERM_MF_4	nucleotide binding	37	0.017643056	1.432955384	18.07682556
GOTERM_MF_DIRECT	ATP binding	28	0.034908099	1.478206628	36.00989912
GOTERM_MF_DIRECT	nucleotide binding	28	0.067499952	1.386274725	58.44219529
Annotation Cluster 7		Enrichment Score: 1.813092422353014			
GOTERM_BP_4	protein complex subunit organization	16	0.012484647	2.021953279	16.70472961
GOTERM_BP_4	protein complex assembly	15	0.01657945	2.014883512	21.5909603
GOTERM_BP_4	macromolecular complex assembly	18	0.017570186	1.842383716	22.73238389
Annotation Cluster 8		Enrichment Score: 1.6701628201625516			
INTERPRO	Peptidase M12A, astacin	5	0.005518436	6.959615385	7.575450771
INTERPRO	Peptidase, metallopeptidase	5	0.009773431	5.923076923	13.04836219
INTERPRO	Metallopeptidase, catalytic domain	6	0.023626785	3.671005917	28.85044101
GOTERM_MF_DIRECT	metallopeptidase activity	7	0.050300561	2.619955777	47.71451029
GOTERM_MF_DIRECT	metalloendopeptidase activity	6	0.069557722	2.710996711	59.57985549
Annotation Cluster 9		Enrichment Score: 1.553280334357495			
GOTERM_MF_DIRECT	neuropeptide receptor binding	4	0.002930166	13.37425044	3.619874138
INTERPRO	FMRFamide-related peptide-like	3	0.021841219	12.84852071	26.97552754
GOTERM_MF_DIRECT	neuropeptide hormone activity	3	0.022660896	12.53835979	25.02375472
GOTERM_MF_4	hormone activity	4	0.055943336	4.560283688	47.52562942

Appendix Table 3.4 cont.

GOTERM_BP_DIRECT	neuropeptide signaling pathway	5	0.076119293	3.112818591	66.42304047
GOTERM_MF_4	G-protein coupled receptor binding	4	0.077558372	3.979883946	59.51705897
Annotation Cluster 10	Enrichment Score: 1.5294418208525264				
GOTERM_BP_4	reproductive system development	27	0.018382337	1.582800659	23.65649828
GOTERM_BP_4	reproductive structure development	27	0.018382337	1.582800659	23.65649828
GOTERM_BP_4	sex differentiation	28	0.023729119	1.530852672	29.48813389
GOTERM_BP_4	animal organ development	32	0.026562715	1.461189691	32.40780422
GOTERM_BP_4	gonad development	26	0.027965565	1.541427345	33.81122615
GOTERM_BP_4	development of primary sexual characteristics	26	0.027965565	1.541427345	33.81122615
GOTERM_BP_4	system development	35	0.118114857	1.254290047	83.93734573

Appendix Table 3.5 DAVID functional annotation results for all genes upregulated under desiccation stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 13.454167310828879			
GOTERM_CC_4	ribosomal subunit	46	1.60E-17	4.258874429	2.17E-14
KEGG_PATHWAY	Ribosome	46	8.78E-15	3.48756368	9.80E-12
GOTERM_CC_4	ribosome	49	4.30E-14	3.377266848	5.85E-11
GOTERM_CC_DIRECT	ribosome	45	2.22E-13	3.453246876	2.96E-10
GOTERM_CC_4	cytosolic part	44	3.42E-13	3.476921024	4.66E-10
GOTERM_MF_DIRECT	structural constituent of ribosome	50	4.10E-12	2.932187556	6.12E-09
Annotation Cluster 2		Enrichment Score: 10.500392446528114			
GOTERM_BP_4	larval development	285	6.94E-12	1.427901916	1.11E-08
GOTERM_BP_4	post-embryonic development	286	1.12E-11	1.42104921	1.79E-08
GOTERM_BP_DIRECT	nematode larval development	251	4.05E-10	1.427206235	6.50E-07
Annotation Cluster 3		Enrichment Score: 4.68712400436698			
INTERPRO	Chaperonin TCP-1, conserved site	7	7.83E-06	10.52909393	0.013111068
INTERPRO	TCP-1-like chaperonin intermediate domain	7	7.83E-06	10.52909393	0.013111068
INTERPRO	GroEL-like equatorial domain	7	2.18E-05	9.359194606	0.036551155
INTERPRO	Chaperone tailless complex polypeptide 1 (TCP-1)	7	2.18E-05	9.359194606	0.036551155
INTERPRO	Chaperonin Cpn60/TCP-1	7	5.07E-05	8.423275145	0.084914935
INTERPRO	GroEL-like apical domain	7	5.07E-05	8.423275145	0.084914935
Annotation Cluster 4		Enrichment Score: 3.129436075464537			
GOTERM_CC_DIRECT	integral component of endoplasmic reticulum membrane	10	4.93E-04	4.041577825	0.653862864
GOTERM_CC_4	intrinsic component of endoplasmic reticulum membrane	12	5.59E-04	3.383532649	0.758385055
GOTERM_CC_4	integral component of endoplasmic reticulum membrane	11	0.001484333	3.24926548	2.000773038
Annotation Cluster 5		Enrichment Score: 3.089742589427437			
GOTERM_BP_4	male sex differentiation	19	5.23E-04	2.462049062	0.831966075

Appendix Table 3.5 cont.

GOTERM_BP_4	nematode male tail tip morphogenesis	17	8.76E-04	2.517584003	1.390961049
GOTERM_BP_DIRECT	nematode male tail tip morphogenesis	17	9.41E-04	2.501949318	1.498727462
GOTERM_BP_4	male anatomical structure morphogenesis	17	0.001014781	2.485307285	1.609042821
Annotation Cluster 6	Enrichment Score: 3.0803791540807115				
GOTERM_BP_4	sex differentiation	127	1.29E-04	1.374006807	0.206679972
GOTERM_BP_4	animal organ development	144	5.15E-04	1.30115463	0.819901771
GOTERM_BP_4	reproductive system development	115	8.77E-04	1.334043824	1.392634411
GOTERM_BP_4	reproductive structure development	115	8.77E-04	1.334043824	1.392634411
GOTERM_BP_4	gonad development	111	0.002532686	1.302214384	3.970601659
GOTERM_BP_4	development of primary sexual characteristics	111	0.002532686	1.302214384	3.970601659
Annotation Cluster 7	Enrichment Score: 2.6670390086084503				
GOTERM_BP_4	response to other organism	25	0.00141075	1.993561994	2.230285705
GOTERM_BP_4	defense response to other organism	24	0.002338911	1.954829932	3.672112377
GOTERM_BP_4	defense response to bacterium	21	0.003022868	2.029378531	4.721805983
Annotation Cluster 8	Enrichment Score: 2.585191773298422				
GOTERM_MF_DIRECT	neuropeptide receptor binding	8	2.24E-04	5.59852344	0.333597829
GOTERM_MF_DIRECT	neuropeptide hormone activity	7	4.08E-04	6.123385013	0.606728379
INTERPRO	FMRFamide-related peptide-like	6	0.00285959	5.553807788	4.680384692
GOTERM_MF_4	hormone activity	8	0.174684036	1.728494624	91.55416271
Annotation Cluster 9	Enrichment Score: 2.55732136384281				
GOTERM_BP_DIRECT	positive regulation of vulval development	16	9.63E-04	2.590253411	1.533951085
GOTERM_BP_4	positive regulation of nematode larval development	17	0.004700021	2.153932981	7.250585003
GOTERM_BP_4	positive regulation of post-embryonic development	17	0.004700021	2.153932981	7.250585003
Annotation Cluster 10	Enrichment Score: 2.5248968998303023				
GOTERM_MF_4	purine nucleoside binding	138	0.001854296	1.262075439	2.36095558
GOTERM_MF_4	purine ribonucleoside triphosphate binding	138	0.001854296	1.262075439	2.36095558
GOTERM_MF_4	ribonucleotide binding	140	0.002002574	1.254913988	2.547511149
GOTERM_MF_4	ribonucleoside binding	138	0.002118124	1.258745425	2.692660743
GOTERM_MF_DIRECT	ATP binding	117	0.002230606	1.292816926	3.276565593

Appendix Table 3.5 cont.

GOTERM_MF_4	purine ribonucleotide binding	138	0.002356765	1.254332645	2.99180384
GOTERM_MF_4	purine nucleotide binding	138	0.00236966	1.253234281	3.007943854
GOTERM_MF_4	nucleotide binding	156	0.034793787	1.144990069	36.61115027

Appendix Table 3.6 DAVID functional annotation results for all genes upregulated under freezing stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 9.273402807530477				
KEGG_PATHWAY	Proteasome	17	1.96E-10	7.111049876	2.11E-07
GOTERM_CC_DIRECT	proteasome complex	17	4.18E-10	6.867753623	5.56E-07
GOTERM_CC_4	proteasome complex	17	1.84E-09	6.352349229	2.50E-06
Annotation Cluster 2	Enrichment Score: 4.323987700286669				
INTERPRO	Proteasome B-type subunit	6	6.09E-06	15.58234661	0.009777261
INTERPRO	Proteasome, subunit alpha/beta	8	1.00E-05	8.904198062	0.016089628
GOTERM_CC_4	proteasome core complex	8	1.31E-05	8.540973753	0.017722318
GOTERM_CC_DIRECT	proteasome core complex	8	1.57E-05	8.310559006	0.02084507
GOTERM_MF_DIRECT	threonine-type endopeptidase activity	8	2.47E-05	7.749029225	0.035777679
GOTERM_BP_DIRECT	proteolysis involved in cellular protein catabolic process	9	0.036732809	2.334009903	44.6139217
Annotation Cluster 3	Enrichment Score: 3.6205363151007397				
GOTERM_MF_DIRECT	nucleotide binding	112	7.22E-06	1.499318587	0.010442924
GOTERM_MF_4	ribonucleotide binding	119	1.66E-04	1.368288976	0.210632399
GOTERM_MF_4	purine ribonucleotide binding	117	2.23E-04	1.364157273	0.281968341
GOTERM_MF_4	purine nucleotide binding	117	2.28E-04	1.362962739	0.288695834
GOTERM_MF_4	purine nucleoside binding	116	2.67E-04	1.360846561	0.338449713
GOTERM_MF_4	purine ribonucleoside triphosphate binding	116	2.67E-04	1.360846561	0.338449713
GOTERM_MF_4	ribonucleoside binding	116	2.96E-04	1.357255937	0.37419345
GOTERM_MF_4	nucleotide binding	135	0.001376461	1.27103009	1.731700628
GOTERM_MF_DIRECT	ATP binding	95	0.001470145	1.356080114	2.106655011
Annotation Cluster 4	Enrichment Score: 3.5743098184461526				
GOTERM_BP_4	cuticle development	23	7.70E-05	2.556979263	0.122270889
GOTERM_BP_DIRECT	collagen and cuticulin-based cuticle development	19	4.41E-04	2.529084477	0.693236608

Appendix Table 3.6 cont.

GOTERM_BP_4	collagen and cuticulin-based cuticle development	19	5.58E-04	2.479641516	0.882821794
Annotation Cluster 5	Enrichment Score: 3.429793702560316				
INTERPRO	Protein kinase, catalytic domain	49	4.71E-05	1.844287401	0.075560651
INTERPRO	Protein kinase-like domain	52	2.67E-04	1.684578012	0.42847813
GOTERM_BP_DIRECT	protein phosphorylation	49	3.03E-04	1.702144251	0.477923334
GOTERM_MF_DIRECT	protein kinase activity	49	7.55E-04	1.628625628	1.087383454
GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	58	0.002460897	1.472519084	3.076508882
Annotation Cluster 6	Enrichment Score: 3.4067832310108113				
GOTERM_BP_4	larval development	193	1.07E-04	1.272676983	0.169118365
GOTERM_BP_4	post-embryonic development	193	1.73E-04	1.262140659	0.274055835
GOTERM_BP_DIRECT	nematode larval development	162	0.003272464	1.222632383	5.043286654
Annotation Cluster 7	Enrichment Score: 3.012729992019925				
GOTERM_BP_4	response to other organism	22	4.87E-04	2.30897793	0.770619556
GOTERM_BP_4	defense response to other organism	21	9.37E-04	2.251253482	1.479050914
GOTERM_BP_4	defense response to bacterium	18	0.002006979	2.289410321	3.14155535
Annotation Cluster 8	Enrichment Score: 2.7408314257325883				
GOTERM_CC_4	cytoplasmic region	18	0.001057581	2.423789849	1.423648017
GOTERM_CC_DIRECT	cell cortex	15	0.001275656	2.660392365	1.680842054
GOTERM_CC_4	cell cortex	16	0.004440771	2.256106274	5.852790808
Annotation Cluster 9	Enrichment Score: 2.5362772626226286				
GOTERM_BP_4	positive regulation of RNA metabolic process	27	0.002199603	1.884770357	3.438158723
GOTERM_BP_4	positive regulation of nucleobase-containing compound metabolic process	28	0.002472266	1.843131506	3.856550041
GOTERM_BP_4	positive regulation of cellular biosynthetic process	29	0.002744058	1.805984813	4.27191373
GOTERM_BP_4	positive regulation of biosynthetic process	29	0.002744058	1.805984813	4.27191373
GOTERM_BP_4	positive regulation of gene expression	31	0.003276106	1.742543269	5.080149552
GOTERM_BP_4	positive regulation of nitrogen compound metabolic process	29	0.00327801	1.783780081	5.083031303
GOTERM_BP_4	positive regulation of macromolecule biosynthetic process	28	0.004007421	1.780652472	6.180661638
Annotation Cluster 10	Enrichment Score: 2.4882410171608873				

Appendix Table 3.6 cont.

INTERPRO	Alpha tubulin	5	0.002578382	7.791173305	4.060996473
INTERPRO	Tubulin/FtsZ, 2-layer sandwich domain	6	0.00258	5.843379978	4.063495961
INTERPRO	Tubulin, C-terminal	6	0.00258	5.843379978	4.063495961
INTERPRO	Tubulin, conserved site	6	0.00258	5.843379978	4.063495961
INTERPRO	Tubulin/FtsZ, GTPase domain	6	0.00258	5.843379978	4.063495961
GOTERM_BP_DIRECT	microtubule-based process	8	0.00292818	4.011039241	4.524138145
INTERPRO	Tubulin/FtsZ, C-terminal	6	0.003464231	5.499651744	5.42015664
INTERPRO	Tubulin	6	0.003464231	5.499651744	5.42015664
GOTERM_MF_DIRECT	structural constituent of cytoskeleton	7	0.010050625	3.650984924	13.60053023

Appendix Table 3.7 DAVID functional annotation results for all genes upregulated under only heat stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 2.0429919180501837				
GOTERM_MF_DIRECT	peptidase activity	6	0.002640291	6.045280612	2.797909797
GOTERM_MF_4	peptidase activity, acting on L-amino acid peptides	7	0.006710432	3.926717557	6.383229806
GOTERM_BP_DIRECT	proteolysis	6	0.011328748	4.285135135	12.42562803
GOTERM_BP_4	proteolysis	7	0.033531119	2.781415929	35.57327167
Annotation Cluster 2	Enrichment Score: 1.6320695396248888				
GOTERM_CC_4	contractile fiber part	4	0.009076846	8.955462185	9.216840141
GOTERM_CC_4	striated muscle dense body	3	0.034046296	10.0034418	30.74249956
GOTERM_CC_DIRECT	striated muscle dense body	3	0.041094767	9.074202128	32.74148185
Annotation Cluster 3	Enrichment Score: 1.1904674750108832				
INTERPRO	Peptidase M12A, astacin	3	0.011161554	18.40169492	12.47541569
INTERPRO	Peptidase, metallopeptidase	3	0.015194742	15.66101695	16.62071693
INTERPRO	Metallopeptidase, catalytic domain	3	0.051485892	8.088657106	46.6083621
GOTERM_MF_DIRECT	metalloendopeptidase activity	3	0.082973441	6.0997426	60.53490806
GOTERM_MF_DIRECT	metallopeptidase activity	3	0.113938626	5.052771855	72.70491943
GOTERM_MF_DIRECT	zinc ion binding	3	0.871924464	0.846339286	99.99999997
Annotation Cluster 4	Enrichment Score: 1.003866263771454				
GOTERM_BP_DIRECT	cell migration	3	0.076770815	6.393145161	60.55313643
GOTERM_BP_4	cell migration	4	0.095942168	3.556435644	72.75087668
GOTERM_BP_4	cell motility	4	0.132189064	3.078857143	83.92035788
Annotation Cluster 5	Enrichment Score: 0.7232769998749886				
GOTERM_BP_4	muscle cell development	3	0.128093208	4.698837209	82.91415526
GOTERM_BP_4	muscle cell differentiation	3	0.146571407	4.321925134	87.03660045
GOTERM_BP_4	cytoskeleton organization	4	0.166308633	2.763076923	90.41199293

Appendix Table 3.7 cont.

