July 2016

Molecular Response of Spartina alterniflora to the Deepwater Horizon Oil Spill

Mariano Alvarez
University of South Florida, malvare2@mail.usf.edu

Follow this and additional works at: http://scholarcommons.usf.edu/etd
Part of the Evolution Commons, and the Genetics Commons

Scholar Commons Citation
Alvarez, Mariano, "Molecular Response of Spartina alterniflora to the Deepwater Horizon Oil Spill" (2016). Graduate Theses and Dissertations.
http://scholarcommons.usf.edu/etd/6166

This is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Molecular Response of *Spartina alterniflora* to the *Deepwater Horizon* Oil Spill

by

Mariano Alvarez

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biology with a concentration in Ecology and Evolution
Department of Integrative Biology
College of Arts and Sciences
University of South Florida

Major Professor: Christina Richards, Ph.D.
Gordon Fox, Ph.D.
Kathleen Scott, Ph.D.
Aaron Schrey, Ph.D.

Date of Approval:
July 07, 2016

Keywords: coastal ecosystems, gene expression, molecular ecology, transgenics

Copyright © 2016, Mariano Alvarez
Dedication

To everyone who has taken a chance on me along the way, and the few that have put up with me all along.
Acknowledgments

Thanks to my advisor, my committee, and all of my collaborators for their support and inspiration.

Travel funding support was provided by the Department of Integrative Biology travel grant program, the Summer Institute for Statistical Genetics, and NSF. Thank you to these funding sources.
Table of Contents

List of Tables iv

List of Figures v

Abstract vii

Chapter One: Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution? 1

Abstract 1

Introduction 2

Why wild systems? 4

Microarray and RNAseq technology 7

Studies reviewed: what have we learned? 9

How much variation is there in gene expression and how is it structured? 10

Variation in gene expression can be shaped by evolutionary processes 11

Macroevolutionary patterns of variation in gene expression: the comparative method 13

How do environmental stimuli affect gene expression? 15

Natural environmental fluctuations impact gene expression 15

Gene expression responds to extreme events or environmental stress 16

Practical implications of rapid response of gene expression to environment 17

A more nuanced understanding of expression: the use of time-course studies 18

Systems biology methods isolate the importance of specific environmental factors 19

How does gene expression affect phenotype? 20

Alternate phenotypes reveal expression differences that translate into phenotypic variation 22

Moving from correlation to causation 24

Transcriptomics in the future: where do we go from here? 25

A unifying workflow 26

Future directions for inquiry 28

Acknowledgments 32

References 32

Figures and Tables 47
Chapter Two: Genome-wide Differential Gene Expression in the Foundation Salt Marsh Grass *Spartina alterniflora* Caused by the *Deepwater Horizon* Oil Spill

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>52</td>
</tr>
<tr>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>Methods</td>
<td>57</td>
</tr>
<tr>
<td>Study species and population</td>
<td>57</td>
</tr>
<tr>
<td>RNA extraction and microarray hybridization</td>
<td>59</td>
</tr>
<tr>
<td>Data analysis</td>
<td>59</td>
</tr>
<tr>
<td>Gene interaction networks</td>
<td>60</td>
</tr>
<tr>
<td>Target genes and technical conformation</td>
<td>61</td>
</tr>
<tr>
<td>Results</td>
<td>62</td>
</tr>
<tr>
<td>Oil-contaminated populations were differentiated from uncontaminated controls</td>
<td>62</td>
</tr>
<tr>
<td>Gene Ontology and network analysis identified genes and categories of interest</td>
<td>63</td>
</tr>
<tr>
<td>Discussion</td>
<td>64</td>
</tr>
<tr>
<td>References</td>
<td>69</td>
</tr>
<tr>
<td>Figures and Tables</td>
<td>77</td>
</tr>
</tbody>
</table>

Chapter Three: Oil responsive genes in *S. alterniflora* modulate phenotype in *Brachypodium distachyon* with applications in *Spartina alterniflora*

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>86</td>
</tr>
<tr>
<td>Introduction</td>
<td>87</td>
</tr>
<tr>
<td>Methods</td>
<td>90</td>
</tr>
<tr>
<td>Generating gene targets</td>
<td>90</td>
</tr>
<tr>
<td>Gene targets</td>
<td>91</td>
</tr>
<tr>
<td>Oil exposure experiment</td>
<td>93</td>
</tr>
<tr>
<td>Analysis</td>
<td>94</td>
</tr>
<tr>
<td>Results</td>
<td>95</td>
</tr>
<tr>
<td>T-DNA insertion genotypes differ from wild type and respond to oil stress, but not alter phenotypic response to oil stress</td>
<td>95</td>
</tr>
<tr>
<td>Discussion</td>
<td>95</td>
</tr>
<tr>
<td>References</td>
<td>100</td>
</tr>
<tr>
<td>Figures and Tables</td>
<td>109</td>
</tr>
</tbody>
</table>

Chapter Four: Oil exposure induced effects in *S. alterniflora*, but not in a genotype specific manner

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>119</td>
</tr>
<tr>
<td>Introduction</td>
<td>120</td>
</tr>
<tr>
<td>Methods</td>
<td>123</td>
</tr>
<tr>
<td>Results</td>
<td>124</td>
</tr>
<tr>
<td>Discussion</td>
<td>124</td>
</tr>
<tr>
<td>References</td>
<td>126</td>
</tr>
<tr>
<td>Tables and Figures</td>
<td>130</td>
</tr>
</tbody>
</table>
Appendix 1: Primary Problems in Transcriptome Studies 135
  Bias in signals 135
  Heterologous arrays 136
  Polyploidy 136
  RNA pooling 137
  Statistical analysis 138
  Unannotated genes 139

Appendix 2: Copyright Permissions 141
**List of Tables**

| Table 1.1: DNA microarrays provided commercially by Agilent and Affymetrix, and custom built. | 49 |
| Table 2.1: Highly oil-responsive target genes and their respective homologs. | 81 |
| Table 3.1: T-DNA insertion lines are shown with their corresponding *S. alterniflora* contigs from a previously constructed microarray. | 109 |
| Table 3.2: Linear model for biomass. | 111 |
| Table 3.3: Linear model for inflorescence weight. | 112 |
| Table 3.4: Linear model for percent inflorescence. | 113 |
| Table 3.5: Generalized linear model for seeds, using poisson distribution. | 114 |
| Table 4.1: ANOVA results of all phenotypes. | 130 |
List of Figures

Figure 1.1: Number of studies each year addressing each of the question categories. 47

Figure 1.2: Venn diagram of question categories in studies examined. 48

Figure 2.1: Map of Gulf of Mexico study sites. 77

Figure 2.2: Principal variance components analysis. 78

Figure 2.3: Results of ANOVA, using oil, population (which is nested within oil), and state. 79

Figure 2.4: Interaction network of all genes highly responsive to oil (P<0.001), visualized in Cytoscape. 80

Figure 2.5: Average fold change in expression relative to control (expressed as 1) in qPCR target genes. 84

Figure 2.6: Normalized change in expression in microarray due to oil stress in target genes. 85

Figure 3.1: Means of percent inflorescence for each T-DNA knockout line in oil treatment and control. 115

Figure 3.2: Means of seed production for each T-DNA knockout line in oil treatment and control. 116

Figure 3.3: Means of inflorescence weight for each T-DNA knockout line in oil treatment and control. 117

Figure 3.4: Means of total biomass for each T-DNA knockout line in oil treatment and control. 118

Figure 4.1: Changes in the number of leaves over the course of the experiment. 131

Figure 4.2: Changes in the number of ramets over the course of the experiment. 132
Figure 4.3: Changes in the number of leaves per ramet over the course of the experiment. 133

Figure 4.4: Mortality over the course of the experiment. 134
Abstract

Although the “genome as a blueprint” metaphor has been pervasive in biology, recent advances in molecular biology have revealed a complex network of regulatory machinery that dynamically regulated molecular processes in response to environmental conditions. However, these patterns, as well as the evolutionary processes that underlie them, remain understudied in natural conditions. In 2010, the Deepwater Horizon oil spill released an estimated 4.9 million barrels of oil into the Gulf of Mexico, making landfall on salt marsh habitat dominated by the foundation species Spartina alterniflora. Despite the severe impacts to phenotype and fitness, S. alterniflora proved remarkably resilient in the face of the crude oil stress. Despite the tragedy of the Deepwater Horizon oil spill, the event represented a unique opportunity to explore the molecular mechanisms of oil tolerance in this highly resilient species. To understand how organisms regulate gene expression in natural settings and to identify best practices in genome-wide expression studies, we first surveyed a ten-year span of transcriptome-wide gene expression studies. We then confirmed the hypothesis that crude oil exposure would induce differential gene expression in affected populations, and whole-transcriptome microarray identified 3,622 genes that responded exclusively to oil stress. To confirm the function of candidate genes involved in resilience to oil stress, we used a highly-differentially expressed subset of these genes to construct gene interaction networks and identify target genes. We obtained T-DNA insertion
genotypes of the emerging model grass species *Brachypodium distachyon* that were disrupted in these target genes for functional confirmation, but were unable to detect significant modulation of oil response through these heterologous knockouts. Finally, we isolated the phenotypic effects of crude oil exposure through greenhouse trials and found evidence that crude oil may have acted as a selective pressure, rather than an inducer of plasticity. Together, these studies identify novel patterns of gene expression in response to a severe but unpredictable stressor that has widespread impacts on a foundational salt marsh grass species. In addition, this dissertation represents a pathway to understanding functional genomics in non-model systems without extensive genomic resources.
Chapter One:
	Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution?¹