GOTERM_BP_4	actin cytoskeleton organization	3	0.214053101	3.381589958	95.51689772
GOTERM_BP_4	single-organism organelle organization	4	0.361912282	1.823350254	99.69462463
Annotation Cluster 6	Enrichment Score: 0.5128329065387658				
GOTERM_BP_4	larval development	11	0.300102774	1.30202109	98.99438239
GOTERM_BP_4	post-embryonic development	11	0.30980195	1.29124183	99.1599403
GOTERM_BP_DIRECT	nematode larval development	10	0.311277066	1.325890617	98.69998113
Annotation Cluster 7	Enrichment Score: 0.3703467040597769				
GOTERM_BP_4	response to nutrient levels	3	0.026406122	11.54571429	29.17511654
GOTERM_BP_DIRECT	regulation of transcription from RNA polymerase II promoter	3	0.085464618	6.005681818	64.66858567
GOTERM_BP_4	positive regulation of cellular metabolic process	4	0.122505465	3.18816568	81.44764665
GOTERM_BP_4	positive regulation of macromolecule metabolic process	4	0.142140363	2.97679558	86.14170016
GOTERM_BP_4	positive regulation of RNA metabolic process	3	0.182387489	3.759069767	92.54059138
GOTERM_BP_4	positive regulation of nucleobase-containing compound metabolic process	3	0.199455084	3.544736842	94.31664218
GOTERM_BP_4	positive regulation of macromolecule biosynthetic process	3	0.210059699	3.424576271	95.21423653
GOTERM_BP_4	positive regulation of biosynthetic process	3	0.216719908	3.353526971	95.7090747
GOTERM_BP_4	positive regulation of cellular biosynthetic process	3	0.216719908	3.353526971	95.7090747
GOTERM_BP_4	positive regulation of nitrogen compound metabolic process	3	0.220726462	3.312295082	95.98355319
GOTERM_BP_4	positive regulation of gene expression	3	0.251634927	3.026966292	97.6161699
GOTERM_BP_DIRECT	regulation of transcription, DNA-templated	3	0.780276816	1.045844327	99.99999783
GOTERM_BP_4	cellular nitrogen compound biosynthetic process	5	0.861665033	0.826380368	100
GOTERM_BP_4	regulation of gene expression	4	0.873685807	0.814512472	100
GOTERM_BP_4	regulation of RNA metabolic process	3	0.894387086	0.794690265	100
GOTERM_BP_4	regulation of nucleobase-containing compound metabolic process	3	0.908941203	0.759586466	100
GOTERM_BP_4	regulation of macromolecule biosynthetic process	3	0.923404105	0.722898032	100
GOTERM_BP_4	regulation of cellular biosynthetic process	3	0.927503877	0.712070485	100
GOTERM_BP_4	nucleobase-containing compound biosynthetic process	3	0.947818696	0.654412955	100

Appendix Table 3.7 cont.

GOTERM_BP_4	aromatic compound biosynthetic process	3	0.953115905	0.637884767	100
GOTERM_BP_4	heterocycle biosynthetic process	3	0.954051952	0.63487824	100
GOTERM_BP_4	organic cyclic compound biosynthetic process	3	0.957768813	0.622650231	100
GOTERM_BP_4	cellular macromolecule biosynthetic process	4	0.95914892	0.630175439	100
GOTERM_BP_4	macromolecule biosynthetic process	4	0.959903899	0.627972028	100
GOTERM_BP_4	gene expression	4	0.982210378	0.547560976	100
GOTERM_BP_4	RNA metabolic process	3	0.98238717	0.521755972	100
Annotation Cluster 8		Enrichment Score: 0.33226617294925337			
GOTERM_BP_DIRECT	gonad development	4	0.134538597	3.063768116	81.41258998
GOTERM_BP_4	development of primary sexual characteristics	5	0.472624833	1.385802469	99.97381714
GOTERM_BP_4	gonad development	5	0.472624833	1.385802469	99.97381714
GOTERM_BP_4	reproductive structure development	5	0.481628002	1.370295015	99.97902892
GOTERM_BP_4	reproductive system development	5	0.481628002	1.370295015	99.97902892
GOTERM_BP_4	sex differentiation	5	0.538352473	1.277988615	99.99529202
GOTERM_BP_4	animal organ development	5	0.685527106	1.067353407	99.99996661
GOTERM_BP_4	system development	5	0.854035347	0.837686567	100
Annotation Cluster 9		Enrichment Score: 0.25366442457538674			
GOTERM_MF_4	ribonucleotide binding	7	0.402328418	1.333794296	99.35419019
GOTERM_MF_4	purine nucleoside binding	6	0.573347459	1.166439909	99.97623135
GOTERM_MF_4	purine ribonucleoside triphosphate binding	6	0.573347459	1.166439909	99.97623135
GOTERM_MF_4	ribonucleoside binding	6	0.57586063	1.163362231	99.97756794
GOTERM_MF_4	purine ribonucleotide binding	6	0.579201049	1.159283836	99.97923995
GOTERM_MF_4	purine nucleotide binding	6	0.580034266	1.158268702	99.97963917
GOTERM_MF_DIRECT	ATP binding	5	0.599155981	1.187844612	99.99452483
GOTERM_MF_4	nucleotide binding	7	0.609716338	1.092144374	99.99007028
Annotation Cluster 10		Enrichment Score: 0.17732553998365105			
GOTERM_BP_4	ion transmembrane transport	3	0.557053168	1.584705882	99.99723727
GOTERM_MF_4	ion transmembrane transporter activity	3	0.702677496	1.215756303	99.99930905
GOTERM_BP_4	ion transport	3	0.750534544	1.110164835	99.99999831

Appendix Table 3.8 DAVID functional annotation results for all genes upregulated under only desiccation stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 9.63078208734866			
GOTERM_CC_4	ribosomal subunit	30	8.57E-13	4.991413227	1.14E-09
GOTERM_CC_4	ribosome	33	1.54E-11	4.087412831	2.04E-08
GOTERM_CC_DIRECT	ribosome	30	9.98E-11	4.160752177	1.29E-07
KEGG_PATHWAY	Ribosome	31	1.02E-10	3.792553191	1.13E-07
GOTERM_MF_DIRECT	structural constituent of ribosome	33	2.05E-09	3.367100462	2.96E-06
GOTERM_BP_DIRECT	translation	35	5.96E-07	2.587062937	9.25E-04
Annotation Cluster 2		Enrichment Score: 9.398673357148654			
GOTERM_BP_4	larval development	180	3.08E-10	1.527297466	4.82E-07
GOTERM_BP_4	post-embryonic development	181	3.39E-10	1.523067913	5.30E-07
GOTERM_BP_DIRECT	nematode larval development	163	6.10E-10	1.571783059	9.47E-07
Annotation Cluster 3		Enrichment Score: 6.088357035785617			
INTERPRO	TCP-1-like chaperonin intermediate domain	7	3.01E-07	18.27777778	4.84E-04
INTERPRO	Chaperonin TCP-1, conserved site	7	3.01E-07	18.27777778	4.84E-04
INTERPRO	Chaperone tailless complex polypeptide 1 (TCP-1)	7	8.67E-07	16.24691358	0.00139304
INTERPRO	GroEL-like equatorial domain	7	8.67E-07	16.24691358	0.00139304
INTERPRO	GroEL-like apical domain	7	2.08E-06	14.62222222	0.003341404
INTERPRO	Chaperonin Cpn60/TCP-1	7	2.08E-06	14.62222222	0.003341404
Annotation Cluster 4		Enrichment Score: 3.2505604378925392			
GOTERM_BP_4	sex differentiation	83	9.03E-05	1.520760646	0.141267828
GOTERM_BP_4	reproductive structure development	76	3.39E-04	1.493081307	0.529741667
GOTERM_BP_4	reproductive system development	76	3.39E-04	1.493081307	0.529741667
GOTERM_BP_4	gonad development	73	0.001057142	1.45037391	1.641820733

Appendix Table 3.8 cont.

GOTERM_BP_4	development of primary sexual characteristics	73	0.001057142	1.45037391	1.641820733
GOTERM_BP_4	animal organ development	88	0.002700289	1.346625087	4.143737274
Annotation Cluster 5		Enrichment Score: 2.708270690801666			
GOTERM_BP_4	male sex differentiation	13	0.001819969	2.852883675	2.810835615
GOTERM_BP_4	nematode male tail tip morphogenesis	12	0.001909144	3.009635526	2.946642599
GOTERM_BP_DIRECT	nematode male tail tip morphogenesis	12	0.00199228	2.995041322	3.046632775
GOTERM_BP_4	male anatomical structure morphogenesis	12	0.002121604	2.971050455	3.269485664
Annotation Cluster 6		Enrichment Score: 2.619257977479449			
INTERPRO	NADP-dependent oxidoreductase domain	6	4.95E-04	8.355555556	0.792110005
INTERPRO	Aldo/keto reductase subgroup	5	0.001883243	8.703703704	2.983255582
INTERPRO	Aldo/keto reductase	5	0.003528063	7.46031746	5.520341502
INTERPRO	Aldo/keto reductase, conserved site	4	0.010140434	8.355555556	15.1052632
Annotation Cluster 7		Enrichment Score: 2.1519363042180535			
GOTERM_MF_4	purine nucleoside binding	84	0.004941921	1.319989736	5.986616733
GOTERM_MF_4	purine ribonucleoside triphosphate binding	84	0.004941921	1.319989736	5.986616733
GOTERM_MF_4	ribonucleoside binding	84	0.005308512	1.316506913	6.41729462
GOTERM_MF_DIRECT	ATP binding	71	0.005361451	1.36499138	7.447398652
GOTERM_MF_4	ribonucleotide binding	85	0.00568194	1.309151445	6.854138711
GOTERM_MF_4	purine ribonucleotide binding	84	0.005763164	1.311891639	6.948907434
GOTERM_MF_4	purine nucleotide binding	84	0.005906382	1.310742872	7.115792852
GOTERM_MF_4	nucleotide binding	94	0.045288059	1.185468489	43.87025085
Annotation Cluster 8		Enrichment Score: 2.0926572440161184			
GOTERM_BP_4	multicellular organism growth	34	0.00236457	1.727900396	3.63745397
GOTERM_BP_4	regulation of multicellular organism growth	34	0.003032864	1.701041841	4.642824404
GOTERM_BP_4	positive regulation of developmental growth	30	0.013906264	1.587273531	19.68225614
GOTERM_BP_4	positive regulation of multicellular organism growth	29	0.017969581	1.568747929	24.70874539
GOTERM_BP_DIRECT	positive regulation of multicellular organism growth	29	0.01920261	1.56114082	25.97865188
Annotation Cluster 9		Enrichment Score: 1.937576083675099			
INTERPRO	Alpha crystallin/Heat shock protein	5	0.007483468	6.14379085	11.36904915

Appendix Table 3.8 cont.

INTERPRO	Alpha crystallin/Hsp20 domain	5	0.009263651	5.802469136	13.88898033
INTERPRO	HSP20-like chaperone	5	0.022201723	4.541062802	30.28360031
Annotation Cluster 10	Enrichment Score: 1.8085552475701687				
GOTERM_BP_4	genitalia development	55	0.007921086	1.414314964	11.70347829
GOTERM_BP_4	hermaphrodite genitalia development	53	0.015180693	1.375708174	21.29161977
GOTERM_BP_DIRECT	hermaphrodite genitalia development	51	0.031207425	1.328085735	38.85264974

Appendix Table 3.9 DAVID functional annotation results for all genes upregulated under only freezing stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 12.422529416478609			
GOTERM_CC_DIRECT	proteasome complex	16	1.61E-13	13.26684164	2.05E-10
KEGG_PATHWAY	Proteasome	15	4.06E-13	13.77495463	3.93E-10
GOTERM_CC_4	proteasome complex	16	8.24E-13	12.11022727	1.08E-09
Annotation Cluster 2		Enrichment Score: 6.33742164702744			
INTERPRO	Proteasome, subunit alpha/beta	8	6.32E-08	18.8428246	9.52E-05
GOTERM_CC_4	proteasome core complex	8	1.11E-07	17.30032468	1.45E-04
GOTERM_CC_DIRECT	proteasome core complex	8	1.23E-07	17.05736783	1.56E-04
INTERPRO	Proteasome B-type subunit	6	1.45E-07	32.97494305	2.19E-04
GOTERM_MF_DIRECT	threonine-type endopeptidase activity	8	1.58E-07	16.41385281	2.14E-04
GOTERM_BP_DIRECT	proteolysis involved in cellular protein catabolic process	9	4.73E-04	4.849972191	0.708027157
Annotation Cluster 3		Enrichment Score: 4.2658739423271355			
GOTERM_BP_4	cuticle development	16	3.03E-05	3.649015873	0.046523858
GOTERM_BP_DIRECT	collagen and cuticulin-based cuticle development	14	6.17E-05	3.872347512	0.092571804
GOTERM_BP_4	collagen and cuticulin-based cuticle development	14	8.54E-05	3.748173913	0.131281447
Annotation Cluster 4		Enrichment Score: 4.26587321351089			
GOTERM_BP_4	cell morphogenesis involved in differentiation	19	3.30E-05	3.145069124	0.050794213
GOTERM_BP_4	cell projection morphogenesis	22	4.08E-05	2.787442681	0.062681904
GOTERM_BP_4	cell part morphogenesis	22	5.20E-05	2.742301909	0.079988963
GOTERM_BP_4	cell morphogenesis	27	1.23E-04	2.290059032	0.189607228
Annotation Cluster 5		Enrichment Score: 4.048610800738773			
GOTERM_BP_4	cell morphogenesis involved in differentiation	19	3.30E-05	3.145069124	0.050794213
GOTERM_BP_4	neuron development	25	7.24E-05	2.474965549	0.11126062
GOTERM_BP_4	neuron projection development	23	1.24E-04	2.51112462	0.190362166

Appendix Table 3.9 cont.

GOTERM_BP_4	nervous system development	29	2.16E-04	2.136049214	0.331545266
Annotation Cluster 6		Enrichment Score: 3.409979076284017			
GOTERM_MF_DIRECT	nucleotide binding	65	1.14E-06	1.843115259	0.001539722
GOTERM_MF_4	ribonucleotide binding	64	4.37E-04	1.518911653	0.518973727
GOTERM_MF_4	purine ribonucleotide binding	63	5.19E-04	1.516145231	0.616705239
GOTERM_MF_4	purine nucleotide binding	63	5.31E-04	1.514817608	0.63131222
GOTERM_MF_4	purine nucleoside binding	62	7.59E-04	1.5012898	0.900341447
GOTERM_MF_4	purine ribonucleoside triphosphate binding	62	7.59E-04	1.5012898	0.900341447
GOTERM_MF_4	ribonucleoside binding	62	8.12E-04	1.497328613	0.96371421
GOTERM_MF_4	nucleotide binding	73	0.001088301	1.418621695	1.288920509
GOTERM_MF_DIRECT	ATP binding	50	0.002917431	1.511802233	3.864171595
Annotation Cluster 7		Enrichment Score: 3.4002679343672444			
GOTERM_CC_DIRECT	cell cortex	12	8.34E-05	4.368350298	0.105716396
GOTERM_CC_4	cell cortex	12	7.16E-04	3.427422813	0.930780379
GOTERM_CC_4	cytoplasmic region	12	0.001054005	3.273034398	1.366503677
Annotation Cluster 8		Enrichment Score: 3.366586528438728			
GOTERM_BP_4	sex differentiation	57	1.29E-04	1.665036595	0.198623915
GOTERM_BP_4	animal organ development	65	1.56E-04	1.585782205	0.24001565
GOTERM_BP_4	development of primary sexual characteristics	51	6.65E-04	1.615449735	1.018320627
GOTERM_BP_4	gonad development	51	6.65E-04	1.615449735	1.018320627
GOTERM_BP_4	reproductive system development	51	8.41E-04	1.597372475	1.285755274
GOTERM_BP_4	reproductive structure development	51	8.41E-04	1.597372475	1.285755274
Annotation Cluster 9		Enrichment Score: 3.3069722173493097			
GOTERM_BP_4	neuron projection extension	10	3.23E-04	4.527731092	0.495148015
GOTERM_BP_4	developmental growth involved in morphogenesis	10	5.52E-04	4.217612524	0.845046575
GOTERM_BP_4	developmental cell growth	10	6.74E-04	4.105142857	1.032092162
Annotation Cluster 10		Enrichment Score: 2.874850610061086			
GOTERM_BP_DIRECT	phosphorylation	28	1.83E-05	2.514800396	0.027470367
GOTERM_MF_DIRECT	kinase activity	27	9.59E-05	2.322019597	0.129206152

Appendix Table 3.9 cont.

INTERPRO	Protein kinase, catalytic domain	26	8.43E-04	2.070890143	1.262475832
INTERPRO	Protein kinase-like domain	28	0.001570937	1.919539305	2.340825097
GOTERM_BP_DIRECT	protein phosphorylation	26	0.003080567	1.876766247	4.526402571
GOTERM_MF_DIRECT	protein kinase activity	26	0.004083772	1.830466429	5.369949656
GOTERM_MF_4	kinase activity	33	0.013994962	1.533247828	15.45723268
GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	29	0.024521022	1.519681617	25.6051212

Appendix Table 3.10 DAVID functional annotation results for all genes upregulated under both heat and desiccation stresses.

Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 3.434235118111703				
GOTERM_CC_DIRECT	intracellular ribonucleoprotein complex	9	1.76E-07	14.02150685	1.69E-04
GOTERM_CC_4	ribosomal subunit	8	1.10E-06	14.13864013	0.001185393
GOTERM_CC_4	cytosolic part	8	3.19E-06	12.06737438	0.00342889
KEGG_PATHWAY	Ribosome	8	3.24E-06	10.13266583	0.002198495
GOTERM_CC_DIRECT	ribosome	8	4.60E-06	11.51696203	0.004432501
GOTERM_CC_4	ribosome	8	7.89E-06	10.5254321	0.008468464
GOTERM_MF_DIRECT	structural constituent of ribosome	8	1.62E-05	9.414276847	0.016210439
GOTERM_CC_4	cytosol	10	8.27E-05	5.193469786	0.088745875
GOTERM_CC_4	intracellular ribonucleoprotein complex	9	8.38E-04	4.297177419	0.896041798
GOTERM_BP_4	peptide metabolic process	8	0.003321734	3.947252747	4.184445065
GOTERM_BP_4	cellular amide metabolic process	8	0.005507888	3.599198397	6.849627434
GOTERM_BP_4	organonitrogen compound biosynthetic process	9	0.006694834	3.094180704	8.267855162
GOTERM_BP_4	cellular macromolecule biosynthetic process	11	0.202452235	1.444152047	94.53142385
GOTERM_BP_4	macromolecule biosynthetic process	11	0.205543872	1.439102564	94.79759371
GOTERM_BP_4	cellular nitrogen compound biosynthetic process	9	0.418617211	1.239570552	99.90579809
GOTERM_BP_4	gene expression	9	0.646228624	1.026676829	99.99984061
Annotation Cluster 2	Enrichment Score: 2.8541081461178837				
GOTERM_CC_DIRECT	cytosolic large ribosomal subunit	5	8.21E-05	21.06111111	0.079035971
GOTERM_CC_4	large ribosomal subunit	5	2.61E-04	15.58040936	0.280013619
GOTERM_MF_4	RNA binding	5	0.127764871	2.473076923	71.779496
Annotation Cluster 3	Enrichment Score: 2.8098048717170943				
GOTERM_BP_DIRECT	nematode larval development	19	8.00E-04	2.216889112	0.912167497
GOTERM_BP_4	larval development	20	0.002049356	1.972759227	2.601070568

Appendix Table 3.10 cont.

GOTERM_BP_4	post-embryonic development	20	0.002268148	1.956427015	2.875046107
Annotation Cluster 4	Enrichment Score: 2.300608902796648				
GOTERM_CC_4	cytoplasm	26	2.55E-04	1.893412601	0.272861688
GOTERM_CC_4	intracellular part	30	0.015250801	1.37235207	15.2073158
GOTERM_CC_4	intracellular organelle	25	0.032296516	1.399989491	29.69923675
Annotation Cluster 5	Enrichment Score: 1.7144924713508911				
GOTERM_CC_DIRECT	cytosolic small ribosomal subunit	3	0.010435461	18.955	9.60998526
GOTERM_CC_4	small ribosomal subunit	3	0.023792202	12.24942529	22.7767053
GOTERM_BP_4	ribosomal small subunit biogenesis	3	0.028945185	11.04098361	31.43204902
Annotation Cluster 6	Enrichment Score: 1.1172279925076105				
GOTERM_BP_4	gonad development	9	0.057148613	2.078703704	53.04585956
GOTERM_BP_4	development of primary sexual characteristics	9	0.057148613	2.078703704	53.04585956
GOTERM_BP_4	reproductive structure development	9	0.060322207	2.055442523	55.03626594
GOTERM_BP_4	reproductive system development	9	0.060322207	2.055442523	55.03626594
GOTERM_BP_4	sex differentiation	9	0.083533779	1.916982922	67.39287763
GOTERM_BP_DIRECT	hermaphrodite genitalia development	7	0.088231074	2.205663957	65.25389426
GOTERM_BP_4	animal organ development	10	0.092183537	1.778922345	71.13287527
GOTERM_BP_4	hermaphrodite genitalia development	7	0.102705286	2.112231183	75.14806644
GOTERM_BP_4	genitalia development	7	0.106231269	2.092543276	76.37387363
Annotation Cluster 7	Enrichment Score: 0.7534990594917631				
GOTERM_BP_DIRECT	protein transport	3	0.117507716	4.983	76.08364777
GOTERM_BP_4	protein transport	4	0.168205838	2.763076923	90.61468898
GOTERM_BP_4	establishment of protein localization	4	0.185563246	2.633431085	92.84202919
GOTERM_BP_4	organic substance transport	5	0.264000258	1.873956594	98.05111892
Annotation Cluster 8	Enrichment Score: 0.5374508421132793				
GOTERM_MF_4	purine ribonucleoside triphosphate binding	7	0.246295454	1.587654321	92.69704456
GOTERM_MF_4	purine nucleoside binding	7	0.246295454	1.587654321	92.69704456
GOTERM_MF_4	ribonucleoside binding	7	0.248263862	1.583465259	92.87167147
GOTERM_MF_4	purine ribonucleotide binding	7	0.250896721	1.57791411	93.09941979

Appendix Table 3.10 cont.

GOTERM_MF_4	purine nucleotide binding	7	0.251556405	1.576532399	93.15545709
GOTERM_MF_4	ribonucleotide binding	7	0.261520265	1.556093345	93.95389228
GOTERM_MF_4	nucleotide binding	7	0.443607121	1.274168436	99.55994109
GOTERM_MF_DIRECT	ATP binding	6	0.454917603	1.330385965	99.77148592
Annotation Cluster 9	Enrichment Score: 0.41234106441687957				
GOTERM_BP_4	organelle assembly	4	0.193307844	2.58045977	93.66887107
GOTERM_BP_4	cellular component assembly involved in morphogenesis	3	0.250765622	3.047511312	97.54968967
GOTERM_BP_4	anatomical structure formation involved in morphogenesis	3	0.358969131	2.338541667	99.66964156
GOTERM_BP_4	cellular component morphogenesis	3	0.675920647	1.280418251	99.99994831
GOTERM_BP_4	single-organism organelle organization	3	0.737605908	1.139593909	99.99999657
Annotation Cluster 10	Enrichment Score: 0.273794525023818				
GOTERM_BP_4	macromolecular complex assembly	4	0.448118153	1.595026643	99.95174831
GOTERM_BP_4	protein complex assembly	3	0.563616205	1.56993007	99.99763727
GOTERM_BP_4	protein complex subunit organization	3	0.597365885	1.476973684	99.99915994

Appendix Table 3.11 DAVID functional annotation results for all genes upregulated under both heat and freezing stresses.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 1.5331320518768734				
GOTERM_MF_4	purine ribonucleoside triphosphate binding	6	0.021933125	3.14041514	16.46703352
GOTERM_MF_4	purine nucleoside binding	6	0.021933125	3.14041514	16.46703352
GOTERM_MF_4	ribonucleoside binding	6	0.022169558	3.132129085	16.6307213
GOTERM_MF_4	purine ribonucleotide binding	6	0.022487526	3.12114879	16.85041478
GOTERM_MF_4	purine nucleotide binding	6	0.022567506	3.118415735	16.9055954
GOTERM_MF_4	ribonucleotide binding	6	0.023790808	3.077986836	17.74559628
GOTERM_MF_DIRECT	nucleotide binding	6	0.028582948	3.119118131	23.76206789
GOTERM_MF_4	nucleotide binding	6	0.052100437	2.52033317	35.2160312
GOTERM_MF_DIRECT	ATP binding	5	0.08299858	2.771637427	55.54268246
Annotation Cluster 2	Enrichment Score: 1.3154297490252276				
GOTERM_BP_DIRECT	phosphorylation	4	0.011141063	7.861392833	11.25475234
GOTERM_MF_DIRECT	kinase activity	4	0.02042415	6.306719894	17.5568775
GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	4	0.043218945	4.530827951	30.12399199
GOTERM_MF_4	kinase activity	4	0.058438551	4.017180789	38.64804282
GOTERM_BP_4	phosphate-containing compound metabolic process	5	0.070317822	2.89055794	59.4979704
GOTERM_MF_DIRECT	ATP binding	5	0.08299858	2.771637427	55.54268246
GOTERM_MF_DIRECT	transferase activity	4	0.184676429	2.519670388	85.19440895
Annotation Cluster 3	Enrichment Score: 1.0051395305078314				
GOTERM_BP_4	larval development	7	0.076496243	2.071397188	62.71104846
GOTERM_BP_4	post-embryonic development	7	0.079154128	2.054248366	64.01977082
GOTERM_BP_DIRECT	embryo development ending in birth or egg hatching	8	0.104904756	1.770148629	69.30590126
GOTERM_BP_4	embryo development	8	0.119396848	1.702369668	79.32241899
GOTERM_BP_DIRECT	nematode larval development	6	0.124279826	2.059030135	75.69090598

Appendix Table 3.11 cont.

Annotation Cluster 4		Enrichment Score: 0.9632363691420783			
GOTERM_BP_4	protein modification process	5	0.095269058	2.608443067	71.09199534
GOTERM_BP_4	cellular protein metabolic process	6	0.110270542	2.117924528	76.50337939
GOTERM_BP_4	macromolecule modification	5	0.122709875	2.386605245	80.26633907
Annotation Cluster 5		Enrichment Score: 0.8455631782557304			
GOTERM_BP_DIRECT	protein phosphorylation	3	0.117830397	4.73862247	73.71361618
INTERPRO	Protein kinase, catalytic domain	3	0.129643096	4.560806553	75.37048044
GOTERM_MF_DIRECT	protein kinase activity	3	0.164624921	3.872140523	81.41583325
INTERPRO	Protein kinase-like domain	3	0.164909234	3.925517491	83.77535343
Annotation Cluster 6		Enrichment Score: 0.8380449987311416			
GOTERM_BP_4	animal organ development	5	0.089167738	2.668383518	68.58034905
GOTERM_BP_4	development of primary sexual characteristics	4	0.147496663	2.771604938	86.16708742
GOTERM_BP_4	gonad development	4	0.147496663	2.771604938	86.16708742
GOTERM_BP_4	reproductive structure development	4	0.151160692	2.740590031	86.8862848
GOTERM_BP_4	reproductive system development	4	0.151160692	2.740590031	86.8862848
GOTERM_BP_4	system development	5	0.174817884	2.094216418	90.76243844
GOTERM_BP_4	sex differentiation	4	0.175569579	2.55597723	90.86620921
Annotation Cluster 7		Enrichment Score: 0.7224945526445601			
GOTERM_CC_4	cytoplasmic part	7	0.05198233	2.243241618	42.16080619
GOTERM_CC_4	cytoplasm	7	0.232534046	1.529294793	93.37583499
GOTERM_CC_4	intracellular part	8	0.562566734	1.097881656	99.97924697
Annotation Cluster 8		Enrichment Score: 0.5585969649511825			
GOTERM_BP_DIRECT	hermaphrodite genitalia development	3	0.270016307	2.780248685	96.50631521
GOTERM_BP_4	hermaphrodite genitalia development	3	0.277664597	2.715725806	98.22606596
GOTERM_BP_4	genitalia development	3	0.281383484	2.690412783	98.33601489
Annotation Cluster 9		Enrichment Score: 0.0254760241500447			
GOTERM_CC_4	integral component of membrane	7	0.910770997	0.816896627	100
GOTERM_CC_DIRECT	integral component of membrane	7	0.956970841	0.733228338	100
GOTERM_CC_DIRECT	membrane	7	0.962197071	0.723394395	100

Appendix Table 3.12 DAVID functional annotation results for all genes upregulated under both freezing and desiccation stresses. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 2.936648661289104				
GOTERM_MF_DIRECT	cholesterol binding	4	9.33E-04	18.42371234	1.238158613
GOTERM_MF_4	cholesterol binding	4	0.001030959	17.77880184	1.205111176
GOTERM_MF_4	sterol binding	4	0.001610637	15.55645161	1.876859336
Annotation Cluster 2	Enrichment Score: 1.9685815073519815				
GOTERM_BP_4	defense response to bacterium	10	0.005045452	3.085203848	7.360510019
GOTERM_BP_4	defense response to other organism	10	0.014732343	2.6003861	20.09524876
GOTERM_BP_4	response to other organism	10	0.016714122	2.545832546	22.490356
Annotation Cluster 3	Enrichment Score: 1.822582448208148				
INTERPRO	Nidogen, extracellular domain	3	0.009759603	18.84895833	13.50467966
GOTERM_BP_DIRECT	cell-matrix adhesion	3	0.01861116	13.71108491	23.65005004
GOTERM_BP_4	cell-matrix adhesion	3	0.018751977	13.65202703	24.88319896
Annotation Cluster 4	Enrichment Score: 1.7299249229276552				
INTERPRO	EGF-like calcium-binding, conserved site	5	0.005974629	6.731770833	8.483041928
INTERPRO	EGF-like calcium-binding	5	0.022609261	4.597306911	28.70119073
INTERPRO	EGF-type aspartate/asparagine hydroxylation site	4	0.04782202	4.864247312	51.56224525
Annotation Cluster 5	Enrichment Score: 1.724608659428519				
INTERPRO	Protein phosphatase 2C (PP2C)-like Protein phosphatase 2C, manganese/magnesium aspartate binding site	4	0.003383962	12.56597222	4.890638969
INTERPRO	Protein phosphatase 2C	3	0.017590342	14.13671875	23.08922574
INTERPRO	Protein phosphatase 2C	3	0.022224514	12.56597222	28.28488175
GOTERM_MF_DIRECT	protein serine/threonine phosphatase activity	3	0.095506571	5.68967587	73.81582593
Annotation Cluster 6	Enrichment Score: 1.6737176200956212				
GOTERM_BP_4	positive regulation of macromolecule biosynthetic process	14	0.013032266	2.159642694	17.98570097
GOTERM_BP_4	positive regulation of RNA metabolic process	13	0.015205576	2.201257071	20.67338104

Appendix Table 3.12 cont.

GOTERM_BP_4	positive regulation of biosynthetic process	14	0.015371554	2.114836829	20.87522494
GOTERM_BP_4	positive regulation of cellular biosynthetic process	14	0.015371554	2.114836829	20.87522494
GOTERM_BP_4	positive regulation of nitrogen compound metabolic process	14	0.01688571	2.088834736	22.69454693
GOTERM_BP_4	positive regulation of nucleobase-containing compound metabolic process	13	0.022963627	2.075746799	29.61154258
GOTERM_BP_4	positive regulation of cellular metabolic process	17	0.023664229	1.8310411	30.37059409
GOTERM_BP_4	positive regulation of gene expression	14	0.032219963	1.908897662	39.04417068
GOTERM_BP_DIRECT	positive regulation of transcription from RNA polymerase II promoter	10	0.033346874	2.256968709	38.5625555
GOTERM_BP_4	positive regulation of macromolecule metabolic process	17	0.039676474	1.70964611	45.76959763
Annotation Cluster 7	Enrichment Score: 1.5199283119377243				
GOTERM_CC_4	endoplasmic reticulum part	13	0.01687967	2.170128446	19.28434884
GOTERM_CC_4	endoplasmic reticulum membrane	11	0.03816616	2.087933031	38.71976972
GOTERM_CC_4	nuclear outer membrane-endoplasmic reticulum membrane network	11	0.042773289	2.046382124	42.3129451
Annotation Cluster 8	Enrichment Score: 1.4309210283787772				
GOTERM_BP_4	sphingolipid metabolic process	5	0.02517533	4.439683586	31.9818172
KEGG_PATHWAY	Sphingolipid metabolism	4	0.039356385	5.169859515	33.31511414
GOTERM_BP_DIRECT	sphingolipid metabolic process	4	0.051433554	4.717792656	53.16077957
Annotation Cluster 9	Enrichment Score: 1.4162315928511962				
INTERPRO	Protein kinase, catalytic domain	20	0.014356925	1.821155395	19.25844387
GOTERM_BP_DIRECT	protein phosphorylation	20	0.028437423	1.688817232	33.9253549
INTERPRO	Protein kinase-like domain	21	0.031385198	1.64585499	37.60675468
GOTERM_MF_DIRECT	protein kinase activity	20	0.048923951	1.58046552	48.8098029
GOTERM_MF_DIRECT	protein serine/threonine kinase activity	15	0.067567177	1.661932814	60.6981625
GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	24	0.075106343	1.425018468	60.05729875
Annotation Cluster 10	Enrichment Score: 1.3067951045441				
GOTERM_BP_DIRECT	positive regulation of necrotic cell death	3	0.047407757	8.437590711	50.22303117
GOTERM_BP_4	positive regulation of necrotic cell death	3	0.047750944	8.401247401	52.26728649
GOTERM_BP_4	necrotic cell death	4	0.047841424	4.854054054	52.33579288

Appendix Table 3.12 cont.