Abstract

Molecular ecology has moved beyond the use of a relatively small number of markers, often non-coding, and it is now possible to use whole genome measures of gene expression with microarrays and RNAseq (i.e. transcriptomics) to capture molecular response to environmental challenges. While transcriptome studies are shedding light on the mechanistic basis of traits as complex as personality or physiological response to catastrophic events, these approaches are still challenging because of the required technical expertise, difficulties with analysis, and cost. Still, we found that in the last ten years, 575 studies used microarrays or RNAseq in ecology. These studies broadly address three questions that reflect the progression of the field: (i) How much variation in gene expression is there and how is it structured? (ii) How do environmental stimuli affect gene expression? (iii) How does gene expression affect phenotype? We discuss technical aspects of RNAseq and microarray technology, and a framework that leverages the advantages of both. Further, we highlight future directions

¹ This chapter has been previously published in Molecular Energy (Alvarez, M., Richards, C., Schrey, A. “Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution?” Molecular Ecology. 24.4 (2015): 710-725.) and is reproduced with permission of the publisher.
of research, particularly related to moving beyond correlation and the development of additional annotation resources. Measuring gene expression across an array of taxa in ecological settings promises to enrich our understanding of ecology and genome function.

**Introduction**

The interactions between organisms and environments are of central importance to many questions in the study of ecology (Scheiner & Willig 2011). While much progress has been made by examining traits and behaviors of individuals within and among populations, the integration of molecular techniques into ecology has allowed investigators unprecedented ability to assess the mechanisms that govern ecological interactions and underlie pattern and process at the most basic levels of biological organization. Now, genome wide approaches can lay the foundation for sophisticated functional studies that explore the specific genomic basis of phenotypic variation and rapid response to environmental change (e.g. mass flowering in a tropical tree Kobayashi *et al.* 2013; plasticity underlying response to osmotic conditions in killifish, Whitehead *et al.* 2013). As a result, the past decade has seen the rise of molecular ecology as a synthetic discipline that uses molecular techniques to answer (and often generate) ecological questions (Andrew *et al.* 2013). DNA microarrays and RNA sequencing (RNAseq) are the most widespread and powerful transcriptomics technologies, and allow ecologists to simultaneously measure genome-wide gene expression on large numbers of individuals in wild populations. These tools measure variation in gene expression at the level of mRNA, which ultimately contributes to the
formation of proteins, cellular phenotype, and organismal phenotype that can be shaped by ecological processes (Oleksiak et al. 2002; Crawford & Oleksiak 2007; Aubin-Horth & Renn 2009; Ekblom & Galindo 2011; Whitehead 2012). The quantification of these patterns on a genome-wide scale allows us to observe the molecular regulation of phenotype in response to ecological phenomena, and begin to understand the ecological transcriptome (Richards et al. 2009).

Application of transcriptomics in an ecological context has become imperative because as the research community continues to acquire abundant genomics data for a variety of organisms in controlled lab settings, we have made little progress in understanding how the genome actually functions to create complex traits and adapt to complex environments (Richards et al. 2009). The nascent field of ecological genomics has already begun to shed light on how genomes function in natural environments including the mechanisms underlying adaptation (e.g. Lai et al. 2006; Elmer & Meyer 2011; Andrew et al. 2013), divergence (e.g. Pavey et al. 2010; Renaut & Bernatchez 2011; Nosil & Feder 2012), genotype-by-environment interactions (e.g. Richards et al. 2012), and phenotypic plasticity (e.g. Wittkopp 2007; Aubin-Horth & Renn 2009; Nicotra et al. 2010; Whitehead et al. 2013; Schneider et al. 2014). These studies capitalize on the statistical strength of ecological experimental design to capture sources of variation while leveraging powerful genomics tools to assess gene activity. The work in Helianthus by Rieseberg and colleagues serves as an illustrative example. With classic ecological design in several greenhouse and field studies, they documented that some adaptive traits in wild populations of the hybrid sunflower species Helianthus deserticola are much greater or much smaller (i.e. transgressive) compared to the parental species.
*H. annuus* and *H. petiolaris* (Rosenthal et al. 2002; Gross et al. 2004). Lai et al. (2006) used microarray data from all three species to suggest that novel gene expression in the hybrid *H. deserticola* may contribute to the transgressive phenotypic patterns. In particular, they identified a number of highly differentially transcribed transporter genes and speculated that the differential expression of these genes was correlated to adaptation in *H. deserticola* to an extreme, arid environment. This was supported by an association between fitness and expression of one of the genes of interest (G protein-coupled receptor: QHB30N12) in *H. deserticola* in the field. The combination of molecular techniques with a traditional ecological design allowed Lai et al. (2006) to identify possible mechanisms of adaptation that resulted from hybridization. This union of ecology and molecular biology is the hallmark of molecular ecology.

**Why wild systems?**

By wild systems we typically mean non-model organisms or non-traditional model organisms (e.g. *Daphnia pulex, Coregonus clupeaformis, Fundulus heteroclitus*) in natural settings, although traditional model organisms like *Arabidopsis thaliana, Drosophila melanogaster* and *Caenorhabditis elegans* do have their own ecology and can be extremely informative when studied in an ecological setting (Kammenga 2007; Pavey et al. 2012; Weigel 2012). Since organisms in natural settings are continuously exposed to multiple environmental signals and must respond appropriately to dynamic conditions, this context provides a unique opportunity to discover information about gene expression patterns that cannot be gleaned through controlled laboratory settings. Transcriptome studies in natural settings have found novel expression of otherwise-
silent genes that only respond to the multiple, simultaneous stimuli that occur in complex, dynamic environments (Colbourne et al. 2011; Pavey et al. 2012). Novel behavior of transcripts may also be exposed by particular environmental challenges, which may contribute to variation between individuals and populations (Dalziel et al. 2009). For example, Whitehead and colleagues (2012) examined pollution-tolerant and naïve populations of Fundulus heteroclitus along the northern Atlantic coast of the United States. Divergence in expression of genes that were responsive to the toxin was revealed only at higher dosages of PCB-126. On the other hand, the authors found that neutral processes explained patterns of population divergence in expression of genes that were not responsive to dioxin-like compound PCB-126. These results suggest that environmental challenge may be necessary to expose adaptive population divergence: without the stimulus of high PCB concentrations, the population differences between F. heteroclitus were obscured (Whitehead et al. 2012). Further, the complex interactions of environmental factors in natural systems may reveal more differentiation between populations than would be observable under controlled conditions.

Besides identifying context dependency of transcription, transcriptome studies in non-model organisms may yield functional information about novel transcripts that either have no homolog in their most closely-related model organism, or have taken on a novel function. In Daphnia pulex, researchers found that unannotated, Daphnia-specific genes made up more than a third of the transcriptome and were the most responsive to a variety of ecological stimuli (Colbourne et al. 2011). The study further revealed the importance of the diversification of duplicated genes within specific metabolic pathways in this species. In many cases, gene duplication allowed for immediate divergence in
expression patterns, which may be particularly effective when the duplicated gene interacts with genes sharing a common regulatory program. This novel behavior or function of duplicated genes was only exposed under specific environmental challenges, suggesting that regardless of the number of laboratory studies on an organism, a large number of gene functions cannot be annotated without exposure to complex natural stimuli.

By revealing context-specific gene expression variation, novel transcripts or novel function of known transcripts, transcriptome experiments in wild settings may be the only way to infer the function of many genes present in an organism (Colbourne et al. 2011; Pavey et al. 2012). This association of gene expression with natural environmental conditions provides an “ecological annotation” that promises to build upon existing biological process, cellular components and molecular function annotations and could be the best option for annotating genes that govern important traits (Landry & Aubin-Horth 2007; sensu the “stress annotation” in Richards et al. 2012).

To assess the impact of transcriptomics in wild systems over the past decade, we summarize the primary questions that have been addressed with transcriptomic approaches in ecology. In particular, we concentrate on studies using the two main whole-genome quantification techniques: DNA microarray and RNAseq. We briefly review the literature, evaluate the experimental evidence, and identify some promising questions for future research. As the availability of next-generation sequencing technology increases, ecological transcriptomics is increasingly performed using RNAseq. However, DNA microarrays represent the dominant method of transcriptome
quantification over the past decade, and continue to offer robust data that may still be appropriate in some systems. Therefore, clarification of the differences and appropriate applications of each type of technology is needed before we can explore the use of transcriptomics in ecology.