GOTERM_BP_4	regulation of necrotic cell death	3	0.054724791	7.801158301	57.28651517
-------------	-----------------------------------	---	-------------	-------------	-------------

Appendix Table 3.13 DAVID functional annotation results for all genes upregulated under heat, freezing, and desiccation stresses. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 2.8552971885509995			
GOTERM_BP_DIRECT	translation	10	2.47E-04	4.707287449	0.313537408
GOTERM_BP_4	peptide metabolic process	12	7.72E-04	3.343555268	1.070536998
GOTERM_BP_4	cellular amide metabolic process	12	0.001631685	3.04873276	2.249558093
GOTERM_BP_4	organonitrogen compound biosynthetic process	12	0.012190328	2.32973606	15.70847473
Annotation Cluster 2		Enrichment Score: 1.9871768061645052			
GOTERM_MF_DIRECT	neuropeptide receptor binding	4	3.03E-04	29.05440613	0.348577123
GOTERM_MF_DIRECT	neuropeptide hormone activity	3	0.005061985	27.23850575	5.675575344
INTERPRO	FMRFamide-related peptide-like	3	0.005171073	27.15947467	6.52123622
GOTERM_BP_DIRECT	neuropeptide signaling pathway	5	0.006510934	6.651601831	7.973193381
GOTERM_MF_4	G-protein coupled receptor binding	4	0.01050715	8.633286713	10.11093752
GOTERM_MF_4	hormone activity	3	0.059868843	7.419230769	46.36711839
GOTERM_CC_DIRECT	extracellular space	4	0.378228662	1.799240626	99.31743275
Annotation Cluster 3		Enrichment Score: 1.7988997757389005			
GOTERM_BP_4	larval development	28	0.011936155	1.559640236	15.40578245
GOTERM_BP_4	post-embryonic development	28	0.013301544	1.546728181	17.02004079
GOTERM_BP_DIRECT	nematode larval development	25	0.025265865	1.535241767	27.78432563
Annotation Cluster 4		Enrichment Score: 1.7363771721423678			
GOTERM_CC_DIRECT	striated muscle thin filament	3	0.003766332	31.65027829	3.882850853
GOTERM_CC_4	striated muscle thin filament	3	0.010294495	19.12141148	11.03370385
GOTERM_CC_4	myofilament	3	0.020259104	13.45580808	20.64532167
GOTERM_CC_4	sarcomere	3	0.14432697	4.430570953	82.81311629
Annotation Cluster 5		Enrichment Score: 1.4709411334878155			
GOTERM_BP_DIRECT	endoplasmic reticulum unfolded protein response	4	0.021181266	6.706272531	23.83905149

Appendix Table 3.13 cont.

GOTERM_BP_4	endoplasmic reticulum unfolded protein response	4	0.022106092	6.58579068	26.76191412
GOTERM_BP_4	response to unfolded protein	4	0.031212812	5.762566845	35.7129971
GOTERM_BP_4	cellular response to topologically incorrect protein	4	0.034943125	5.51202046	39.07730316
GOTERM_BP_4	response to topologically incorrect protein	4	0.046331884	4.923358081	48.36470637
GOTERM_BP_4	response to endoplasmic reticulum stress	4	0.063141475	4.334238311	59.69672579
Annotation Cluster 6	Enrichment Score: 1.2657589696625473				
GOTERM_MF_4	ribonucleotide binding	17	0.024206201	1.74419254	21.90795898
GOTERM_MF_4	purine ribonucleoside triphosphate binding	16	0.041439403	1.674888075	34.76001198
GOTERM_MF_4	purine nucleoside binding	16	0.041439403	1.674888075	34.76001198
GOTERM_MF_4	ribonucleoside binding	16	0.042284583	1.670468845	35.33818742
GOTERM_MF_4	purine ribonucleotide binding	16	0.043431031	1.664612688	36.11507709
GOTERM_MF_4	purine nucleotide binding	16	0.043721154	1.663155059	36.31034127
GOTERM_MF_4	nucleotide binding	18	0.065071779	1.512199902	49.28822297
GOTERM_MF_DIRECT	nucleotide binding	14	0.125066259	1.505781167	78.5255856
GOTERM_MF_DIRECT	ATP binding	13	0.149351306	1.490949788	84.47013785
Annotation Cluster 7	Enrichment Score: 1.2412802804493126				
GOTERM_BP_4	regulation of multicellular organism growth	8	0.030648923	2.627491618	35.189684
GOTERM_BP_4	positive regulation of multicellular organism growth	7	0.059674361	2.485813149	57.56812797
GOTERM_BP_4	positive regulation of developmental growth	7	0.065012507	2.431329573	60.80382115
GOTERM_BP_DIRECT	positive regulation of multicellular organism growth	7	0.069015724	2.399793602	59.73342004
GOTERM_BP_4	multicellular organism growth	7	0.075761863	2.335356037	66.63627864
Annotation Cluster 8	Enrichment Score: 1.2335388930433702				
GOTERM_MF_DIRECT	structural constituent of ribosome	6	0.02314647	3.652090156	23.63406551
GOTERM_CC_4	ribosomal subunit	5	0.023858309	4.51874152	23.87738541
KEGG_PATHWAY	Ribosome	5	0.030060052	3.987391647	22.71754196
GOTERM_CC_DIRECT	ribosome	5	0.046020056	3.672500646	39.00931627
GOTERM_CC_4	ribosome	5	0.059516062	3.36395202	50.00607
GOTERM_CC_DIRECT	cytosolic large ribosomal subunit	3	0.077434504	6.447278912	57.0815317
GOTERM_CC_4	large ribosomal subunit	3	0.127686621	4.780352871	78.63497086

Appendix Table 3.13 cont.

GOTERM_CC_DIRECT	intracellular ribonucleoprotein complex	4	0.128633675	3.179480011	76.42755833
GOTERM_CC_4	cytosolic part	4	0.136779144	3.085408222	81.02028994
Annotation Cluster 9		Enrichment Score: 1.1184810221078847			
GOTERM_BP_DIRECT	positive regulation of vulval development	4	0.018966413	6.993684211	21.61762849
GOTERM_BP_4	positive regulation of post-embryonic development	4	0.033050246	5.634509804	37.3911265
GOTERM_BP_4	positive regulation of nematode larval development	4	0.033050246	5.634509804	37.3911265
GOTERM_BP_4	regulation of post-embryonic development	6	0.102191889	2.392008879	77.73019222
GOTERM_BP_4	vulval development	5	0.299656844	1.785584093	99.30055906
GOTERM_BP_4	post-embryonic organ development	5	0.306717564	1.765689005	99.39265257
Annotation Cluster 10		Enrichment Score: 0.8775612560895987			
GOTERM_BP_4	cellular macromolecule biosynthetic process	19	0.108311671	1.408627451	79.75443231
GOTERM_BP_4	macromolecule biosynthetic process	19	0.111056146	1.40370218	80.60555112
GOTERM_BP_4	cellular nitrogen compound biosynthetic process	17	0.193686066	1.322208589	95.0186926

Appendix Table 3.14 DAVID functional annotation results for all genes downregulated under heat stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 2.354603785807994				
GOTERM_MF_DIRECT	ATPase activity	7	0.00369477	4.672746479	4.518948017
GOTERM_MF_4	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	7	0.004493945	4.471932439	4.876322642
GOTERM_MF_4	primary active transmembrane transporter activity	7	0.005199642	4.340405014	5.622094627
Annotation Cluster 2	Enrichment Score: 2.250122005576583				
INTERPRO	ABC transporter, transmembrane domain, type 1	5	0.001082019	10.88421053	1.484703073
INTERPRO	ABC transporter, conserved site	5	0.003019063	8.272	4.091685431
INTERPRO	ABC transporter-like	5	0.004852789	7.256140351	6.500513483
GOTERM_MF_DIRECT	ATPase activity, coupled to transmembrane movement of substances	5	0.008346185	6.180881586	9.940731897
INTERPRO	AAA+ ATPase domain	5	0.042442766	3.794495413	45.07714421
Annotation Cluster 3	Enrichment Score: 2.1242852295091934				
GOTERM_CC_4	vacuolar membrane	6	0.002913794	6.092035061	3.469370442
GOTERM_CC_4	vacuolar part	6	0.003405215	5.877022059	4.043489618
GOTERM_CC_4	vacuole	7	0.042711121	2.72338639	41.03273003
Annotation Cluster 4	Enrichment Score: 1.6022911750149864				
GOTERM_MF_DIRECT	GTPase activator activity	5	0.020183437	4.768108652	22.48670676
GOTERM_MF_4	GTPase activator activity	5	0.023850612	4.517564403	23.50419433
GOTERM_MF_4	GTPase regulator activity	5	0.032406476	4.10687673	30.62529933
Annotation Cluster 5	Enrichment Score: 1.5730639262940194				
GOTERM_MF_DIRECT	nucleotide binding	25	0.014076014	1.647421548	16.22988222
GOTERM_MF_4	purine ribonucleotide binding	28	0.016650575	1.552046666	17.00305622
GOTERM_MF_4	purine nucleotide binding	28	0.016829695	1.550687606	17.17070328
GOTERM_MF_4	ribonucleotide binding	28	0.019708009	1.530583618	19.82274412

Appendix Table 3.14 cont.

GOTERM_MF_4	purine ribonucleoside triphosphate binding	27	0.027114808	1.505854801	26.2959286
GOTERM_MF_4	purine nucleoside binding	27	0.027114808	1.505854801	26.2959286
GOTERM_MF_4	ribonucleoside binding	27	0.027930421	1.501881569	26.97885436
GOTERM_MF_DIRECT	ATP binding	22	0.041498353	1.545871016	41.10899145
GOTERM_MF_4	nucleotide binding	29	0.105020797	1.298040444	70.81530599
Annotation Cluster 6	Enrichment Score: 1.563246278319628				
INTERPRO	Peptidase aspartic, active site	3	0.010270869	19.08923077	13.29394605
INTERPRO	Aspartic peptidase	3	0.03080394	10.78956522	35.09927651
GOTERM_MF_DIRECT	aspartic-type endopeptidase activity	3	0.064572515	7.152162978	56.56455961
Annotation Cluster 7	Enrichment Score: 1.4999362901236082				
GOTERM_CC_4	cytoplasmic, membrane-bounded vesicle	6	0.0238423	3.646327555	25.32304698
GOTERM_CC_4	membrane-bounded vesicle	6	0.024509996	3.619904891	25.93879126
GOTERM_CC_4	intracellular vesicle	6	0.041379165	3.141804245	40.03222908
GOTERM_CC_4	cytoplasmic vesicle	6	0.041379165	3.141804245	40.03222908
Annotation Cluster 8	Enrichment Score: 1.4588348869621588				
GOTERM_BP_4	positive regulation of RNA metabolic process	8	0.02280313	2.823714379	28.96423854
GOTERM_BP_4	positive regulation of gene expression	9	0.023981966	2.557999683	30.2241669
GOTERM_BP_4	positive regulation of nucleobase-containing compound metabolic process	8	0.030131277	2.662713121	36.46542646
GOTERM_BP_4	positive regulation of cellular metabolic process	10	0.033335969	2.245187099	39.50775313
GOTERM_BP_4	positive regulation of macromolecule biosynthetic process	8	0.035354771	2.572451659	41.35393361
GOTERM_BP_4	positive regulation of biosynthetic process	8	0.038907894	2.519081293	44.47621643
GOTERM_BP_4	positive regulation of cellular biosynthetic process	8	0.038907894	2.519081293	44.47621643
GOTERM_BP_4	positive regulation of nitrogen compound metabolic process	8	0.041148671	2.488108982	46.36482525
GOTERM_BP_DIRECT	positive regulation of transcription from RNA polymerase II promoter	6	0.061347498	2.814572743	58.41656607
Annotation Cluster 9	Enrichment Score: 1.370389007769635				
GOTERM_CC_DIRECT	integral component of plasma membrane	15	0.024876851	1.900991754	24.4708582
GOTERM_CC_4	integral component of plasma membrane	15	0.054525065	1.699139031	49.25969078
GOTERM_CC_4	intrinsic component of plasma membrane	15	0.057074404	1.687658361	50.89065677

Appendix Table 3.14 cont.

Annotation Cluster 10	Enrichment Score: 1.1798696349424356				
KEGG_PATHWAY	Inositol phosphate metabolism	3	0.035295024	9.730769231	28.79672372
KEGG_PATHWAY	Phosphatidylinositol signaling system	3	0.05160586	7.90625	39.39574396
GOTERM_BP_4	phospholipid metabolic process	4	0.158481509	2.890945674	92.25510289

Appendix Table 3.15 DAVID functional annotation results for all genes downregulated under desiccation stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 10.734486380892319				
GOTERM_CC_4	intrinsic component of plasma membrane	138	3.29E-13	1.848731615	4.45E-10
GOTERM_CC_4	integral component of plasma membrane	137	4.25E-13	1.847820282	5.76E-10
GOTERM_CC_DIRECT	integral component of plasma membrane	109	4.48E-08	1.678769338	5.90E-05
Annotation Cluster 2	Enrichment Score: 7.306583464703739				
GOTERM_CC_DIRECT	neuronal cell body	38	1.91E-08	2.718485809	2.51E-05
GOTERM_CC_4	somatodendritic compartment	47	6.75E-08	2.318079371	9.14E-05
GOTERM_CC_4	neuronal cell body	41	9.35E-08	2.463351656	1.27E-04
Annotation Cluster 3	Enrichment Score: 6.886393846908003				
GOTERM_CC_4	M band	15	1.06E-07	5.310797342	1.44E-04
GOTERM_CC_DIRECT	M band	15	1.12E-07	5.288783482	1.48E-04
GOTERM_CC_4	A band	16	1.84E-07	4.806539817	2.49E-04
Annotation Cluster 4	Enrichment Score: 6.267636314327617				
INTERPRO	Neurotransmitter-gated ion-channel transmembrane domain	29	2.96E-08	3.199237921	4.92E-05
INTERPRO	Neurotransmitter-gated ion-channel, conserved site	28	5.51E-08	3.195433833	9.17E-05
INTERPRO	Neurotransmitter-gated ion-channel ligand-binding	31	6.53E-08	2.959507228	1.09E-04
INTERPRO	Neurotransmitter-gated ion-channel	29	3.52E-07	2.879314129	5.86E-04
GOTERM_MF_DIRECT	extracellular ligand-gated ion channel activity	31	6.92E-06	2.372351934	0.010495079
GOTERM_MF_4	ligand-gated channel activity	35	9.55E-05	1.978678976	0.125067879
Annotation Cluster 5	Enrichment Score: 3.5401312472941315				
GOTERM_BP_4	cell morphogenesis involved in differentiation	42	3.63E-06	2.128862924	0.005830509
GOTERM_BP_4	neuron projection guidance	32	2.01E-05	2.251420065	0.032248716
GOTERM_BP_4	neuron projection development	53	4.41E-05	1.771895534	0.070945442
GOTERM_BP_4	cell projection morphogenesis	42	0.001716644	1.629500016	2.725681694

Appendix Table 3.15 cont.

GOTERM_BP_4	cell part morphogenesis	42	0.002345861	1.603111352	3.707181256
GOTERM_BP_4	cell morphogenesis	50	0.044472438	1.298598006	51.89157758
Annotation Cluster 6		Enrichment Score: 3.312513511849342			
GOTERM_CC_4	postsynaptic membrane	14	4.39E-04	3.017148635	0.593196959
GOTERM_CC_DIRECT	postsynaptic membrane	14	4.59E-04	3.00464221	0.602840427
GOTERM_CC_4	synaptic membrane	16	5.73E-04	2.688403626	0.772836192
Annotation Cluster 7		Enrichment Score: 3.073320951074031			
INTERPRO	ABC transporter, conserved site	16	7.67E-05	3.177174211	0.127467625
INTERPRO	ABC transporter-like	17	1.08E-04	2.961182105	0.179895286
INTERPRO	ABC transporter, transmembrane domain, type 1	13	2.32E-04	3.396650061	0.384371896
GOTERM_MF_DIRECT	ATPase activity, coupled to transmembrane movement of substances	17	6.84E-04	2.50556644	1.032989062
GOTERM_MF_DIRECT	ATPase activity	22	0.011499234	1.750948783	16.0932221
INTERPRO	AAA+ ATPase domain	19	0.023982439	1.730685494	33.20621183
Annotation Cluster 8		Enrichment Score: 2.991689109518408			
GOTERM_BP_4	muscle cell differentiation	41	1.02E-05	2.067062473	0.016474962
GOTERM_BP_4	myofibril assembly	35	8.97E-05	2.012035081	0.144254875
GOTERM_BP_4	muscle cell development	35	2.36E-04	1.918452054	0.378540871
GOTERM_BP_4	striated muscle myosin thick filament assembly	30	7.63E-04	1.911044903	1.220795195
GOTERM_BP_DIRECT	striated muscle myosin thick filament assembly	29	0.001724657	1.840217213	2.759815431
GOTERM_BP_4	actin cytoskeleton organization	41	0.002260533	1.617325031	3.574626217
GOTERM_BP_4	cellular component assembly involved in morphogenesis	37	0.005710318	1.578413558	8.799620079
GOTERM_BP_4	organelle assembly	41	0.316402541	1.110749087	99.77978858
Annotation Cluster 9		Enrichment Score: 2.980927745172803			
GOTERM_CC_4	plasma membrane receptor complex	21	1.84E-05	2.891434109	0.024910376
GOTERM_CC_4	acetylcholine-gated channel complex	18	3.62E-05	3.07659984	0.049063323
GOTERM_CC_DIRECT	acetylcholine-gated channel complex	18	3.84E-05	3.063846983	0.050615574
GOTERM_BP_DIRECT	chemical synaptic transmission, postsynaptic	14	0.001054987	2.759650759	1.696801882
GOTERM_MF_DIRECT	acetylcholine-activated cation-selective channel activity	15	0.00263305	2.387657431	3.92088877

Appendix Table 3.15 cont.

GOTERM_MF_DIRECT	acetylcholine binding	15	0.003218707	2.340840618	4.773243381
GOTERM_MF_4	acetylcholine receptor activity	15	0.0038656	2.289614243	4.950860891
INTERPRO	Nicotinic acetylcholine-gated receptor, transmembrane domain	13	0.005535126	2.435334006	8.814614292
GOTERM_MF_4	ammonium ion binding	15	0.005677406	2.201552157	7.192507668
GOTERM_BP_DIRECT	cation transmembrane transport	19	0.006112611	1.977823537	9.462710923
GOTERM_MF_DIRECT	acetylcholine receptor activity	13	0.009528204	2.249242507	13.51886457
Annotation Cluster 10	Enrichment Score: 2.89698332113607				
GOTERM_CC_4	myosin complex	12	1.31E-04	3.83747937	0.17693611
INTERPRO	Myosin head, motor domain	8	0.001207575	4.41274196	1.988731334
GOTERM_CC_DIRECT	myosin complex	9	0.001795878	3.702148437	2.34093005
GOTERM_MF_DIRECT	motor activity	9	0.009104881	2.865188917	12.9564236

Appendix Table 3.16 DAVID functional annotation results for all genes downregulated under freezing stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 3.770875785720256			
INTERPRO	FAD-binding, type 2	6	9.34E-05	10.31054131	0.153500427
INTERPRO	CO dehydrogenase flavoprotein-like, FAD-binding, subdomain 2	6	9.34E-05	10.31054131	0.153500427
GOTERM_MF_DIRECT	oxidoreductase activity, acting on CH-OH group of donors	6	5.58E-04	7.348062016	0.829256519
Annotation Cluster 2		Enrichment Score: 3.5258221672213854			
INTERPRO	ABC transporter-like	14	1.77E-04	3.37655154	0.290581783
INTERPRO	ABC transporter, conserved site	13	1.87E-04	3.574320988	0.30636566
INTERPRO	ABC transporter, transmembrane domain, type 1	11	2.84E-04	3.979507172	0.466895155
GOTERM_MF_DIRECT	ATPase activity, coupled to transmembrane movement of substances	14	8.40E-04	2.857579673	1.245171931
Annotation Cluster 3		Enrichment Score: 3.1070583659518456			
GOTERM_CC_4	endoplasmic reticulum part	32	3.49E-04	1.972057735	0.465801001
GOTERM_CC_4	nuclear outer membrane-endoplasmic reticulum membrane network	29	5.96E-04	1.9916802	0.794795611
GOTERM_CC_4	endoplasmic reticulum membrane	28	9.54E-04	1.96204729	1.268202378
GOTERM_CC_DIRECT	endoplasmic reticulum membrane	22	0.001879552	2.079435524	2.386331478
Annotation Cluster 4		Enrichment Score: 3.0027187432190843			
GOTERM_BP_DIRECT	phosphorylation	52	9.38E-06	1.902922033	0.014958592
GOTERM_MF_DIRECT	kinase activity	49	8.27E-04	1.617013647	1.226424348
GOTERM_MF_4	kinase activity	64	0.126518476	1.168634411	82.5525231
Annotation Cluster 5		Enrichment Score: 2.754253982975061			
GOTERM_CC_DIRECT	myosin complex	8	0.00108697	4.66297663	1.386523799
INTERPRO	Myosin head, motor domain	7	0.001260014	5.346206605	2.052393097
GOTERM_CC_4	myosin complex	9	0.001278199	4.007730236	1.696332245

Appendix Table 3.16 cont.

GOTERM_MF_DIRECT	motor activity	8	0.005492785	3.527069767	7.886879129
Annotation Cluster 6		Enrichment Score: 2.733104523819273			
GOTERM_BP_4	muscle cell differentiation	32	2.84E-05	2.248806574	0.045091695
GOTERM_BP_4	myofibril assembly	27	2.58E-04	2.163533611	0.408481081
GOTERM_BP_4	muscle cell development	27	5.50E-04	2.062904141	0.869912327
GOTERM_BP_4	striated muscle myosin thick filament assembly	24	7.62E-04	2.131048121	1.202691163
GOTERM_BP_DIRECT	striated muscle myosin thick filament assembly	23	0.002225462	2.00618914	3.490968142
GOTERM_BP_4	cellular component assembly involved in morphogenesis	28	0.009493143	1.66498179	14.05148191
GOTERM_BP_4	actin cytoskeleton organization	29	0.014315091	1.594570875	20.45852212
GOTERM_BP_4	organelle assembly	33	0.14699933	1.246173255	91.98627368
Annotation Cluster 7		Enrichment Score: 2.7085600088282296			
GOTERM_BP_4	cellular nitrogen compound catabolic process	20	0.001385475	2.22736668	2.17689579
GOTERM_BP_4	heterocycle catabolic process	20	0.001385475	2.22736668	2.17689579
GOTERM_BP_4	organic cyclic compound catabolic process	20	0.002065422	2.154338265	3.228928209
GOTERM_BP_4	aromatic compound catabolic process	19	0.003694485	2.098216848	5.706493223
Annotation Cluster 8		Enrichment Score: 2.582905467944192			
GOTERM_CC_4	A band	9	0.001977775	3.764837494	2.61344515
GOTERM_CC_DIRECT	M band	8	0.002894299	3.996837111	3.652744993
GOTERM_CC_4	M band	8	0.003115755	3.94411547	4.088390725
Annotation Cluster 9		Enrichment Score: 2.2882927468660386			
GOTERM_BP_4	regulation of cell maturation	6	0.001924634	6.065290807	3.011971142
GOTERM_BP_4	regulation of oocyte maturation	5	0.007089781	5.973392461	10.68044832
GOTERM_BP_4	regulation of oocyte development	5	0.010003245	5.475609756	14.75145964
Annotation Cluster 10		Enrichment Score: 2.237218937483266			
INTERPRO	FAD-binding, type 2, subdomain 1	4	0.001447898	13.74738841	2.355019485
INTERPRO	FAD linked oxidase, N-terminal	4	0.003423719	10.99791073	5.484763432
INTERPRO	FAD-linked oxidase, C-terminal	3	0.015064131	13.74738841	22.0931555
INTERPRO	FAD-linked oxidase-like, C-terminal	3	0.015064131	13.74738841	22.0931555

Appendix Table 3.17 DAVID functional annotation results for all genes downregulated under only heat stress.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 1.24526226167108				
GOTERM_MF_4	purine nucleoside binding	7	0.049070938	2.381481481	35.96103192
GOTERM_MF_4	purine ribonucleoside triphosphate binding	7	0.049070938	2.381481481	35.96103192
GOTERM_MF_4	ribonucleoside binding	7	0.049621579	2.375197889	36.28874467
GOTERM_MF_4	purine ribonucleotide binding	7	0.050362032	2.366871166	36.72707901
GOTERM_MF_4	purine nucleotide binding	7	0.050548267	2.364798599	36.83690435
GOTERM_MF_4	ribonucleotide binding	7	0.053395843	2.334140017	38.49523256
GOTERM_MF_4	nucleotide binding	7	0.118175624	1.911252654	67.17419949
Annotation Cluster 2	Enrichment Score: 1.1096618032995584				
GOTERM_BP_4	positive regulation of RNA metabolic process	3	0.064913188	6.834672304	57.52297285
GOTERM_BP_4	positive regulation of nucleobase-containing compound metabolic process	3	0.071932398	6.444976077	61.41587263
GOTERM_BP_4	positive regulation of macromolecule biosynthetic process	3	0.076372563	6.226502311	63.70571125
GOTERM_BP_4	positive regulation of biosynthetic process	3	0.079192381	6.097321765	65.09418238
GOTERM_BP_4	positive regulation of cellular biosynthetic process	3	0.079192381	6.097321765	65.09418238
GOTERM_BP_4	positive regulation of nitrogen compound metabolic process	3	0.080900344	6.022354694	65.91119901
GOTERM_BP_4	positive regulation of gene expression	3	0.094374654	5.503575077	71.76505096
Annotation Cluster 3	Enrichment Score: 0.9412369536912591				
GOTERM_BP_4	nucleobase-containing compound biosynthetic process	6	0.084001927	2.379683474	67.34996156
GOTERM_BP_4	cellular nitrogen compound biosynthetic process	7	0.085285465	2.103513664	67.92882599
GOTERM_BP_4	aromatic compound biosynthetic process	6	0.091618578	2.319580972	70.64904206
GOTERM_BP_4	heterocycle biosynthetic process	6	0.093087889	2.308648147	71.24896541
GOTERM_BP_4	organic cyclic compound biosynthetic process	6	0.099349036	2.264182659	73.68088442
Annotation Cluster 4	Enrichment Score: 0.5048691735741905				
INTERPRO	Protein kinase-like domain	3	0.188939978	3.611476091	88.76168465

Appendix Table 3.17 cont.

GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	3	0.37340732	2.208778626	98.40862389
GOTERM_MF_4	kinase activity	3	0.433396892	1.958375635	99.34741301
Annotation Cluster 5		Enrichment Score: 0.39829641609880023			
GOTERM_BP_4	regulation of gene expression	5	0.252598247	1.851164708	97.56268015
GOTERM_BP_4	regulation of RNA metabolic process	4	0.317921813	1.926521856	99.24107356
GOTERM_BP_4	gene expression	6	0.334832275	1.493348115	99.4490544
GOTERM_BP_4	regulation of nucleobase-containing compound metabolic process	4	0.344026314	1.841421736	99.53868772
GOTERM_BP_4	regulation of macromolecule biosynthetic process	4	0.374052803	1.752480078	99.74625142
GOTERM_BP_4	regulation of cellular biosynthetic process	4	0.383488066	1.726231478	99.79094775
GOTERM_BP_4	cellular macromolecule biosynthetic process	5	0.43294812	1.432216906	99.92806631
GOTERM_BP_4	macromolecule biosynthetic process	5	0.435772738	1.427209154	99.93250594
GOTERM_BP_4	RNA metabolic process	4	0.598859019	1.264862961	99.99913063
GOTERM_BP_4	nucleic acid metabolic process	4	0.693682802	1.108185932	99.99997214
Annotation Cluster 6		Enrichment Score: 0.26291411723558794			
GOTERM_BP_4	protein modification process	4	0.468725899	1.517639603	99.96867779
GOTERM_BP_4	cellular protein metabolic process	5	0.524384858	1.283590623	99.9923668
GOTERM_BP_4	macromolecule modification	4	0.531385157	1.388570324	99.99368235
GOTERM_BP_4	protein metabolic process	5	0.679774734	1.071812214	99.99995091
Annotation Cluster 7		Enrichment Score: 0.02407757245051425			
GOTERM_BP_DIRECT	nematode larval development	3	0.923978896	0.729239839	100
GOTERM_BP_DIRECT	reproduction	3	0.949513524	0.659126984	100
GOTERM_BP_4	larval development	3	0.954726288	0.645630292	100
GOTERM_BP_4	post-embryonic development	3	0.956417739	0.640285205	100
Annotation Cluster 8		Enrichment Score: 0.002747610737745106			
GOTERM_CC_4	integral component of membrane	7	0.988594074	0.64491839	100
GOTERM_CC_DIRECT	integral component of membrane	7	0.993732813	0.599914095	100
GOTERM_CC_DIRECT	membrane	6	0.998779304	0.50731555	100

Appendix Table 3.18 DAVID functional annotation results for all genes downregulated under only desiccation stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 9.734192582727484				
GOTERM_CC_4	intrinsic component of plasma membrane	92	2.48E-12	2.140427018	3.26E-09
GOTERM_CC_4	integral component of plasma membrane	91	4.14E-12	2.131563966	5.46E-09
GOTERM_CC_DIRECT	integral component of plasma membrane	70	6.11E-07	1.850940804	7.75E-04
Annotation Cluster 2	Enrichment Score: 7.011550804142004				
GOTERM_CC_DIRECT	neuronal cell body	28	2.67E-08	3.438994363	3.38E-05
GOTERM_CC_4	neuronal cell body	30	6.81E-08	3.130268762	8.97E-05
GOTERM_CC_4	somatodendritic compartment	32	5.09E-07	2.740931851	6.70E-04
Annotation Cluster 3	Enrichment Score: 5.848365131094264				
INTERPRO	Neurotransmitter-gated ion-channel transmembrane domain	21	1.45E-07	4.040350877	2.30E-04
INTERPRO	Neurotransmitter-gated ion-channel, conserved site	20	3.96E-07	3.980641258	6.27E-04
INTERPRO	Neurotransmitter-gated ion-channel ligand-binding	22	3.99E-07	3.662955466	6.32E-04
INTERPRO	Neurotransmitter-gated ion-channel	21	8.81E-07	3.636315789	0.001396138
GOTERM_MF_DIRECT	extracellular ligand-gated ion channel activity	22	9.81E-06	2.983888507	0.014109439
GOTERM_MF_4	ligand-gated channel activity	25	4.11E-05	2.511228276	0.051561528
Annotation Cluster 4	Enrichment Score: 4.518762562419328				
GOTERM_CC_4	M band	10	1.78E-05	6.148742211	0.023508913
GOTERM_CC_DIRECT	M band	10	2.03E-05	6.053331914	0.025737194
GOTERM_CC_4	A band	10	7.67E-05	5.217114603	0.101047203
Annotation Cluster 5	Enrichment Score: 4.024000891621929				
INTERPRO	Immunoglobulin-like domain	16	4.50E-05	3.463157895	0.071282657
INTERPRO	Immunoglobulin subtype 2	12	1.25E-04	4.074303406	0.198071825
INTERPRO	Immunoglobulin subtype	12	1.51E-04	3.995951417	0.238252142
Annotation Cluster 6	Enrichment Score: 3.845783097290106				

Appendix Table 3.18 cont.

GOTERM_BP_4	cell morphogenesis involved in differentiation	30	2.68E-06	2.637427187	0.004194429
GOTERM_BP_4	neuron projection guidance	24	5.65E-06	2.928725115	0.008857645
GOTERM_BP_4	neuron development	37	1.55E-04	1.945420568	0.24304628
GOTERM_BP_4	cell projection morphogenesis	30	3.99E-04	2.018771427	0.623328133
GOTERM_BP_4	neuron projection development	33	5.06E-04	1.913537597	0.789752373
GOTERM_BP_4	cell part morphogenesis	30	5.25E-04	1.986078772	0.8202252
GOTERM_BP_4	cell morphogenesis	36	0.004821343	1.621690766	7.293999859
Annotation Cluster 7	Enrichment Score: 3.638763340160885				
GOTERM_CC_4	plasma membrane receptor complex	17	2.64E-06	4.065001795	0.003483712
GOTERM_CC_4	acetylcholine-gated channel complex	14	2.07E-05	4.155701632	0.027305817
GOTERM_CC_DIRECT	acetylcholine-gated channel complex	14	2.46E-05	4.091217432	0.031190354
GOTERM_BP_DIRECT	chemical synaptic transmission, postsynaptic	12	9.49E-05	4.169376046	0.147278084
GOTERM_MF_4	acetylcholine receptor activity	13	2.04E-04	3.525764499	0.255981619
GOTERM_MF_DIRECT	acetylcholine-activated cation-selective channel activity	12	6.02E-04	3.385357143	0.86328582
GOTERM_MF_DIRECT	acetylcholine binding	12	7.20E-04	3.318977591	1.030856453
INTERPRO	Nicotinic acetylcholine-gated receptor, transmembrane domain	11	7.83E-04	3.593843098	1.23326358
GOTERM_MF_DIRECT	acetylcholine receptor activity	11	0.00116835	3.373091356	1.667964079
GOTERM_MF_4	ammonium ion binding	12	0.001171714	3.129376774	1.458739107
GOTERM_BP_DIRECT	cation transmembrane transport	13	0.007745712	2.385289105	11.37733315
Annotation Cluster 8	Enrichment Score: 3.0528935103976216				
GOTERM_CC_4	synaptic membrane	12	4.98E-04	3.501656581	0.654862005
GOTERM_CC_4	postsynaptic membrane	10	0.001115081	3.74271265	1.459535828
GOTERM_CC_DIRECT	postsynaptic membrane	10	0.001248534	3.684636817	1.57193015
Annotation Cluster 9	Enrichment Score: 3.0056769814043722				
INTERPRO	Spectrin repeat	5	3.34E-04	12.36842105	0.528458544
INTERPRO	Spectrin/alpha-actinin	5	3.34E-04	12.36842105	0.528458544
INTERPRO	Actinin-type, actin-binding, conserved site	4	0.00860423	8.657894737	12.79820418
Annotation Cluster 10	Enrichment Score: 3.0008323683571954				

Appendix Table 3.18 cont.