**Microarray and RNAseq technology**

Over the past ten years, transcriptomic workflows have become increasingly refined. Microarray experiments have been the subject of multiple reviews (Allison et al. 2006; Kammenga et al. 2007; Crawford & Oleksiak 2007), and there is currently consensus on most aspects of experimental design. RNAseq is maturing, and the application to ecological studies has been discussed, but there is no current consensus on cDNA library preparation methods and data processing (Ekblom & Galindo 2011; Vijay et al. 2013; Wolf 2013). Previous authors explore the main differences between microarrays and RNAseq in expense, statistical analysis, bias in signals and the specific problems of using heterologous arrays and RNA pooling, which we summarize in box 1.

DNA microarrays have been a convenient and popular tool for use in ecology, particularly because of the ease of analysis (Allison et al. 2006; Richards et al. 2009; see box 1: Statistical analysis). Briefly, a DNA microarray consists of thousands of probes, representing sections of DNA code to be quantified, that are affixed to a surface. Level of expression for each gene is estimated from single probes or averages of multiple probes that are designed to target segment(s) of a gene (Aryoles & Gibson 2006). Complementary DNA (cDNA) is reverse transcribed from mRNA transcripts that are extracted from experimental material. The cDNA is labeled with a fluorescent dye
before being washed over probes designed to hybridize with a specific DNA sequence. The array is scanned with lasers; relative light intensity of the fluorescent dye at a probe is proportional to the number of cDNA transcripts that are hybridizing at that particular probe (Allison et al. 2006; although see box 1 Bias in signals). An image of the illuminated probes serves as the raw microarray data, which is preprocessed, normalized, and analyzed. The raw intensity data are typically first converted to a logarithmic scale (base 2) because the distribution is highly skewed: most transcripts show low expression and a minority have high expression (Aryoles & Gibson 2006).

RNAseq is a newer, increasingly popular technique for genome-wide ecological transcriptomics. RNAseq uses next-generation sequencing methods to characterize RNA transcripts by using high-throughput sequencing of a cDNA library to generate hundreds of thousands of fragments of DNA sequences. In the RNAseq study design phase, a user must select a next-generation sequencing platform: each platform differs in read length, sequencing depth and quality, and the impact of highly differentially expressed genes on the detection of other differentially expressed genes (Wolf et al. 2010, Wolf 2013; see box 1: Bias in signal). Platform characteristics, coverage, costs, and even available platforms, are rapidly and continually evolving, and users will need to obtain current, up-to-date information from manufacturers. Initial raw data processing requires considerable time, computing power and bioinformatics expertise, and importantly for non-model systems the requirements increase when assembling transcripts de novo (Wolf 2013). Raw RNAseq data are parsed with scripting languages due to the size of the resulting files (Malone & Oliver 2011) and then aligned to a reference genome or transcriptome via software such as the Tuxedo Suite Tools
(Trapnell et al. 2012), or assembled de novo via software such as Trinity (Grabherr et al. 2011).

**Studies reviewed: what have we learned?**

Over the past decade, both RNAseq and microarrays have made critical contributions to ecology. We used a systematic review approach (Doerr et al. 2014) to characterize the development of this field in an objective, repeatable fashion. We found 575 studies published between 2004 and 2013 through the Web of Science database (see Appendix S1). Each study shared the wild card search term “transcriptom*” which initially returned 307,000 studies. We further refined the search by choosing studies only from the Web of Science research areas “evolutionary biology” and “environmental science ecology” which returned 2303 empirical studies. From these studies, we concentrated on ecological and evolutionary studies and manually excluded studies that primarily referred to toxicology, agriculture, or other applications in environmental science without an obvious ecological context. We also excluded meta-analyses and studies that performed analyses on previously-generated data. Rather than classifying studies by the ecological phenomenon or study organism, we identified a more general summation based on three questions that reflects how ecological transcriptomics has made a transition over the past ten years from largely descriptive investigations to those that are more functional and mechanistic (Figure 1.1): (i) How much variation in gene expression is there in natural populations, and how is it structured? (ii) How do environmental stimuli affect gene expression? (iii) How does variation in gene expression translate into phenotype? In the following, we elaborate on how each of
these questions has been addressed. Note that some studies addressed more than one of these questions, and were classified in multiple categories (Figure 1.2; Appendix S1). We also identified which technology (55% of the total were DNA microarray, 45% were RNAseq) each study used, and whether organisms in each study were reared in or acclimated to laboratory conditions before sampling, or were sampled from a field setting.

**How much variation is there in gene expression, and how is it structured?**

One of the most important questions for understanding the importance of any trait in ecology and evolution is how much variation exists in natural populations, and how is it structured. Gene expression is highly variable, and transcription levels vary within multiple biological scales: within and among individuals, populations, and species (Whitehead & Crawford 2006a, Whitehead & Crawford 2006b, Crawford & Oleksiak 2007). Within individuals, gene expression varies between tissues and even among cell types within the same tissue (Birnbaum et al. 2003, Whitehead & Crawford 2006b). Gene expression further varies temporally, fluctuating with developmental time, day-night cycles, and life history events (Aubin-Horth & Renn, 2009, Francesconi & Lehner 2013). A majority of studies that we reviewed (66%) quantified variation in gene expression in one or more natural populations, even if this was not the main focus of the study. Studies developing transcriptome resources for the first time appear in this category, as they often represent a “first look” at a non-model organism’s transcriptome. This category is the most descriptive in nature, and several of the studies in this category represent some of the earliest and most fundamental questions in ecological
transcriptomics: How does gene expression vary within individuals, within and among populations, and among species? Understanding existing levels of variation in gene expression is an important question because changes in gene expression may allow organisms to respond to novel stressors and variation in gene expression may translate into phenotypic variation (Oleksiak et al. 2002; Whitehead & Crawford 2006a; Whitehead 2012).

Variation in gene expression can be shaped by evolutionary processes

Variation in gene expression is potentially heritable and may be acted upon by natural selection. As with any other trait, the structure of gene expression variation within and among natural populations may reflect both adaptive and non-adaptive processes (Oleksiak et al. 2002; Whitehead 2012). Expression variation may be facilitated by regulatory elements or epigenetic mechanisms that alter gene expression even before genetic variants arise in the population (West-Eberhard 2003). Therefore, population-level differences in expression may reflect the early processes that underlie adaptive divergence (e.g. in Oleksiak et al. 2002; Derome et al. 2006; Jeukens et al. 2010). In order to quantify the expression variation that is correlated with the early stages of divergence within populations, Derome and colleagues (2006) used microarrays to investigate transcriptional differences between differentiated “normal” and “dwarf” types of the fish Coregonus clupeaformis. Previous work demonstrated that the “normal” and “dwarf” types had physiological variation in swimming activity which might be partly due to expression of genes related to energy metabolism (Bernatchez & Dodson 1985). The authors sampled individuals of both types from two populations and compared
transcription in muscle tissue. They found 51 differentially expressed genes between the two types, which as hypothesized, were primarily annotated as being involved in energetic metabolism. The authors were able to detect expression variation that may reflect adaptive divergence between two sympatric sub-populations of *C. clupeaformis*, and identify candidate genes for future analyses.

To specifically examine how adaptive or neutral processes have contributed to variation in gene expression in natural populations, several studies have adopted classic approaches like McDonald-Kreitman tests, $Q_{st}$-$F_{st}$ tests, and Quantitative Trait Loci (QTL) mapping. In particular, the salmonid fishes have been a useful system to tease apart the importance of selection and drift on gene expression in natural populations. In one of the first applications of $Q_{st}$-$F_{st}$ analysis to transcriptomics, Roberge et al. (2007) used both a genome scan and $Q_{st}$-$F_{st}$ analysis to identify genes whose transcriptional profiles (assessed via microarray) had been shaped by selection in two diverging subpopulations of salmon (*Salmo salar*). $Q_{st}$ usually quantifies the amount of variation in quantitative traits in populations (Spitze 1993), which can be compared to variation at neutral loci ($F_{st}$) to identify selection and drift in phenotypic divergence (Koskinen et al. 2002, Roberge et al. 2007). The authors adapted the $Q_{st}$ framework to gene expression by treating gene expression as a quantitative trait and estimated transcription level $Q_{st}$ values for 1044 genes with transcriptional profiles that were highly heritable. After narrowing their search to genes that were divergent between the two populations and testing for a neutral model of genetic variation, the authors identified only 16 genes that likely diverged between the subpopulations due to selection, rather than neutral processes. Conversely, they found 11 divergent genes
whose expression did not reflect directional selection, but may be influenced by neutral processes, suggesting a role for drift in shaping the expression patterns of *S. salar* (Roberge *et al.* 2007). This combination of natural populations, captive breeding, and genomic techniques is a powerful model for investigators attempting to disentangle the role of selection and neutral processes in natural populations.