INTERPRO	Calmodulin-binding domain	4	7.32E-04	17.31578947	1.153885685
INTERPRO	Potassium channel, calcium-activated, SK small conductance calcium-activated potassium channel	4	7.32E-04	17.31578947	1.153885685
GOTERM_MF_DIRECT	activity	4	0.001338267	14.10565476	1.908374236
GOTERM_MF_DIRECT	calcium-activated potassium channel activity	5	0.001383581	8.816034226	1.972394778

Appendix Table 3.19 DAVID functional annotation results for all genes downregulated under only freezing stress. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 2.9607095398013454			
INTERPRO	FAD-binding, type 2, subdomain 1	4	6.24E-05	39.55191257	0.094154276
INTERPRO	CO dehydrogenase flavoprotein-like, FAD-binding, subdomain 2	4	8.10E-04	19.77595628	1.215866025
INTERPRO	FAD-binding, type 2	4	8.10E-04	19.77595628	1.215866025
INTERPRO	FAD-linked oxidase-like, C-terminal	3	0.001870362	39.55191257	2.785874438
INTERPRO	FAD-linked oxidase, C-terminal	3	0.001870362	39.55191257	2.785874438
GOTERM_MF_DIRECT	oxidoreductase activity, acting on CH-OH group of donors	4	0.002180065	14.23273273	2.921902479
INTERPRO	FAD linked oxidase, N-terminal	3	0.006028444	23.73114754	8.721721161
Annotation Cluster 2		Enrichment Score: 2.7514124971864993			
GOTERM_BP_DIRECT	phosphorylation	26	1.02E-05	2.723147047	0.015073671
GOTERM_MF_DIRECT	kinase activity	23	7.30E-04	2.205221314	0.987557789
GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	26	0.025005449	1.562673314	26.01784201
GOTERM_MF_4	kinase activity	27	0.052784358	1.438806589	47.55024179
Annotation Cluster 3		Enrichment Score: 2.4051976905060903			
GOTERM_BP_4	methionine metabolic process	5	6.60E-04	11.85739437	0.997100222
GOTERM_BP_DIRECT	methionine biosynthetic process	4	0.001024711	18.22413793	1.497621292
GOTERM_BP_DIRECT	cellular amino acid biosynthetic process	4	0.012192111	8.099616858	16.51866837
GOTERM_BP_DIRECT	L-methionine biosynthetic process from methylthioadenosine	3	0.029024785	10.93448276	35.17707278
Annotation Cluster 4		Enrichment Score: 1.7441607919867892			
GOTERM_CC_DIRECT	endoplasmic reticulum membrane	10	0.009146148	2.814820315	9.981378488
GOTERM_CC_4	endoplasmic reticulum part	12	0.020925878	2.19581044	23.28457387
GOTERM_CC_4	endoplasmic reticulum membrane	11	0.022076241	2.288695822	24.40672198
GOTERM_CC_4	nuclear outer membrane-endoplasmic reticulum membrane network	11	0.024975247	2.243149636	27.16796712

Appendix Table 3.19 cont.

Annotation Cluster 5		Enrichment Score: 1.6793922956915996			
GOTERM_BP_4	regulation of oocyte maturation	4	0.002528626	13.79769526	3.768163921
GOTERM_BP_4	regulation of oocyte development	4	0.003306449	12.64788732	4.900332153
GOTERM_BP_4	regulation of cell maturation	4	0.004215558	11.67497291	6.207820848
GOTERM_BP_4	oocyte development	5	0.059043975	3.387826962	60.27859487
GOTERM_BP_4	oocyte differentiation	5	0.062247285	3.328391401	62.28130664
GOTERM_BP_4	cell maturation	4	0.097388339	3.613682093	78.86938853
GOTERM_BP_4	oocyte maturation	3	0.139107186	4.553239437	89.69321506
Annotation Cluster 6		Enrichment Score: 1.529166347087554			
INTERPRO	Cytochrome P450, conserved site	7	0.008915577	3.899484338	12.64224292
INTERPRO	Cytochrome P450, E-class, group I	7	0.010164485	3.79264915	14.28896294
INTERPRO	Cytochrome P450	7	0.012263745	3.642939316	16.99175065
GOTERM_MF_DIRECT	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	7	0.048798444	2.637241653	49.3273128
GOTERM_MF_DIRECT	heme binding	9	0.107545921	1.871511934	78.69037975
GOTERM_MF_4	heme binding	9	0.114589294	1.840551285	76.50219329
Annotation Cluster 7		Enrichment Score: 1.4608465464294533			
GOTERM_BP_4	oligosaccharide metabolic process	4	0.021519677	6.598897734	28.11017841
GOTERM_BP_DIRECT	N-glycan processing	3	0.041070102	9.112068966	46.05666136
GOTERM_BP_4	protein N-linked glycosylation	4	0.046892008	4.895956383	51.74210267
Annotation Cluster 8		Enrichment Score: 1.4565871679307534			
GOTERM_BP_4	negative regulation of intracellular signal transduction	6	0.016600122	3.994069681	22.42705095
GOTERM_BP_4	negative regulation of signal transduction	7	0.047795559	2.656056338	52.43150669
GOTERM_BP_4	negative regulation of cell communication	7	0.053794723	2.578695474	56.78068052
Annotation Cluster 9		Enrichment Score: 1.4314244614668197			
GOTERM_BP_4	negative regulation of nematode larval development	13	0.020080048	2.11702835	26.48870982
GOTERM_BP_4	negative regulation of post-embryonic development	13	0.020080048	2.11702835	26.48870982
GOTERM_BP_4	regulation of post-embryonic development	16	0.020793117	1.909115068	27.2960688
GOTERM_BP_DIRECT	negative regulation of vulval development	12	0.039853955	1.988087774	45.04095188

Appendix Table 3.19 cont.

GOTERM_BP_4	vulval development	15	0.084940496	1.603253323	73.98917019
GOTERM_BP_4	post-embryonic organ development	15	0.090869457	1.585389776	76.43184388
Annotation Cluster 10	Enrichment Score: 1.416151586170258				
GOTERM_BP_4	pteridine-containing compound metabolic process	4	0.005262008	10.84104628	7.692041705
GOTERM_BP_4	folic acid-containing compound metabolic process	3	0.032326741	10.34827145	39.2574457
GOTERM_BP_4	cellular modified amino acid metabolic process	3	0.331765962	2.529577465	99.77918914

Appendix Table 3.20 DAVID functional annotation results for all genes downregulated under both heat and desiccation stresses. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 2.0712923038485074				
GOTERM_BP_4	protein transport	6	0.005363871	5.101065089	6.968284356
GOTERM_BP_4	establishment of protein localization	6	0.006557271	4.861718926	8.456230781
GOTERM_BP_4	organic substance transport	7	0.017374874	3.228971363	20.97393171
Annotation Cluster 2	Enrichment Score: 1.145260305102106				
GOTERM_MF_DIRECT	GTPase activator activity	3	0.031853227	10.41648352	28.46749112
GOTERM_MF_4	GTPase activator activity	3	0.039824486	9.185714286	31.489798
GOTERM_MF_4	GTPase regulator activity	3	0.047314382	8.350649351	36.30498573
GOTERM_BP_4	positive regulation of catalytic activity	3	0.158939032	4.103579589	90.21757841
GOTERM_BP_4	regulation of hydrolase activity	3	0.196866186	3.572944297	94.73596781
Annotation Cluster 3	Enrichment Score: 0.9096040587938682				
GOTERM_MF_DIRECT	nucleotide binding	9	0.044343896	2.159389475	37.46274921
GOTERM_MF_DIRECT	ATP binding	8	0.080292827	2.046747638	57.94630624
GOTERM_MF_4	purine nucleoside binding	9	0.131468822	1.701058201	73.06368677
GOTERM_MF_4	purine ribonucleoside triphosphate binding	9	0.131468822	1.701058201	73.06368677
GOTERM_MF_4	ribonucleoside binding	9	0.132956722	1.696569921	73.49007055
GOTERM_MF_4	purine ribonucleotide binding	9	0.134954954	1.690622261	74.05321869
GOTERM_MF_4	purine nucleotide binding	9	0.135457071	1.689141856	74.19303743
GOTERM_MF_4	ribonucleotide binding	9	0.143110892	1.667242869	76.2426638
GOTERM_MF_4	nucleotide binding	9	0.304112195	1.365180467	96.57490311
Annotation Cluster 4	Enrichment Score: 0.6570886992353261				
GOTERM_BP_4	cellular chemical homeostasis	3	0.044869423	8.634615385	46.01833795
GOTERM_BP_4	chemical homeostasis	3	0.110082097	5.14859054	79.11768074
GOTERM_BP_4	cellular homeostasis	3	0.123924281	4.791462872	83.08217739

Appendix Table 3.20 cont.

GOTERM_BP_4	ion transport	4	0.478541833	1.518174134	99.98406663
GOTERM_BP_4	ion transmembrane transport	3	0.542870307	1.625339367	99.99728112
GOTERM_MF_4	ion transmembrane transporter activity	3	0.717840939	1.181985294	99.99923033
Annotation Cluster 5		Enrichment Score: 0.5048866717117628			
GOTERM_BP_DIRECT	germ cell development	3	0.129350021	4.674484053	80.98167144
GOTERM_BP_4	gamete generation	4	0.333884968	1.912163961	99.57312814
GOTERM_BP_4	cellular process involved in reproduction in multicellular organism	3	0.470484122	1.858571921	99.98042311
GOTERM_BP_4	germ cell development	3	0.470484122	1.858571921	99.98042311
Annotation Cluster 6		Enrichment Score: 0.48786267874301137			
GOTERM_BP_4	neuron projection development	3	0.262122502	2.939443535	98.31324457
GOTERM_BP_4	neuron development	3	0.300303602	2.665347514	99.17365014
GOTERM_BP_4	nervous system development	3	0.436864865	1.983069562	99.95525219
Annotation Cluster 7		Enrichment Score: 0.3783264618741519			
GOTERM_CC_DIRECT	integral component of plasma membrane	4	0.378338399	1.77425897	98.99341862
GOTERM_CC_4	integral component of plasma membrane	4	0.437898241	1.611035525	99.78437644
GOTERM_CC_4	intrinsic component of plasma membrane	4	0.442350978	1.60015015	99.80189939
Annotation Cluster 8		Enrichment Score: 0.2886700553690473			
GOTERM_BP_DIRECT	phosphorylation	3	0.33753902	2.44470143	99.28048
GOTERM_MF_DIRECT	kinase activity	3	0.389256248	2.183095348	99.39211145
GOTERM_BP_DIRECT	protein phosphorylation	3	0.442082762	1.964794683	99.90811226
INTERPRO	Protein kinase-like domain	3	0.46599914	1.880977131	99.93785868
GOTERM_MF_DIRECT	transferase activity	4	0.663593576	1.162924795	99.99873104
GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	3	0.697811492	1.227099237	99.99854295
GOTERM_MF_4	kinase activity	3	0.760713369	1.087986464	99.99983395
Annotation Cluster 9		Enrichment Score: 0.18238607829642642			
GOTERM_BP_DIRECT	nematode larval development	8	0.541229731	1.138325603	99.99118859
GOTERM_BP_4	larval development	8	0.719617296	0.971204542	99.99999617
GOTERM_BP_4	post-embryonic development	8	0.728380862	0.963164069	99.9999975

Appendix Table 3.20 cont.

Annotation Cluster 10	Enrichment Score: 0.15814786808690545				
GOTERM_BP_4	cell migration	3	0.289780465	2.735719726	98.99021442
GOTERM_BP_4	cell motility	3	0.351208939	2.368351648	99.70035877
GOTERM_BP_DIRECT	hermaphrodite genitalia development	3	0.731179251	1.15278604	99.99998543
GOTERM_BP_4	hermaphrodite genitalia development	3	0.748653352	1.114143921	99.99999912
GOTERM_BP_4	genitalia development	3	0.753415281	1.10375909	99.99999932
GOTERM_BP_4	animal organ development	4	0.83850001	0.875777155	100
GOTERM_BP_4	development of primary sexual characteristics	3	0.869170666	0.852801519	100
GOTERM_BP_4	gonad development	3	0.869170666	0.852801519	100
GOTERM_BP_4	reproductive system development	3	0.873409546	0.843258471	100
GOTERM_BP_4	reproductive structure development	3	0.873409546	0.843258471	100
GOTERM_BP_4	sex differentiation	3	0.897946923	0.786454532	100

Appendix Table 3.21 DAVID functional annotation results for all genes downregulated under both heat and freezing stresses.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 1.489613140970943			
GOTERM_BP_DIRECT	nematode larval development	7	0.005305607	3.403119251	4.634010428
GOTERM_BP_4	larval development	7	0.015222354	2.761862917	15.78695676
GOTERM_BP_4	post-embryonic development	7	0.015858119	2.738997821	16.39393686
GOTERM_BP_DIRECT	embryo development ending in birth or egg hatching	6	0.141570101	1.880782918	74.37317165
GOTERM_BP_4	embryo development	6	0.196559584	1.702369668	91.38286124
Annotation Cluster 2		Enrichment Score: 1.080320947872462			
GOTERM_BP_4	protein complex assembly	3	0.06857336	6.27972028	54.87402312
GOTERM_BP_4	protein complex subunit organization	3	0.076333042	5.907894737	58.91060639
GOTERM_BP_4	macromolecular complex assembly	3	0.109690536	4.785079929	72.78574764
Annotation Cluster 3		Enrichment Score: 0.556456118381719			
GOTERM_MF_DIRECT	DNA binding	3	0.113384293	4.421175373	58.161358
GOTERM_MF_4	DNA binding	3	0.120999004	3.99378882	50.64100875
GOTERM_CC_DIRECT	nucleus	3	0.616449232	1.390115711	99.93882894
GOTERM_CC_4	nucleus	3	0.70298089	1.197954137	99.99920616
Annotation Cluster 4		Enrichment Score: 0.522224679268457			
GOTERM_CC_4	intracellular organelle part	5	0.127271269	2.265518707	73.20411398
GOTERM_CC_4	intracellular membrane-bounded organelle	6	0.302579216	1.476038781	96.93821352
GOTERM_CC_4	intracellular organelle	6	0.458471637	1.259990541	99.73511296
GOTERM_CC_4	intracellular part	7	0.461550623	1.200808061	99.7493286
Annotation Cluster 5		Enrichment Score: 0.033823793497920			
GOTERM_CC_DIRECT	membrane	6	0.870708794	0.858534008	99.99998619
GOTERM_CC_DIRECT	integral component of membrane	5	0.952257651	0.725170883	99.99999999
GOTERM_CC_4	integral component of membrane	5	0.954775394	0.729371989	100

Appendix Table 3.22 DAVID functional annotation results for all genes downregulated under both desiccation and freezing stresses.

Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 4.081795950772048				
GOTERM_CC_4	myosin complex	9	4.14E-05	6.611042184	0.052891141
GOTERM_CC_DIRECT	myosin complex	8	5.01E-05	7.674089069	0.061185928
INTERPRO	Myosin head, motor domain	7	7.72E-05	8.950008832	0.120447297
GOTERM_MF_DIRECT	motor activity	8	2.94E-04	5.810881226	0.417462993
Annotation Cluster 2	Enrichment Score: 3.9350858919790386				
GOTERM_CC_4	A band	9	6.75E-05	6.21037296	0.086151923
GOTERM_CC_DIRECT	M band	8	1.47E-04	6.57779063	0.179909892
GOTERM_CC_4	M band	8	1.57E-04	6.506105006	0.200919703
Annotation Cluster 3	Enrichment Score: 2.9414567332487223				
GOTERM_CC_4	integral component of plasma membrane	52	6.86E-04	1.611035525	0.872497254
GOTERM_CC_4	intrinsic component of plasma membrane	52	7.85E-04	1.60015015	0.998277572
GOTERM_CC_DIRECT	integral component of plasma membrane	44	0.002783595	1.580311634	3.34675578
Annotation Cluster 4	Enrichment Score: 2.6705394911735434				
GOTERM_MF_DIRECT	peptidase activity	26	5.65E-04	2.107741516	0.801084703
GOTERM_BP_DIRECT	proteolysis	32	0.001166634	1.848489666	1.756586725
GOTERM_MF_4	peptidase activity, acting on L-amino acid peptides	34	0.014779685	1.520596776	16.93456074
Annotation Cluster 5	Enrichment Score: 2.58190908749288				
GOTERM_BP_4	muscle cell differentiation	25	3.22E-06	2.970395281	0.004982512
GOTERM_BP_4	myofibril assembly	20	1.38E-04	2.709580085	0.213597721
GOTERM_BP_4	muscle cell development	20	2.58E-04	2.583553105	0.398707048
GOTERM_BP_4	striated muscle myosin thick filament assembly	17	9.82E-04	2.552131513	1.507189045

190

Appendix Table 3.22 cont.