An expression quantitative trait loci (eQTL) mapping approach is another useful tool for identifying important expression variation within populations (Brem *et al.* 2002; Wittkopp 2007). eQTL analyses treat gene expression as a quantitative trait, and use classic QTL methods to map genetic loci that underlie variation in gene expression. Further, eQTL mapping can identify whether genes are modified by *cis-* or *trans-*regulation (Brem *et al.* 2002; Wittkopp 2007; Derome *et al.* 2008, Whiteley *et al.* 2008). Although eQTL is usually performed in model organisms, Derome and colleagues (2008) were able to make use of a linkage map generated for *C. clupeaformis* to identify 34 transcripts that may play a role in the ongoing divergence between the two types, and thus may be under selection (Derome *et al.* 2008). The power of eQTL and *Qst-Fst* analyses can be leveraged in natural environments to identify the effects of evolutionary forces on transcripts of interest that would otherwise be unidentifiable under laboratory conditions.

**Macroevolutionary patterns of variation in gene expression: the comparative method**

Because neutral and adaptive processes contribute to variation in gene expression, examining the contributions of these processes to differential expression patterns in a phylogenetic comparative framework may contribute to our understanding
of divergence between populations or species (Whittkopp 2007; Whitehead 2012; Schraiber et al. 2013). Comparisons between closely-related species can identify genes (e.g. Chelaifa et al. 2010; Lai et al. 2006), or changes in regulatory pathways (e.g. Brem et al. 2002; Schraiber et al. 2013) that may have been important in speciation events and test hypotheses about the importance of response to environmental challenges to speciation (Whittkopp 2007). For example, Chelaifa et al. (2010) used microarrays to explore the transcriptomes of *Spartina maritima* and *Spartina alterniflora*, two recently diverged sister species that occupy overlapping niches. The authors found differential expression in 13% of the transcriptome, including genes annotated as transporter genes, developmental genes, and cellular growth genes. These divergent expression patterns may play a role in allowing the species to inhabit their different ecological niches and in shaping differentiated phenotypes (Chelaifa et al. 2010).

Although this category is the most descriptive of the three categories examined, studies in this category have identified gene expression variation across time, space, and phylogenetic distance in complex natural environments. Additionally, advanced approaches, such as $Q_{st}$-$F_{st}$, have been able to quantify the influence of selection, drift, or bottleneck events on the evolution of gene expression. Further application of these approaches to new or existing datasets may shed light on the relative influence of these factors in diverging or recently diverged populations. Additionally, the use of captive breeding populations may aid in the development of resources, such as genetic maps (as in Derome et al. 2007). These resources can then be combined with sampling in natural systems for more powerful discovery of the forces that shape diversity in gene expression.
How do environmental stimuli affect gene expression?

A central theme in molecular ecology explores the mechanisms by which organisms respond to their environment, a question which has taken on increasing importance in the shadow of global climate change (Nicotra et al. 2010; Meier et al. 2014, Palumbi et al. 2014). This question is the logical next step in ecological studies after describing patterns of variation, and just under half of the studies we reviewed (41%) addressed how environmental stimuli affect gene expression. Our definition of environmental stimulus includes, but is not limited to, abiotic stress (such as temperature or pollution; e.g. Chapman et al. 2011; Whitehead et al. 2012; Palumbi et al. 2014), environmental heterogeneity in time or space (e.g. Richards et al. 2012; Meier et al. 2014), host-parasite interactions (e.g. Webster et al. 2011), and other, potentially selective biotic and abiotic interactions.

Natural environmental fluctuations impact gene expression

Ten years of transcriptomics in the wild have described the significant impacts that stress response can have on many categories of genes, but have also shown that transcription may be affected by even small changes in environment (Krasnov et al. 2005; Richards et al. 2012). For example, Richards et al. (2012) used DNA microarrays to explore how gene expression changes through development and in response to environmental conditions. The authors found that variation in gene expression in two accessions of *A. thaliana* grown in field conditions, was equally explained by differences in accession and developmental status, and that temperature and precipitation
significantly predicted expression. The authors also created a “stress annotation” of the
*A. thaliana* genome based on published microarray studies to show that genes
previously annotated as stress-related, were expressed during the life cycle of the
organisms even under normal field conditions (Richards *et al.* 2012). Using a simple
design and several environmental measurements, this study teased apart the influence
of development and abiotic environmental variables in a complex natural setting. In
addition to identifying the molecular level basis of response to environmental
challenges, this type of study elucidates genotype-by-environment interactions at the
molecular level (Wray 2013) and reveals the molecular mechanisms of phenotypic
plasticity (e.g. Chapman *et al.* 2011; Gunter *et al.* 2013; Whitehead *et al.* 2013).

**Gene expression responds to extreme events or environmental stress**

Transcriptomic data can also provide insight into mechanisms of organismal
response to specific pollution events, stresses, or climatic conditions in wild populations.
For example, Whitehead and colleagues (2013) recently used a combination of RNaseq
and microarrays to examine the molecular and physiological response to osmotic
challenge in two species of killifish, *Fundulus heteroclitus* and *F. majalis*, which vary in
tolerance to changes in salinity. They reported rapid divergence in gene expression
between the two species in response to osmotic stress. The authors also found a
greater capacity for the morphological remodeling of gills in *F. heteroclitus*, and
suggested that the expression and correlated morphological variation they observed
could have played a role in the divergence of *F. heteroclitus* from *F. majalis* allowing it to
inhabit a broader range of osmotic environments (Whitehead *et al.* 2013). In another
study, Chapman et al. (2011) found that environmental conditions (pH and temperature) were the primary drivers of differential transcription in populations of the eastern oyster (*Crassostrea virginica*), as opposed to the minimal effects attributed to pollutants (metals and organic contaminants). The authors were able to develop statistical models of predictive value that were parameterized by data from natural populations, rather than simulated data or data obtained from controlled studies.

*Practical implications of rapid response of gene expression to environment*

The responsiveness of gene expression has the practical implication that sampling requires consistent handling with minimal exposure to unwanted stimuli before and during collection. When possible, samples should be collected at approximately the same time and flash frozen immediately upon collection. RNA preservation additives, such as RNAlater (Qiagen), may also prevent RNA degradation between sampling times. To minimize batch effects, each sample must be handled consistently to prevent differential environmental stimuli from having effects on the samples. As an alternative to immediate freezing of samples, some experiments have attempted to minimize intra-individual variations by letting organisms sit in a uniform space for a period of time before samples are sacrificed and frozen. This may be necessary when logistics prevent immediate preservation, but may introduce unwanted variation into the sample populations by introducing additional handling, feeding, and other environmental variables that may affect gene expression.

Another technical difficulty is that transcriptome assays represent only a “snapshot” of gene expression at a particular moment: temporal variation is necessarily
controlled for by careful sampling during the same time of day, climate cycle, or time during a stimulus response in order to get relevant comparisons between groups. In all of the studies addressing environmental response, ecological experimental design and analysis allowed investigators to discriminate between the effects of multiple, complex environmental inputs in wild populations. However, most studies used data obtained from a single time point to describe organismal response to stimulus. Unfortunately, this methodology condenses the temporo-spatial variation in a transcriptional response into a single time point. Future transcriptome studies should explore temporal levels of variation, as temporal differences in gene expression may help to pinpoint the primary regulatory loci that allow organisms to modulate gene networks and subsequently phenotype in response to endogenous or exogenous stimuli (West-Eberhard 2003; Aubin-Horth & Renn 2009; Dalziel et al. 2009, Johnson et al. 2012). Gene expression response may involve variation in regulatory loci early in the response, followed by a generalized downstream response in other genes (Wittkopp 2007; de Nadal et al. 2011). Alternatively, gene networks may respond sequentially to a complex environmental stimulus as various conditions are met (Aubin-Horth & Renn 2009).

A more nuanced understanding of expression: the use of time course studies

Aubin-Horth & Renn (2009) suggest using a time course approach to understand temporal patterns in transcription. Time course designs allow for a description of the entirety of a transcriptional response, and assist in ecological annotation by teasing apart general response genes from major regulatory genes. The replication normally associated with ecological studies, combined with the replication required for surveying
multiple time points, may quickly become overwhelming in both cost and scale. However, measuring behavior of previously identified candidate genes with qPCR (rather than using whole-genome assays) may provide useful data on temporal response without incurring prohibitive costs. A recent study used this approach to explore the role of candidate genes in the environmentally responsive network that underlies the diet-induced plasticity of the lower jaw in cichlids (Schneider et al. 2014). The authors identified a pattern across development in response to diet: first mechanically responsive genes, then osteoblast differentiation genes, then matrix related genes were differentially expressed. In addition, the time course design allowed for identifying so called immediate early genes (e.g. AP1) that influenced expression at different levels of the regulatory cascade.

Systems biology methods isolate the importance of specific environmental factors

As ecological transcriptomics identifies the stimuli that affect patterns of transcription, the effects of climatic and meteorological fluctuations on loci of interest should become an important focus of study. Climatic variation can have large impacts on gene expression (e.g. Richards et al. 2012) and is a critical component in surveying organismal response to climate change. When response to these variables is not the primary interest, the incorporation of climatic data will allow investigators to control for the impact of climate across taxa and habitats. When these variables are of interest, a clear understanding of climate variables will allow researchers to make predictions about which factors (temperature, rainfall, CO₂) are driving variation in gene expression. Nagano and colleagues (2012) provide an example in rice (Oryza sativa) where the
authors used microarrays to model the expression of each gene in response to climate variables at time points across development. In soybean \((Glycine\ max)\), Leakey \textit{et al.} (2009) used microarray data and physiological measurements to determine the mechanism of respiratory regulation in a free-air CO\(_2\) enrichment (FACE) experiment. The authors found variation in metabolic gene expression during periods of elevated CO\(_2\) and a concurrent increase in nighttime respiration.

Although these experiments used agricultural species, their approaches are useful to determine the impact of climate on wild populations. Aikawa \textit{et al.} (2010) used a similar design to model the expression of a single gene in the flowering time network over two years in \textit{Arabidopsis halleri}. The authors found that at any time point, expression follows cues from the prior six weeks of temperature data. Higher-throughput expression analysis may reveal other patterns of response to specific environmental factors within the flowering time network. These studies often use systems biology approaches of unsupervised classifications such as principal components analysis and clustering (e.g. Richards \textit{et al.} 2012), and regression analysis for each gene (e.g. Nagano \textit{et al.} 2012), to identify specific environmental factors that impact gene expression in an ecological context (Richards \textit{et al.} 2009; Shimizu \textit{et al.} 2011).

**How does gene expression affect phenotype?**

For gene expression to play a functional role in ecology, it must affect phenotype. Characterizing the relationship between gene expression and phenotype provides critical insight into understanding how ecological and evolutionary processes, such as
adaptive divergence, take place at the molecular level and exert influence on phenotype. Despite this imperative, a minority of the studies we reviewed made the connection between gene expression and phenotype (15%); this was the most underrepresented research question, which has increased only in recent years (Figure 1.1). The relationship between gene expression and phenotype is complex, as gene expression may have interactive effects with other larger scale systems, such as the proteome and metabolome, and may not immediately impact fitness (Crawford & Oleksiak 2007; Dalziel et al. 2009, Rees et al. 2011). Most studies in this category relied on patterns of correlation between expression of functional genetic elements and the production of a particular phenotype (including proteins, metabolites or traits). In one example, Aubin-Horth et al. (2012) linked differential expression in genes that are involved in stress response to behavioral variation in Gasterosteus aculeatus. The authors were able to demonstrate a correlation between expression levels in candidate genes with variation in boldness and aggressiveness (Aubin-Horth et al. 2012). Few studies have confirmed the causal relationship between functional elements and phenotype through additional protein- or metabolism-based assays (Rees et al. 2011; Whitehead et al. 2011), by knocking out genes of interest (Dowen et al. 2012; Richards et al. 2012), or through transgenic expression of genes of interest (Kobayashi et al. 2013).
Alternate phenotypes reveal expression differences that translate into phenotypic variation

In organisms with alternate phenotypes, analysis of differential gene expression can help to explain the processes by which the phenotypes diverge and resources are allocated to create the alternative types (Derome et al. 2008; Cardoen et al. 2011; Gunter et al. 2013; Schneider et al. 2014). For instance, honeybee workers appear in two phenotypes: non-altruistic reproductive forms and altruistic, non-reproductive forms. Cardoen et al. (2011) hypothesized that environmental signals, received by the worker bees, control the activation of the ovaries. The authors found 1,292 genes, involved in multiple metabolic pathways, which were differentially transcribed between the two phenotypes and identified candidate genes which were potentially linked to the phenotypic differentiation between non-altruistic reproductive forms and altruistic reproductive forms. In another study, Filteau and colleagues (2013) used a weighted gene co-expression network analysis to identify the architecture of gene networks that were divergent between the two previously-discussed “normal” and “dwarf” types of the fish *C. clupeaformis*. The authors sampled brain and muscles tissue and found 14 and 17 co-expression modules, respectively, that differed between the two types. A gene network-based approach may lend additional functional information about morphological differentiation and divergence when gene annotation information is insufficient, and provides further, stronger correlation between gene expression and phenotype.

Other studies have used large-scale environmental disturbance as a natural experiment to reveal genes that contribute to phenotypes such as pollution or drought
tolerance. Natural disturbance events encapsulate numerous biotic and abiotic interactions that may be difficult or impossible to model under controlled conditions. For example, Whitehead et al. (2011) used microarrays to examine the impact of the Deepwater Horizon oil spill on the transcription and physiology of the Gulf killifish (F. grandis). The authors sampled individuals from six different field sites across three time points and identified more than 1,500 genes that were differentially expressed in response to oil stress. Additionally, individuals exposed to oil showed altered gill morphology. One responsive gene, cyp1a, is known to cause developmental abnormalities and decrease larval survival. A follow-up study under controlled conditions confirmed that the CYP1A protein was expressed in response to oil exposure, and this expression was particularly localized to areas of the gills, which showed altered morphology. This combined approach provided a quantitative link between differential transcription, differential protein expression, and individual phenotype. Leveraging the natural “treatment and control” design created by the oil spill allowed the authors to test hypotheses about organismal response in situ rather than relying on extrapolations from laboratory studies.

A few studies have combined transcriptomics with controlled studies of evolution to explore the molecular mechanisms of adaptation over ecological and evolutionary time. For example, Dhar et al. (2011) monitored adaptation to salt stress in Saccharomyces cerevisiae using both microarrays to examine expression differences and DNA sequencing to quantify sequence changes. Adaptation was quantified by measuring changes in population growth rate, a measure of fitness. Adaptive changes were correlated to a single SNP and differences in genome size, both of which may
have contributed to the differential expression of 1,431 genes. This indicated that the evolution of gene expression may have played an important role in adaptation to this stress (Dhar et al. 2011). Controlled studies of evolution allow investigators to track the effects of expression changes and phenotypic differentiation over evolutionary time, and quantify outcomes such as population divergence and adaptation.

Moving from correlation to causation

Ultimately, it will be imperative to move beyond correlating patterns of gene expression variation with patterns of trait variation. Incorporating manipulations of transcription (through e.g. transgenics, RNAi, or CRISPER/CAS) into transcriptome studies addresses the relative lack of data on whether differential transcription can “trickle up” to affect phenotype and ultimately populations (Ungerer et al. 2007; Dalziel et al. 2009, Pavey et al. 2012). Identifying the impacts of controlled changes in transcription in concert with ecologically-relevant traits in a natural setting will refine our understanding of well-known genetic pathways of interest in model and non-model systems and reveal how these pathways may have diversified across taxa. For example, based on annotation of flowering time genes in the model plant A. thaliana, Kobayashi et al. (2013) identified transcriptional changes in homologs of a floral pathway integrator (SbFT) and a floral repressor (SbSVP) before a community level mass flowering event in the tropical tree Shorea beccariana. The function of these genes was confirmed using transgenic A. thaliana: when compared with the wild type, the transgenic A. thaliana overexpressing SbFT showed early flowering, whereas late flowering was observed for those overexpressing SbSVP. Another study by Zhu et al.
(2008) examined gene expression data from segregating populations of yeast to construct gene networks. The authors used co-expression data, along with transcription factor binding site and protein-protein interaction information, to build gene networks. These gene networks were then linked to phenotype via eQTL analysis, and confirmed through analysis of gene knockout lines, allowing the authors to describe the causal effects of expression networks on phenotype. Although these two studies used transgenic individuals and knockout lines to confirm gene function, other manipulations such as RNAi or CRISPER/CAS have been used for emerging non-traditional model species (e.g. Hwang et al. 2013) and may be easily applied to non-model species of ecological interest. By focusing on this type of inquiry, ecological transcriptomics can continue its progression from a discipline that describes pattern, to one that elucidates process and informs ecological and evolutionary theory.

**Transcriptomics in the future: where do we go from here?**

Ten years of ecological transcriptomics have yielded descriptions of transcriptional variation in natural populations of a variety of organisms and in response to a variety of stimuli. We have described some of the reasons that microarrays or RNAseq have been appropriate based on study system and research question (box 1). As the field progresses, future studies, particularly of organisms with no genomic resources, will most likely rely on RNAseq (already 45% of studies reviewed use RNAseq), but microarrays may still offer valuable data, depending on the study system, the question, and design issues.
A unifying workflow

Although microarrays represented the primary method of whole-genome transcription quantification for the past decade, RNAseq studies have increased greatly in recent years, and previous statistical and technical limitations are rapidly being addressed. Given the advantages of each of the major transcriptomic technologies (see box 1), microarrays and RNAseq may be combined to test hypotheses about the importance of global gene expression patterns in natural populations (Malone & Oliver 2011). If a commercial or custom microarray is already designed for a given species, it may be easily applied to a new study on the same species, unless the question explicitly involves differences in expression of candidate genes that were not included in the array design. However, the user must decide whether the available probes are relevant for the question being investigated. For instance, a microarray based on locust (Schistocerca gregaria) ganglia (Badisco et al. 2011) will not be able to identify all relevant transcripts in other locust tissue types. A nonexhaustive list of commercially available and custom microarrays (Table 1.1) gives an indication of the wide range of taxa represented by existing microarrays. It is important to remember that unlike RNAseq, microarrays cannot give information about previously unidentified transcripts, transcript sequence or alternate isoforms. However, for quantifying variation in response to stimuli or surveying patterns of gene expression in the wild, microarrays are still a useful and viable choice, especially if they are already available.