GOTERM_BP_DIRECT	striated muscle myosin thick filament assembly	16	0.003384025	2.342264303	5.016104351
GOTERM_BP_4	cellular component assembly involved in morphogenesis	20	0.005045295	2.010729113	7.520901257
GOTERM_BP_4	actin cytoskeleton organization	21	0.005539799	1.95225812	8.22885284
GOTERM_BP_4	cytoskeleton organization	27	0.027993161	1.538207772	35.52483173
GOTERM_BP_4	organelle assembly	22	0.103453542	1.404621401	81.51281412
GOTERM_BP_4	single-organism organelle organization	28	0.489448973	1.052655817	99.99693188
Annotation Cluster 6		Enrichment Score: 2.534876179985385			
GOTERM_CC_4	vacuolar proton-transporting V-type ATPase complex	5	0.001360397	9.488069801	1.724112091
GOTERM_CC_4	proton-transporting V-type ATPase complex	6	0.001418652	6.831410256	1.797326401
GOTERM_CC_DIRECT	vacuolar proton-transporting V-type ATPase complex	4	0.001467911	15.34817814	1.777922498
GOTERM_MF_DIRECT	hydrogen ion transmembrane transporter activity	5	0.025600818	4.323572341	30.87897768
Annotation Cluster 7		Enrichment Score: 2.4684047011686014			
GOTERM_BP_4	cellular nitrogen compound catabolic process	14	0.002410739	2.636099948	3.662271997
GOTERM_BP_4	heterocycle catabolic process	14	0.002410739	2.636099948	3.662271997
GOTERM_BP_4	organic cyclic compound catabolic process	14	0.003236309	2.549670441	4.887326324
GOTERM_BP_4	aromatic compound catabolic process	13	0.00711264	2.427237287	10.44703918
Annotation Cluster 8		Enrichment Score: 2.4079107183635386			
INTERPRO	ABC transporter-like	10	7.16E-04	4.037597969	1.111378481
INTERPRO	ABC transporter, conserved site	9	0.00128375	4.142575517	1.984658686
GOTERM_MF_DIRECT	ATPase activity, coupled to transmembrane movement of substances	10	0.002475868	3.362778487	3.468502599
GOTERM_MF_DIRECT	ATPase activity	13	0.008498031	2.360670498	11.44370165
INTERPRO	AAA+ ATPase domain	10	0.047212973	2.111404443	52.98638344
Annotation Cluster 9		Enrichment Score: 1.9129856663784628			
GOTERM_CC_4	proton-transporting V-type ATPase complex	6	0.001418652	6.831410256	1.797326401
GOTERM_BP_4	ATP hydrolysis coupled transmembrane transport	6	0.003743039	5.554639175	5.632024727
GOTERM_BP_DIRECT	ATP hydrolysis coupled proton transport	6	0.004427804	5.343290441	6.515210328
GOTERM_CC_4	proton-transporting two-sector ATPase complex	7	0.008371264	3.887794455	10.18305859
KEGG_PATHWAY	Phagosome	7	0.021714244	3.136363636	21.36357984

Appendix Table 3.22 cont.

GOTERM_MF_DIRECT	proton-transporting ATPase activity, rotational mechanism	5	0.034752111	3.947609529	39.56999053
KEGG_PATHWAY	Oxidative phosphorylation	7	0.273712618	1.582309582	96.98360994
Annotation Cluster 10	Enrichment Score: 1.681202278686573				
GOTERM_CC_4	Golgi stack	6	0.015394489	4.018476621	17.97947207
GOTERM_CC_4	Golgi subcompartment	7	0.021294728	3.187991453	24.0418391
GOTERM_CC_4	organelle subcompartment	7	0.027590259	3.007539107	30.05290252

Appendix Table 3.23 DAVID functional annotation results for all genes downregulated under heat, desiccation, and freezing stresses. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1	Enrichment Score: 2.5808034929957766				
GOTERM_BP_DIRECT	monocarboxylic acid transport	3	0.001306225	53.1722561	1.665175958
GOTERM_MF_DIRECT	monocarboxylic acid transmembrane transporter activity	3	0.00181266	44.99525316	2.107291298
GOTERM_MF_4	organic acid transmembrane transporter activity	4	0.007642848	9.690423862	7.736365183
Annotation Cluster 2	Enrichment Score: 2.493156474968349				
GOTERM_CC_DIRECT	integral component of plasma membrane	12	0.001534498	3.041586806	1.441009026
GOTERM_CC_4	integral component of plasma membrane	12	0.004530996	2.636239951	4.713230899
GOTERM_CC_4	intrinsic component of plasma membrane	12	0.004768376	2.618427518	4.954516277
Annotation Cluster 3	Enrichment Score: 2.1297204620340615				
INTERPRO	ABC transporter, transmembrane domain, type 1	4	0.001728944	16.56292906	2.191042795
INTERPRO	ABC transporter, conserved site	4	0.003804781	12.58782609	4.763200244
INTERPRO	ABC transporter-like	4	0.005503435	11.04195271	6.821440275
GOTERM_MF_4	hydrolase activity, acting on acid anhydrides, catalyzing transmembrane movement of substances	5	0.008998105	5.995337995	9.050241148
GOTERM_MF_DIRECT	ATPase activity	5	0.0091653	5.999367089	10.2450929
GOTERM_MF_DIRECT	ATPase activity, coupled to transmembrane movement of substances	4	0.009839563	8.887951242	10.95947552
GOTERM_MF_4	primary active transmembrane transporter activity	5	0.009972573	5.819004525	9.984467292
INTERPRO	AAA+ ATPase domain	4	0.031290406	5.77423215	33.43544551
Annotation Cluster 4	Enrichment Score: 1.6711700479616796				
INTERPRO	Pyridoxal phosphate-dependent transferase, major region, subdomain 2	3	0.014212451	16.27736132	16.74490548
INTERPRO	Pyridoxal phosphate-dependent transferase, major region, subdomain 1	3	0.026116228	11.80108696	28.73733386
INTERPRO	Pyridoxal phosphate-dependent transferase	3	0.026116228	11.80108696	28.73733386

Appendix Table 3.23 cont.

Annotation Cluster 5		Enrichment Score: 0.9253756456887949			
GOTERM_BP_4	neurotransmitter transport	4	0.015818535	7.463896104	20.30164735
GOTERM_BP_4	neurotransmitter secretion	3	0.04283431	8.98	46.36874601
GOTERM_BP_4	signal release	3	0.049483009	8.289230769	51.43361647
GOTERM_BP_4	synaptic signaling	3	0.371234073	2.292765957	99.86442076
GOTERM_BP_4	secretion	3	0.415052535	2.092427184	99.95150241
GOTERM_BP_4	cell-cell signaling	3	0.542726561	1.638935361	99.99854189
Annotation Cluster 6		Enrichment Score: 0.7481208437884848			
GOTERM_MF_DIRECT	nucleotide binding	13	0.124349675	1.539817811	78.96121728
GOTERM_MF_4	purine ribonucleotide binding	14	0.142163581	1.456536102	79.99782601
GOTERM_MF_4	purine nucleotide binding	14	0.14285574	1.455260676	80.16655896
GOTERM_MF_4	ribonucleotide binding	14	0.153467816	1.436393857	82.59735757
GOTERM_MF_DIRECT	ATP binding	12	0.154359091	1.51562958	86.02870266
GOTERM_MF_4	purine ribonucleoside triphosphate binding	13	0.223272872	1.360846561	92.94717699
GOTERM_MF_4	purine nucleoside binding	13	0.223272872	1.360846561	92.94717699
GOTERM_MF_4	ribonucleoside binding	13	0.226009757	1.357255937	93.2036738
GOTERM_MF_4	nucleotide binding	15	0.274304588	1.260166585	96.54370109
Annotation Cluster 7		Enrichment Score: 0.7383666849095571			
INTERPRO	Protein kinase, ATP binding site	6	0.023626785	3.645123384	26.36969621
INTERPRO	Protein kinase, catalytic domain	6	0.119402264	2.280403277	80.36582872
GOTERM_BP_DIRECT	protein phosphorylation	6	0.183707273	1.964794683	92.62961916
INTERPRO	Protein kinase-like domain	6	0.185356742	1.962758745	92.75301334
GOTERM_MF_DIRECT	protein serine/threonine kinase activity	5	0.217578276	2.061638175	94.38818426
GOTERM_MF_DIRECT	protein kinase activity	6	0.244715733	1.764519732	96.29200439
GOTERM_MF_4	phosphotransferase activity, alcohol group as acceptor	6	0.441037848	1.359248385	99.77676226
GOTERM_MF_4	kinase activity	6	0.549238918	1.205154237	99.9766572
Annotation Cluster 8		Enrichment Score: 0.6428443784800345			

Appendix Table 3.23 cont.

GOTERM_CC_DIRECT	membrane	43	0.142813931	1.142667882	76.69555915
GOTERM_CC_4	integral component of membrane	42	0.199170195	1.113949946	90.56981158
GOTERM_CC_DIRECT	integral component of membrane	39	0.414449806	1.050461823	99.36458038
Annotation Cluster 9	Enrichment Score: 0.6314823468491184				
GOTERM_BP_4	positive regulation of catalytic activity	4	0.161471059	2.845148515	91.84270671
GOTERM_BP_DIRECT	positive regulation of GTPase activity	3	0.166042902	4.051219512	90.29628621
GOTERM_BP_4	regulation of hydrolase activity	3	0.475595733	1.857931034	99.98975704
Annotation Cluster 10	Enrichment Score: 0.6043106739486809				
GOTERM_BP_4	positive regulation of multicellular organism growth	5	0.2296957	2.01232493	97.56200309
GOTERM_BP_DIRECT	positive regulation of multicellular organism growth	5	0.237510422	1.985891918	96.93051048
GOTERM_BP_4	positive regulation of developmental growth	5	0.241557574	1.968219178	98.04513415
GOTERM_BP_4	multicellular organism growth	5	0.26417383	1.890526316	98.72939753
GOTERM_BP_4	regulation of multicellular organism growth	5	0.273338397	1.861139896	98.93696561

Appendix Table 4.1 List of DNA methyltransferase genes used in reciprocal Blastp analyses.

DNA (5-cytosine-)-methyltransferase genes		DNA N6-methyltransferase genes	
Accession Number	Description	Accession Number	Description
NP_001305660.1	DNA (cytosine-5)-methyltransferase 1 isoform d [Homo sapiens]	XP_003373027.1	n(6)-adenine-specific DNA methyltransferase 2-like protein [Trichinella spiralis]
NP_001305659.1	DNA (cytosine-5)-methyltransferase 1 isoform c [Homo sapiens]	XP_003373020.1	N(6)-adenine-specific DNA methyltransferase 2, partial [Trichinella spiralis]
NP_001124295.1	DNA (cytosine-5)-methyltransferase 1 isoform a [Homo sapiens]	XP_003372519.1	putative N(6)-adenine-specific DNA methyltransferase 1 [Trichinella spiralis]
NP_001370.1	DNA (cytosine-5)-methyltransferase 1 isoform b [Homo sapiens]	XP_003373021.1	n(6)-adenine-specific DNA methyltransferase 2, partial [Trichinella spiralis]
NP_787045.1	DNA (cytosine-5)-methyltransferase 3B isoform 3 [Homo sapiens]	NP_495127.1	DNA N6-methyl methyltransferase [Caenorhabditis elegans]
NP_001193985.1	DNA (cytosine-5)-methyltransferase 3B isoform 8 [Homo sapiens]	XP_023954455.1	DNA N6-methyl methyltransferase [Bicyclus anynana]
NP_001193984.1	DNA (cytosine-5)-methyltransferase 3B isoform 7 [Homo sapiens]	XP_013187493.1	PREDICTED: DNA N6-methyl methyltransferase [Amyeloidis transitella]

Appendix Table 4.1 cont.

NP_787046.1	DNA (cytosine-5)-methyltransferase 3B isoform 6 [Homo sapiens]	XP_013163157.1	PREDICTED: DNA N6-methyl methyltransferase [Papilio xuthus]
NP_787044.1	DNA (cytosine-5)-methyltransferase 3B isoform 2 [Homo sapiens]		
NP_008823.1	DNA (cytosine-5)-methyltransferase 3B isoform 1 [Homo sapiens]		
NP_001307822.1	DNA (cytosine-5)-methyltransferase 3A isoform d [Homo sapiens]		
NP_072046.2	DNA (cytosine-5)-methyltransferase 3A isoform a [Homo sapiens]		
NP_715640.2	DNA (cytosine-5)-methyltransferase 3A isoform b [Homo sapiens]		
NP_872592.2	DNA (cytosine-5)-methyltransferase 1 [Bos taurus]		
NP_477475.2	methyltransferase 2, isoform D [Drosophila melanogaster]		
NP_001036355.1	methyltransferase 2, isoform C [Drosophila melanogaster]		
NP_001036980.1	DNA cytosine-5 methyltransferase [Bombyx mori]		

Appendix Table 4.1 cont.

NP_001177350.1	DNA methyltransferase 3 [Apis mellifera]
XP_026302146.1	DNA methyltransferase 3 isoform X1 [Apis mellifera]
XP_026302147.1	DNA methyltransferase 3 isoform X2 [Apis mellifera]
XP_026302148.1	DNA methyltransferase 3 isoform X3 [Apis mellifera]
XP_026302149.1	DNA methyltransferase 3 isoform X4 [Apis mellifera]
XP_026302150.1	DNA methyltransferase 3 isoform X5 [Apis mellifera]
NP_571461.1	DNA (cytosine-5-)-methyltransferase 3 beta, duplicate b.2 [Danio rerio]
NP_001018144.1	DNA (cytosine-5-)-methyltransferase 3 alpha a [Danio rerio]
NP_001018315.1	DNA methyltransferase dnmt5 [Danio rerio]
NP_001018312.2	DNA (cytosine-5-)-methyltransferase 3 beta, duplicate a [Danio rerio]

Appendix Table 4.1 cont.

NP_001186361.1	DNA (cytosine-5)-methyltransferase 1 isoform 3 [Mus musculus]
NP_034198.3	DNA (cytosine-5)-methyltransferase 3B isoform 3 [Mus musculus]
NP_001258682.1	DNA (cytosine-5)-methyltransferase 3A isoform 1 [Mus musculus]
NP_714965.1	DNA (cytosine-5)-methyltransferase 3A isoform 2 [Mus musculus]
NP_001258675.1	DNA (cytosine-5)-methyltransferase 3B isoform 4 [Mus musculus]
NP_001258674.1	DNA (cytosine-5)-methyltransferase 3B isoform 2 [Mus musculus]
NP_001258673.1	DNA (cytosine-5)-methyltransferase 3B isoform 1 [Mus musculus]
NP_001116469.1	DNA (cytosine-5)-methyltransferase 3B isoform 5 [Mus musculus]
NP_001186360.2	DNA (cytosine-5)-methyltransferase 1 isoform 1 [Mus musculus]
NP_034196.5	DNA (cytosine-5)-methyltransferase 1 isoform 2 [Mus musculus]

Appendix Table 4.1 cont.

NP_001300940.1	DNA (cytosine-5)-methyltransferase 1 isoform 5 [Mus musculus]
NP_001186362.1	DNA (cytosine-5)-methyltransferase 1 isoform 4 [Mus musculus]
NP_445806.3	DNA (cytosine-5)-methyltransferase 1 [Rattus norvegicus]
NP_001027526.1	DNA (cytosine-5)-methyltransferase 1 [Sus scrofa]
NP_001009473.1	DNA (cytosine-5)-methyltransferase 1 [Ovis aries]
NP_996835.1	DNA (cytosine-5)-methyltransferase 1 [Gallus gallus]

Appendix Table 4.4 NCBI Blastp results from mapping DNA (5-cytosine-)-methyltransferase (DNMT) genes against *Oscheius tipulae* predicted genes. DNMT genes were obtained from NCBI, and the *O. tipulae* predicted genes were obtained after Illumina HiSeq high throughput sequencing, genome annotation via CLC Genomics, and gene prediction via Augustus.

Methyltransferase gene query accession	<i>O. tipulae</i> subject	Percent ID	Alignment length (aa)	E-value	Methyltransferase query description
NP_001305660.1	g13819	51.16	43	1.00E-07	DNA (cytosine-5)-methyltransferase 1 isoform d [Homo sapiens]
NP_001305659.1	g13819	51.16	43	1.00E-07	DNA (cytosine-5)-methyltransferase 1 isoform c [Homo sapiens]
NP_001124295.1	g13819	51.16	43	1.00E-07	DNA (cytosine-5)-methyltransferase 1 isoform a [Homo sapiens]
NP_001370.1	g13819	51.16	43	1.00E-07	DNA (cytosine-5)-methyltransferase 1 isoform b [Homo sapiens]
NP_872592.2	g13819	48.84	43	6.00E-07	DNA (cytosine-5)-methyltransferase 1 [Bos taurus]
NP_477475.2	g7662	30.58	327	6.00E-44	methyltransferase 2, isoform D [Drosophila melanogaster]
NP_001036355.1	g7662	30.58	327	6.00E-44	methyltransferase 2, isoform C [Drosophila melanogaster]
NP_001036980.1	g13819	44.19	43	1.00E-07	DNA cytosine-5 methyltransferase [Bombyx mori]

201

Appendix Table 4.4 cont.