In organisms without previous genomic resources, RNAseq will most efficiently quantify standing transcriptomic variation in a species and identify gene targets of interest. Once genes and gene networks of interest have been identified in the study
organism, these transcripts could be used to generate annotations, provide information about alternative gene isoforms, or to construct a specialized microarray for future studies. In an example of this integrated technique, Vera et al. (2008), used high-throughput RNAseq and *de novo* assembly of the Glanville fritillary butterfly (*Melitaea cinxia*) transcriptome from 80 individuals across eight populations. The sequencing results were used to construct a microarray, which was used in two follow-up studies, one examining differential gene expression between older and more recently-established populations of *M. cinxia* (Wheat et al. 2011) and another investigating heritable gene expression variation in *M. cinxia* larval development (Kvist et al. 2013). This series of investigations leveraged the ability of RNAseq to characterize previously-unexplored genomes to create a robust microarray for follow-up experiments. This is a powerful experimental pipeline for ecological transcriptomics of non-model organisms when genes of interest have already been identified with RNAseq. If the identification of novel transcripts is still of interest, or when generating ecological annotations, RNAseq can identify previously-unknown transcripts while still providing data on expression variation. This can be a bioinformatics challenge and many researchers will benefit from commercial options for bioinformatics.

While it is now possible for transcriptomics to probe genome-wide patterns, many groups still use single-locus or multi-locus assays, in the form of single- and multi-locus qPCR, that quantify the expression of a select sub-set of genes used to probe the transcriptome. These techniques are important in surveying ecologically-relevant candidate genes of interest for disease and response to environmental conditions (e.g. Aikawa et al. 2010; Schneider et al. 2014) and are important in the validation of gene
expression. The relatively low cost and high reliability of qPCR makes it valuable for validating genome-wide expression techniques for two purposes. First, technical validation confirms that the platform used to survey genome-wide expression is accurate. This validation is commonly performed after a genome-wide survey of expression. A second, less common biological validation confirms that the phenomenon of interest actually causes the observed variation in gene expression or vice versa (Kammenga et al. 2007). Ideally, biological validation of gene function uses independent biological samples to confirm the up- or down-regulation of genes in response to a given treatment or condition of interest. Therefore, although we did not include studies that relied solely on qPCR in our survey of transcriptomics, the use of qPCR for confirmation of the expression of genes of interest is essential.

Future directions for inquiry

Thanks to the power of genome-wide expression studies, we are now at a point in our understanding of genome function where we can and should move beyond telling single gene stories and start assembling a systems-level understanding of how organisms respond to environmental challenges (Whittkopp 2007; Richards et al. 2009; Schraiber et al. 2013). One trend that emerges from our survey of the literature is that as the molecular revolution in ecology progresses, ecological transcriptomics is moving from a largely descriptive discipline to one which identifies the causal elements of phenotypic change in wild populations (Aubin-Horth & Renn 2009; Andrew et al. 2013; Figure 1.1). In the first ten years of ecological transcriptomics work, authors rarely followed up to test the predicted importance of gene expression variation on response
to phenotype. Instead, most relied heavily on Gene Ontology types of analyses to infer relevant biological function without experimentally confirming that these inferences were true. Validating our findings from transcriptome studies may require assays at other molecular levels (e.g. Rees et al. 2011; Whitehead et al. 2011), and the use of knockouts (e.g. Dowen et al. 2012; Richards et al. 2012) or transgenic organisms (e.g. Kobayashi et al. 2013) grown in ecologically relevant experiments. Taking full advantage of the power of transcriptomics in ecology requires the integration of robust experimental designs and a synthetic approach that includes molecular, morphological, physiological or behavioral measurements at other levels of biological organization (Vasemagi & Primmer 2005; Richards et al. 2009).

As molecular ecology shifts from describing correlation to identifying causation, ecological transcriptomics will help elucidate the role of genomic elements that precede, regulate, and follow transcriptional modulation. Understanding the role of different genomic elements will allow investigators to more fully examine the pathways through which differential gene expression modulates phenotypic traits. Epigenetic mechanisms such as DNA methylation, which can result in mitotically or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence, have been correlated to a variety of ecological stimuli across taxa, and have potentially heritable effects on phenotype (Kilvitis et al. 2014). As a proximate cause of transcriptional variation, DNA methylation assays may add functionally relevant information to discriminate between environmental stimuli. For example, Dowen et al. (2012) identified changes in the model plant A. thaliana in response to bacterial pathogen, avirulent bacteria, and the defense hormone salicylic acid under laboratory conditions. The
authors identified differentially methylated regions throughout the genome, and used
RNAseq to identify differentially transcribed genes located near differentially methylated
regions. The authors were thus able to identify stress-response genes whose
transcription was altered by differential methylation in response to ecologically-relevant
stimuli. Studies have shown that DNA sequence-based differences can be context
dependent, but DNA methylation is even more labile to environmental influence, and
natural settings may induce alternative methylation profiles that would not be visible in
controlled settings. Combining genome-wide expression surveys of wild populations
with DNA methylation quantification may allow a more complete picture of the genetic
architecture of environmental response.

Proteomics may also provide an avenue for linking transcriptional variation to
larger biological processes (Vasemagi & Primmer 2005; Diz et al. 2012). Quantifying
protein expression may add functional information about a gene’s response as
understanding actual translation to protein product is crucial to mapping the ultimate
effects of differential gene expression on phenotype. Proteins may also be modified
post-translation to enhance or temper the cellular impact of differentially expressed
genes, making proteomics an important tool for measuring the final impact of gene
regulation on phenotype (Diz et al. 2012). Rees et al. (2011) examined gene and protein
expression in *Fundulus* species to correlate response at these two levels among three
populations of *Fundulus*. The authors found that although mRNA is generally positively
correlated with protein expression, the relationship is nonlinear. Regulatory mechanisms
amongst proteins may alter or enhance gene expression differences (Rees et al. 2011).
Further studies integrating these additional molecular markers, along with functional and
phenotypic analyses, may allow for a better mechanistic explanation of heritable differences between populations.

As the field of ecological transcriptomics now represents a major data stream in molecular ecology, investigators and institutions must build an infrastructure to support increased gene annotation. Traditional model organisms, such as *Mus musculus*, *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Arabidopsis thaliana* have a wide array of genomic resources available to them, including gene annotations, which identifies the function of a putative transcript, and predicted gene interactions. However, ecologists are typically interested in non-model organisms, which almost always lack these genomic resources. Despite increased access to genomic tools over the past decade, molecular ecologists are still limited in their ability to use genomic data because of a lack of information on relevant genes (Pavey *et al.* 2012; Andrew *et al.* 2013). Although there has been a growing push toward an ecological annotation of genes (Landry & Aubin-Horth 2007; Aubin-Horth & Renn 2009; Pavey *et al.* 2012; Andrew *et al.* 2013), investigators working on non-model organisms are still largely limited to using annotations from their closest model relative. As genetic distance increases, the likelihood increases that a putative ortholog, or a gene that is related by vertical descent, has diverged and an annotation from a model organism is not accurate. Further, even in model organisms, gene annotations are not available for the entire genome. Ecological gene annotations from wild species may greatly enhance annotations from related model organisms. As mentioned earlier, genes that are species- and context-specific may be vitally important in explaining ecological processes and interactions (Colbourne *et al.* 2011). A long-term solution to alleviate the
problem of poorly-annotated genes is the creation of a database for proposed ecological annotations (Pavey et al. 2012). While we have made much progress in ecology by examining traits and behaviors of individuals within and among populations, the integration of molecular techniques into ecology allows investigators unprecedented ability to examine the mechanistic underpinnings of the diverse phenotypes that contribute to phenotypic variation and rapid response to environment. An enhanced focus on ecological transcriptomics promises to contribute a powerful component to our understanding of the molecular basis of ecological interactions and evolutionary processes.

Acknowledgements

We thank Christy Foust, Holly Kilvitis, Marta Robertson, Larissa Williams and Koen Verhoeven and three thoughtful anonymous reviewers for valuable feedback on the manuscript. This work was supported by the University of South Florida (MA & CLR).

References


Dowen, R. H., Pelizzola, M., Schmitz, R. J., Lister, R., Dowen, J. M., Nery, J. R., . . .


Zhu, J., Zhang, B., Smith, E. N., Drees, B., Brem, R. B., Kruglyak, L., Bumgarner, R. E.,
Figures and Tables

Figure 1.1. Number of studies each year addressing each of the question categories.