NP_571461.1	g18780	31.19	109	1.00E-14	DNA (cytosine-5-)-methyltransferase 3 beta, duplicate b.2 [Danio rerio]
NP_001018312.2	g18780	41.09	129	5.00E-28	DNA (cytosine-5-)-methyltransferase 3 beta, duplicate a [Danio rerio]
NP_001186361.1	g13819	38.98	59	7.00E-08	DNA (cytosine-5-)-methyltransferase 1 isoform 3 [Mus musculus]
NP_001186360.2	g13819	38.98	59	7.00E-08	DNA (cytosine-5-)-methyltransferase 1 isoform 1 [Mus musculus]
NP_034196.5	g13819	38.98	59	7.00E-08	DNA (cytosine-5-)-methyltransferase 1 isoform 2 [Mus musculus]
NP_001300940.1	g13819	38.98	59	7.00E-08	DNA (cytosine-5-)-methyltransferase 1 isoform 5 [Mus musculus]
NP_001186362.1	g13819	38.98	59	7.00E-08	DNA (cytosine-5-)-methyltransferase 1 isoform 4 [Mus musculus]
NP_445806.3	g13819	48.84	43	1.00E-07	DNA (cytosine-5-)-methyltransferase 1 [Rattus norvegicus]
NP_001027526.1	g13819	51.16	43	1.00E-07	DNA (cytosine-5-)-methyltransferase 1 [Sus scrofa]
NP_001009473.1	g13819	48.84	43	4.00E-06	DNA (cytosine-5-)-methyltransferase 1 [Ovis aries]
NP_996835.1	g13819	48.84	43	1.00E-07	DNA (cytosine-5-)-methyltransferase 1 [Gallus gallus]

202

Appendix Table 4.5 NCBI Blastp results from blasting the *Oscheius tipulae* predicted genes obtained as a hit in initial blast (Appendix Table 4.4) against DNA (5-cytosine-)-methyltransferase (DNMT) genes. *O. tipulae* predicted genes were obtained after Illumina HiSeq high throughput sequencing, annotation, and gene prediction. DNMT genes were obtained from NCBI.

<i>O. tipulae</i> gene query accession	Methyltransferase subject	Percent ID	Alignment length (aa)	E-value	Methyltransferase hit description
g7662	NP_001036355.1	30.58	327	4.00E-46	methyltransferase 2, isoform C [Drosophila melanogaster]
g7662	NP_477475.2	30.58	327	4.00E-46	methyltransferase 2, isoform D [Drosophila melanogaster]
g7662	NP_996835.1	24.6	187	7.00E-08	DNA (cytosine-5)-methyltransferase 1 [Gallus gallus]
g7662	NP_001036980.1	23.94	188	1.00E-07	DNA cytosine-5 methyltransferase [Bombyx mori]
g7662	NP_001027526.1	25.13	187	1.00E-07	DNA (cytosine-5)-methyltransferase 1 [Sus scrofa]
g7662	NP_872592.2	25.13	187	2.00E-07	DNA (cytosine-5)-methyltransferase 1 [Bos taurus]
g7662	NP_001009473.1	25.13	187	3.00E-07	DNA (cytosine-5)-methyltransferase 1 [Ovis aries]

Appendix Table 4.5 cont.

g7662	NP_445806.3	24.06	187	5.00E-07	DNA (cytosine-5)-methyltransferase 1 [Rattus norvegicus]
g7662	NP_001305659.1	24.06	187	9.00E-07	DNA (cytosine-5)-methyltransferase 1 isoform c [Homo sapiens]
g7662	NP_001370.1	24.06	187	9.00E-07	DNA (cytosine-5)-methyltransferase 1 isoform b [Homo sapiens]
g7662	NP_001305660.1	24.06	187	1.00E-06	DNA (cytosine-5)-methyltransferase 1 isoform d [Homo sapiens]
204 g7662	NP_001124295.1	24.06	187	1.00E-06	DNA (cytosine-5)-methyltransferase 1 isoform a [Homo sapiens]
g7662	NP_001186362.1	23.53	187	1.00E-06	DNA (cytosine-5)-methyltransferase 1 isoform 4 [Mus musculus]
g7662	NP_034196.5	23.53	187	1.00E-06	DNA (cytosine-5)-methyltransferase 1 isoform 2 [Mus musculus]
g7662	NP_001186361.1	23.53	187	1.00E-06	DNA (cytosine-5)-methyltransferase 1 isoform 3 [Mus musculus]
g7662	NP_001186360.2	23.53	187	1.00E-06	DNA (cytosine-5)-methyltransferase 1 isoform 1 [Mus musculus]
g7662	NP_001300940.1	23.53	187	1.00E-06	DNA (cytosine-5)-methyltransferase 1 isoform 5 [Mus musculus]

Appendix Table 4.5 cont.

g13819	NP_034196.5	38.98	59	6.00E-11	DNA (cytosine-5)-methyltransferase 1 isoform 2 [Mus musculus]
g13819	NP_001186361.1	38.98	59	6.00E-11	DNA (cytosine-5)-methyltransferase 1 isoform 3 [Mus musculus]
g13819	NP_001186360.2	38.98	59	6.00E-11	DNA (cytosine-5)-methyltransferase 1 isoform 1 [Mus musculus]
g13819	NP_001186362.1	38.98	59	6.00E-11	DNA (cytosine-5)-methyltransferase 1 isoform 4 [Mus musculus]
g13819	NP_001300940.1	38.98	59	6.00E-11	DNA (cytosine-5)-methyltransferase 1 isoform 5 [Mus musculus]
g13819	NP_445806.3	48.94	47	9.00E-11	DNA (cytosine-5)-methyltransferase 1 [Rattus norvegicus]
g13819	NP_001027526.1	51.16	43	1.00E-10	DNA (cytosine-5)-methyltransferase 1 [Sus scrofa]
g13819	NP_001370.1	51.16	43	1.00E-10	DNA (cytosine-5)-methyltransferase 1 isoform b [Homo sapiens]
g13819	NP_001124295.1	51.16	43	1.00E-10	DNA (cytosine-5)-methyltransferase 1 isoform a [Homo sapiens]
g13819	NP_001305659.1	51.16	43	1.00E-10	DNA (cytosine-5)-methyltransferase 1 isoform c [Homo sapiens]

205

Appendix Table 4.5 cont.

g13819	NP_996835.1	48.84	43	2.00E-10	DNA (cytosine-5)-methyltransferase 1 [Gallus gallus]
g13819	NP_001305660.1	51.16	43	2.00E-10	DNA (cytosine-5)-methyltransferase 1 isoform d [Homo sapiens]
g13819	NP_001036980.1	44.19	43	3.00E-10	DNA cytosine-5 methyltransferase [Bombyx mori]
g13819	NP_872592.2	48.84	43	7.00E-10	DNA (cytosine-5)-methyltransferase 1 [Bos taurus]
g13819	NP_001009473.1	48.84	43	5.00E-09	DNA (cytosine-5)-methyltransferase 1 [Ovis aries]
g18780	NP_001018312.2	39.44	142	3.00E-31	DNA (cytosine-5)-methyltransferase 3 beta, duplicate a [Danio rerio]
g18780	NP_571461.1	32.08	106	3.00E-17	DNA (cytosine-5)-methyltransferase 3 beta, duplicate b.2 [Danio rerio]

206

Appendix Table 4.6 NCBI Blastp results from mapping DNA N6-methyltransferase genes against *Oscheius tipulae* predicted genes. Methyltransferase genes were obtained from NCBI, and the *O. tipulae* predicted genes were obtained after Illumina HiSeq high throughput sequencing, annotation, and gene prediction.

Methyltransferase gene query accession	<i>O. tipulae</i> subject	Percent ID	Alignment length (aa)	E-value	Methyltransferase query description
XP_003373027.1	g10318	40	170	1.00E-31	n(6)-adenine-specific DNA methyltransferase 2-like protein [Trichinella spiralis]
XP_003373020.1	g10318	40	170	7.00E-31	N(6)-adenine-specific DNA methyltransferase 2, partial [Trichinella spiralis]
XP_003372519.1	g3229	36.36	132	3.00E-22	putative N(6)-adenine-specific DNA methyltransferase 1 [Trichinella spiralis]
NP_495127.1	g5321	32.31	260	5.00E-30	DNA N6-methyl methyltransferase [Caenorhabditis elegans]
XP_023954455.1	g5321	30.13	156	7.00E-12	DNA N6-methyl methyltransferase [Bicyclus anynana]
XP_013187493.1	g5321	25.66	265	3.00E-10	PREDICTED: DNA N6-methyl methyltransferase [Amyeloidis transitella]
XP_013163157.1	g5321	26.61	124	3.00E-08	PREDICTED: DNA N6-methyl methyltransferase [Papilio xuthus]

207

Appendix Table 4.6 cont.

AAS45233.1	g3229	50	136	9.00E-33	putative N6-DNA methyltransferase A transcript variant [Mus musculus]
------------	-------	----	-----	----------	---

Appendix Table 4.7 NCBI Blastp results from blasting the *Oscheius tipulae* predicted genes obtained as a hit in initial blast (Appendix Table 4.6) against DNA N6-methyltransferase genes. *O. tipulae* predicted genes were obtained after Illumina HiSeq high throughput sequencing, annotation, and gene prediction. Methyltransferase genes were obtained from NCBI.

<i>O. tipulae</i> gene query accession	Methyltransferase subject	Percent ID	Alignment length (aa)	E-value	Methyltransferase subject description
g3229	AAS45233.1	50	136	2.00E-30	putative N6-DNA methyltransferase A transcript variant [Mus musculus]
g3229	XP_003372519.1	36.36	132	5.00E-25	putative N(6)-adenine-specific DNA methyltransferase 1 [Trichinella spiralis]
g5321	NP_495127.1	32.31	260	1.00E-33	DNA N6-methyl methyltransferase [Caenorhabditis elegans]
g5321	XP_023954455.1	30.13	156	2.00E-15	DNA N6-methyl methyltransferase [Bicyclus anynana]
g5321	XP_013187493.1	25.66	265	6.00E-14	PREDICTED: DNA N6-methyl methyltransferase [Amyeloidis transitella]
g5321	XP_013163157.1	26.61	124	8.00E-12	PREDICTED: DNA N6-methyl methyltransferase [Papilio xuthus]
g10318	XP_003373027.1	40	170	4.00E-35	n(6)-adenine-specific DNA methyltransferase 2-like protein [Trichinella spiralis]

209

Appendix Table 4.7 cont.

g10318	XP_003373020.1	40	170	2.00E-34	N(6)-adenine-specific DNA methyltransferase 2, partial [Trichinella spiralis]
--------	----------------	----	-----	----------	---

Appendix Table 4.9 DAVID functional annotation results for all genes both upregulated under at least one abiotic stress and hypermethylated. The three stresses were heat, freezing, and desiccation stress, and hypermethylation was determined by presence of at least one hypermethylated peak summit obtained from MACS v1.4.2 falling within the gene body. Only the top 10 annotation clusters, based on highest enrichment score, are included.

Category	Term	Count	P value	Fold Enrichment	FDR
Annotation Cluster 1		Enrichment Score: 11.043919529175794			
INTERPRO	Nematode cuticle collagen, N-terminal	15	2.85E-13	16.2457	3.45E-10
GOTERM_CC_DIRECT	collagen trimer	14	4.44E-12	15.00113	4.58E-09
GOTERM_MF_DIRECT	structural constituent of cuticle	15	1.43E-11	11.92027	1.60E-08
INTERPRO	Collagen triple helix repeat	12	3.69E-10	14.91858	4.47E-07
Annotation Cluster 2		Enrichment Score: 3.143708367279805			
GOTERM_BP_DIRECT	nematode larval development	25	3.86E-04	2.054197	0.470958
GOTERM_BP_4	larval development	27	9.18E-04	1.852678	1.240274
GOTERM_BP_4	post-embryonic development	27	0.001045	1.83734	1.411153
Annotation Cluster 3		Enrichment Score: 3.0488355693844147			
GOTERM_BP_DIRECT	collagen and cuticulin-based cuticle development	6	5.81E-04	8.695251	0.70766
GOTERM_BP_4	collagen and cuticulin-based cuticle development	6	7.74E-04	8.148204	1.046753
GOTERM_BP_4	cuticle development	6	0.001587	6.941063	2.135323
Annotation Cluster 4		Enrichment Score: 2.088973782953777			
GOTERM_BP_4	reproductive system development	15	0.002885	2.383122	3.850259
GOTERM_BP_4	reproductive structure development	15	0.002885	2.383122	3.850259
GOTERM_BP_4	development of primary sexual characteristics	14	0.00704	2.249419	9.152902
GOTERM_BP_4	gonad development	14	0.00704	2.249419	9.152902
GOTERM_BP_4	animal organ development	16	0.010839	1.980018	13.76375

Appendix Table 4.9 cont.

GOTERM_BP_4	sex differentiation	14	0.013472	2.074416	16.83159
GOTERM_BP_4	system development	17	0.039556	1.651092	42.2122
Annotation Cluster 5		Enrichment Score: 1.7108673390422167			
GOTERM_CC_DIRECT	proteasome complex	3	0.015369	15.53689	14.76996
GOTERM_CC_4	proteasome complex	3	0.017382	14.53227	17.74521
KEGG_PATHWAY	Proteasome	3	0.027584	10.89474	20.86415
Annotation Cluster 6		Enrichment Score: 0.7675370184696046			
GOTERM_CC_DIRECT	cytosolic large ribosomal subunit	3	0.032865	10.35792	29.16498
GOTERM_CC_DIRECT	intracellular ribonucleoprotein complex	4	0.041702	5.108017	35.56655
GOTERM_CC_4	cytosolic part	4	0.045066	4.936653	40.17416
GOTERM_CC_4	large ribosomal subunit	3	0.056608	7.648565	47.75435
GOTERM_CC_4	ribosomal subunit	3	0.147534	4.337992	83.10745
GOTERM_CC_DIRECT	ribosome	3	0.202711	3.54005	90.34381
GOTERM_CC_4	ribosome	3	0.231555	3.229394	94.68327
KEGG_PATHWAY	Ribosome	3	0.25477	2.93617	91.45666
GOTERM_BP_4	ribosome biogenesis	3	0.33784	2.465904	99.63077
GOTERM_MF_DIRECT	structural constituent of ribosome	3	0.382382	2.237548	99.54293
GOTERM_BP_4	cellular amide metabolic process	5	0.386907	1.564869	99.87029
GOTERM_BP_4	peptide metabolic process	4	0.552361	1.372957	99.99819
GOTERM_CC_DIRECT	intracellular	3	0.617725	1.426857	99.99509
Annotation Cluster 7		Enrichment Score: 0.750149937549465			
GOTERM_BP_DIRECT	hermaphrodite genitalia development	8	0.154579	1.775182	87.16053
GOTERM_BP_4	hermaphrodite genitalia development	8	0.187594	1.679289	94.05671
GOTERM_BP_4	genitalia development	8	0.193723	1.663637	94.63791
Annotation Cluster 8		Enrichment Score: 0.5658519903537345			
GOTERM_BP_4	positive regulation of macromolecule biosynthetic process	4	0.186595	2.647015	93.95665
GOTERM_BP_4	positive regulation of biosynthetic process	4	0.194592	2.592098	94.71592

Appendix Table 4.9 cont.

GOTERM_BP_4	positive regulation of cellular biosynthetic process	4	0.194592	2.592098	94.71592
GOTERM_BP_4	positive regulation of nitrogen compound metabolic process	4	0.199433	2.560228	95.13147
GOTERM_BP_4	positive regulation of gene expression	4	0.237433	2.339684	97.48564
GOTERM_BP_4	positive regulation of cellular metabolic process	4	0.359882	1.848212	99.76691
GOTERM_BP_4	positive regulation of RNA metabolic process	3	0.394763	2.179171	99.89114
GOTERM_BP_4	positive regulation of macromolecule metabolic process	4	0.401296	1.725679	99.90607
GOTERM_BP_4	positive regulation of nucleobase-containing compound metabolic process	3	0.423526	2.05492	99.94382
Annotation Cluster 9	Enrichment Score: 0.4938009589001138				
GOTERM_BP_4	nucleoside metabolic process	3	0.244362	3.123478	97.77892
GOTERM_BP_4	glycosyl compound metabolic process	3	0.263112	2.965327	98.42133
GOTERM_BP_4	nucleobase-containing small molecule metabolic process	3	0.513361	1.728863	99.99438
Annotation Cluster 10	Enrichment Score: 0.47551383620081333				
GOTERM_BP_DIRECT	cell migration	3	0.171361	3.961949	89.95145
GOTERM_BP_4	cell motility	4	0.380654	1.784845	99.85111
GOTERM_BP_4	cell migration	3	0.574137	1.546276	99.99908