Question categories: 1 corresponds to studies that investigated the extent and structure of natural variation, 2 corresponds to studies that investigated organismal response to stimulus and 3 corresponds to studies that investigated the effect of differential gene expression on phenotype.
Figure 1.2. Venn diagram of question categories in studies examined. Black numbers represent question categories. Red numbers indicate the number of studies in the question category. Overlapping areas represent studies that were represented by more than one question category.
<table>
<thead>
<tr>
<th>Agilent</th>
<th>Affymetrix</th>
<th>Custom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis thaliana</td>
<td>Arabidopsis thaliana</td>
<td>Eucalyptus grandis</td>
</tr>
<tr>
<td>Bovinae spp. (bovine)</td>
<td>Bovinae spp. (bovine)</td>
<td>Montastraea faveolata</td>
</tr>
<tr>
<td>Brassica spp.</td>
<td>Caenorhabditis elegans</td>
<td>Salmo salar</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>Callitrichidae spp.</td>
<td>Schistocerca gregaria</td>
</tr>
<tr>
<td>Canis spp. (canine)</td>
<td>Canis spp. (canine)</td>
<td>Acropora millepora</td>
</tr>
<tr>
<td>Culicidae spp. (mosquito)</td>
<td>Danio rerio (Zebrafish)</td>
<td>Carpodacus mexicanus</td>
</tr>
<tr>
<td>Danio rerio (Zebrafish)</td>
<td>Drosophila melanogaster</td>
<td>Salvelinus fontinalis</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>Equus ferus (horse)</td>
<td>Amphiura filiformis</td>
</tr>
<tr>
<td>Equus ferus (horse)</td>
<td>Felidae spp. (feline)</td>
<td>Apis mellifera carnica</td>
</tr>
<tr>
<td>Gallus gallus (chicken)</td>
<td>Gallus gallus (chicken)</td>
<td>Anopheles gambiae</td>
</tr>
<tr>
<td>Gossypium spp. (cotton)</td>
<td>Glycine max (soybean)</td>
<td>Crassostrea gigas</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>Homo sapiens</td>
<td>Crassostrea virginica</td>
</tr>
<tr>
<td>Hordeum vulgare (barley)</td>
<td>Macaca mulatta (rhesus)</td>
<td>Folsomia candida</td>
</tr>
<tr>
<td>Leporidae spp. (rabbit)</td>
<td>Medicago spp.</td>
<td>Tribolium castaneum</td>
</tr>
<tr>
<td>Macaca mulatta (rhesus)</td>
<td>Mus musculus</td>
<td>Mytilus Californianus</td>
</tr>
<tr>
<td>Magnaporthea spp.</td>
<td>Oryza sativa (rice)</td>
<td>Ambrosia artemisiifolia</td>
</tr>
<tr>
<td>Medicago spp.</td>
<td>Ovis aries (sheep)</td>
<td>Loxodonta africana</td>
</tr>
<tr>
<td>Mus musculus (mouse)</td>
<td>Saccharomyces cerevisiae</td>
<td>Daphnia magna</td>
</tr>
<tr>
<td>Nicotiana tabacum (tobacco)</td>
<td>Schizosaccharomyces</td>
<td>Pimephales promelas</td>
</tr>
<tr>
<td>Oryza sativa (rice)</td>
<td>Suidae spp. (porcine)</td>
<td>Daphnia pulex</td>
</tr>
<tr>
<td>Ovis aries (sheep)</td>
<td>Taeniopygia guttata (Zebra Finch)</td>
<td>Karenia brevis</td>
</tr>
<tr>
<td>Rattus norvegicus (rat)</td>
<td></td>
<td>Meliteaea cinxia</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td></td>
<td>Helianthus annus</td>
</tr>
<tr>
<td>Salmonidae spp. (salmon)</td>
<td></td>
<td>Brugia malayi</td>
</tr>
<tr>
<td>Solanum lycopersicum (tomato)</td>
<td></td>
<td>Pinus taeda</td>
</tr>
</tbody>
</table>
Table 1.1 (Continued)

<table>
<thead>
<tr>
<th>Agilent</th>
<th>Affymetrix</th>
<th>Custom</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Suidae spp.</em> (porcine)</td>
<td><em>Tursiops truncatus</em></td>
<td></td>
</tr>
<tr>
<td><em>Triticum spp.</em> (wheat)</td>
<td><em>Ruditapes philippinarum</em></td>
<td></td>
</tr>
<tr>
<td><em>Xenopus spp.</em></td>
<td><em>Anemonia viridis</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Solanum lycopersicum</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Dreissena polymorpha</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Gasterosteus aculeatus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Solenopsis invicta</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Pinus pinaster</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Acropora palmata</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Neurospora tetrasperma</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Onthophagus taurus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Petrolisthes cinctipes</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Lepeophtheirus salmonis</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Calanus finmarchicus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Azospirillum brasilense</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Picea glauca</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Metarhizium robertsii</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Locusta migratoria</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Neurospora crassa</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Platichthys flesus</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Megachile rotundata</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Neotoma lepida</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Laternula elliptica</em></td>
<td></td>
</tr>
<tr>
<td>Agilent</td>
<td>Affymetrix</td>
<td>Custom</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Coregonus cupleaformis</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Drosophila mojavensis</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Ruditapes philippinarum</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Lagopus lagopus</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Ostreococcus tauri</em></td>
</tr>
</tbody>
</table>
Chapter Two:
Genome-wide differential gene expression in the foundation salt marsh grass
*Spartina alterniflora* caused by the Deepwater Horizon oil spill\(^2\)

Abstract

In 2010, the *Deepwater Horizon* oil spill released an estimated 4.9 million barrels of oil into the Gulf of Mexico, making landfall on Gulf salt marsh habitat dominated by the foundation species *Spartina alterniflora*. Despite the severe impact, *S. alterniflora* proved remarkably resilient in the face of the crude oil stress. However, the molecular mechanisms of tolerance in this highly resilient species remain understudied. We tested the hypothesis that crude oil exposure would induce differential gene expression in affected populations, and that we could use these data to identify novel transcript behavior in response to the natural stressor. To understand the molecular response of *S. alterniflora* to crude oil stress, we used whole-transcriptome microarrays, and identified genes likely to be orchestrating the response to crude oil stress. We expected to see diverse categories of genes involved in the response to crude oil that may not have been previously annotated, given the cryptic expression and behavior of genes in

---

\(^2\) This chapter has been submitted for publication and is co-authored by Mariano Alvarez, Julie Ferreira de Carvalho, Armel Salmon, Malika Ainouche, Aaron Schrey, Sydney Moyer, and Christina Richards. M. Alvarez and C. Richards performed sample collection, gene expression analyses, and wrote the manuscript. J. Ferreira de Carvalho, A. Salmon, and M. Ainouche created the custom microarray. M. Alvarez and S. Moyer performed technical confirmations.
complex natural environments. We identified 3,622 genes that responded exclusively to oil stress, and used a highly-differentially expressed subset of these genes to construct gene interaction networks. These genes and gene categories exhibited previously unobserved behavior and patterns in response to crude oil exposure, confirming our hypothesis and providing a novel resource for the *Spartina* genus.

**Introduction**

Human-induced environmental change is now a dominant evolutionary force on earth, altering populations and communities through direct and indirect effects of development and commerce (Palumbi *et al.* 2001). Anthropogenic stressors, such as urbanization, pollution, and exploitation, have altered community composition and created novel selection pressures (Crowe *et al.* 2000, Medina *et al.* 2008, Orsini *et al.* 2012). Coastal ecosystems are among the most vulnerable environments and have experienced rapid evolutionary changes in response to severe anthropogenic selection pressures (Smith and Bernatchez 2007, Halpern *et al.* 2008). In particular, pollutant releases, like oil spills, pose a major threat to coastal ecosystems by degrading habitat and inducing heritable changes in natural populations via selection and genotoxicity (Medina *et al.* 2008). Oil spills threaten coastal ecosystems because of their toxicity and because of the difficulty in containing them (Lin and Mendelssohn 1998, 2012, Silliman *et al.* 2012). Both the chemical effects, such as polyaromatic hydrocarbon (PAH) toxicity, and the physical effects, such as coating the foliage, impose immediate and long-term stresses on the organisms that come in contact with crude oil (reviewed in Pezeshki *et al.* 2000). In order to protect these important ecosystems from the deleterious effects of
oil spills, it is imperative to understand the mechanisms by which organisms cope with these stressors (Smith and Bernatchez 2007).

The incorporation of molecular techniques into ecology and evolution has given researchers unprecedented access to underlying mechanisms that shape phenotype and govern the pattern and process of microevolution (Alvarez et al. 2015). As transcriptome data becomes easier to generate in free-living, non-model organisms, researchers have used gene expression data to identify the genomic response to ecologically relevant environmental factors (Allison et al. 2006, Kammenga et al. 2007, Allendorf et al. 2010, Alvarez et al. 2015). For example, Williams and Oleksiak (2008) looked for signatures of selection in populations of Fundulus heteroclitus living in polluted Superfund populations. The authors found 1-6% of loci showed signatures of selection in F. heteroclitus when compared to those living in unpolluted populations (Williams and Oleksiak 2008). Other studies have discovered novel gene expression patterns and functions that became apparent only in response to complex natural stimuli. These cryptic expression signatures would otherwise remain unidentified in controlled laboratory settings, even in well-studied model organisms (Colbourne et al. 2011). As molecular ecologists identify and annotate genes and gene networks that are responsive to natural stimuli (e.g Richards et al. 2012, Whitehead et al. 2011), authors have called for an “ecological annotation” to complement existing “biological”, “cellular” and “molecular” annotations that have been adopted across model species (Pavey et al. 2012). Work in non-model species that inhabit diverse ecologies would be insightful in this effort, however, use of these technologies is often significantly slower than in model
species, and there is a need to expand molecular tools to more diverse species of ecological interest (Alvarez et al. 2015).

In 2010, the Deepwater Horizon oil spill released an estimated 4.9 million barrels of oil into the Gulf of Mexico (National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling 2011). This oil eventually made landfall on the northern shores of the Gulf of Mexico, impacting the shorelines of Louisiana, Mississippi, and Alabama (Thibodeaux et al. 2011, Mendelssohn et al. 2012). Of the 1,773 kilometers of coastline oiled by the Deepwater Horizon disaster, nearly half was salt marsh habitat dominated by the foundation species Spartina alterniflora (Michel et al. 2013). Salt marsh habitats are dynamic and heterogeneous environments that provide ecosystem services such as erosion mitigation and sediment filtering (Turner 1976, Turner and Boesch 1987, Bertness and Bird 1999, Pennings and Bertness 2001). Tidal and rainfall influences create wide salinity gradients across the salt marsh which structures variation within and between species (Callaway et al. 1990, Pennings and Bertness 2001). In addition to natural environmental gradients, salt marshes have long been the target of exploitation and development (Gedan et al. 2009), and are thus frequently impacted by anthropogenic stress. Despite unpredictable disturbance and stress, S. alterniflora-dominated salt marshes are also remarkably resilient in the face of a variety of anthropogenic impacts, including the crude oil stress imposed by the Deepwater Horizon oil spill (Gedan et al. 2009, Silliman et al. 2012).

Although tragic, the patchwork of populations that were either affected or unaffected by the Deepwater Horizon spill creates a unique opportunity to study the effects of this recurrent yet unpredictable stressor on the molecular processes of S.
alterniflora, and to dissect the molecular mechanisms of tolerance in this highly resilient species. There have been several gene expression studies in S. alterniflora under controlled conditions that have identified differential gene expression in response to stress. Baisakh and colleagues (2008) sequenced expressed sequence tags produced under high salinity conditions and found differential expression in seven genes of 9 surveyed due to salt stress. Other studies identified 17 genes that were response to heat stress and 28 genes that responsive to crude oil stress through a “gene fishing” approach using 20 primers (Baisakh and Subudhi 2009, RamanaRao et al. 2012). However, these studies have relied on low-coverage sequencing and quantitative PCR (qPCR) to quantify the response of a limited numbers of candidate genes rather than capturing the response of the entire transcriptome.

To understand the molecular response of S. alterniflora to crude oil stress, we examined the differential regulation of gene expression of natural S. alterniflora populations in response to the Deepwater Horizon oil spill with whole-transcriptome microarrays, created from 454 pyrosequencing (de Carvalho et al. 2013). In order to dissect the regulatory mechanisms of oil response, we explored the genes and gene networks that are likely to be involved in the regulation of phenotype in response to the oil stress. Our previous work has found high genetic diversity within and among S. alterniflora populations (Richards et al. 2004, Foust et al. 2016, Robertson et al., in prep). Therefore, we expected to see high levels of expression differentiation between populations. However, we also expected to see large differences in gene expression due to exposure to oil, particularly in genes annotated as being part of stress response pathways. We expected to see diverse categories of genes involved in the response to
crude oil that may not have been previously annotated as such for two reasons: 1) 
*Spartina alterniflora* has a complex polyploid genome with potential for gene copies to 
undergo functional diversification (Fortune *et al.* 2007) and 2) previous studies of 
response to complex environments has identified novel transcript behavior (Colbourne 

**Methods**

*Study species and population*

*Spartina alterniflora*, or smooth cordgrass, is a halophyte grass that grows in dense stands in the heterogeneous gradient of the salt marsh, and comprises up to 90% of the biomass in native habitats along the east coast of the United States (Richards *et al.* 2005). *Spartina alterniflora* is native to the east coast of the United States and the Gulf of Mexico (Pennings and Bertness 2001, Hughes and Lotterhos 2014), and invasive worldwide (Callaway and Josselyn, 1992, Ayers *et al.* 2004, An *et al.* 2007, Ainouche *et al.* 2009). *Spartina alterniflora* is a foundation species in its native salt marshes, providing refuge for invertebrates (Silliman and Bertness 2002), nurseries for birds and fish (Mendelssohn *et al.* 2012), and habitat-building services (Pennings and Bertness 2001). Despite the dynamic abiotic gradients of the salt marsh and the frequent disturbance imposed by anthropogenic events, *S. alterniflora* persists across a wide range of environmental conditions and shows remarkable resilience to a variety of stressors (Nestler 1977, Pennings and Bertness 2001, Richards *et al.* 2005, Gedan *et al.* 2009, Silliman *et al.* 2012). The resilience of *S. alterniflora* has also been demonstrated in controlled salinity and heat experiments, under conditions substantially
harsher than those found in the salt marsh (Baisakh et al. 2008, Baisakh and Subudhi 2009, Subudhi and Baisakh 2011). Even in response to severe oil stress, *S. alterniflora* has shown up to 100% recovery after 7 months (Lin and Mendelssohn 2012), despite reductions in carbon fixation and transpiration (Lin and Mendelssohn 2012, RamanaRao et al. 2012, Silliman et al. 2012). The extreme resilience of *S. alterniflora* may mitigate anthropogenic damage in Gulf of Mexico salt marsh ecosystems, and understanding the molecular underpinnings of this ecologically important response may provide valuable information for predicting response to climate change and conservation of these ecosystems. However, the molecular mechanisms of stress response in *S. alterniflora*, and crude oil response in particular, remain poorly understood.

In August 2010, four months after the *Deepwater Horizon* oil spill, we traveled to two estuarine locations in Grand Isle, Louisiana and one location in Bay St. Louis, Mississippi. We collected samples of leaf tissue from 10 individuals, spaced 5 meters apart, in a total of three contaminated and three uncontaminated populations of *S. alterniflora* (Figure 2.1). Contamination was assessed by the visual presence of oil on the sediment at each of these locations. Nearby uncontaminated populations did not have any visual signs of the presence of oil. From each plant, we collected the 3rd fully expanded leaf to standardize age and minimize developmental bias from sampling. Leaf samples were immediately frozen in liquid nitrogen to prevent RNA degradation and kept frozen during transport to the University of South Florida for processing and analysis.
RNA extraction and microarray hybridization

We extracted total RNA from each of nine plants separately per population from homogenized leaf tissue using RNeasy Plant Mini Kits (QIAGEN). The Interdisciplinary Center for Biotechnology Research at the University of Florida standardized RNA concentrations and created three pools of three individuals for each population: a total of 54 samples were combined into 18 population-specific RNA pools. Pooling is a common strategy in ecological research that sacrifices measures of individual-level variation to increase sample size and capture population-level response (Alvarez et al. 2015). Twenty RNA pools (18 sample pools and 2 technical replicates) were reverse transcribed into cDNA and hybridized to a custom 4x44k Agilent microarray containing 17,049 unique 60-mer probes corresponding to 16,608 unique annotations and 441 unannotated contigs. Of these, 9356 probes were designed from *S. alterniflora* 454 EST assemblies, 7170 from the EST co-assemblies of 5 *Spartina* species, and 523 from *Spartina maritime* EST assemblies.

Data analysis

Raw florescence data was imported into the statistical program JMP/Genomics (Version 6 for Windows; SAS Institute, Cary, NC, USA) for analysis. We filtered out intensity values less than 2 and subsequent probes that appeared in less than half of the pools (<10). Of the total 17,049 probes, 15,907 passed our filtering protocol. We then median normalized the raw data and visualized normalized data using principal components analysis. To visualize differentiation both between population and in response to oil treatment, we used a principal variance components analysis (PVCA) on
the probe-level data. This strategy uses a principal components analysis to reduce the
dimensionality of the data before calculating variance components, via a mixed linear
model, for each principal component (Richards et al. 2012). To understand the effects of
oil exposure, population and state (Louisiana or Mississippi) on gene expression, we fit
a mixed-model ANOVA on these data using a model that incorporated microarray slide,
state, population, and treatment, with population nested within treatment (expression =
slide + treatment + state + population-within-oil). Populations were nested within
treatment because populations affected by oil stress were not the same populations as
control populations. State was included to control for the comparatively large
geographic distance between the four Louisiana populations and the two Mississippi
populations.

Gene interaction networks

To explore the functionality of differentially expressed genes, we used Gene
Ontology (GO) based on annotation data from the model species A. thaliana (TAIR 10).
We first used gene set enrichment, as measured by the Kolmogorov-Smirnov test, to
look for overrepresented GO terms within the oil-responsive genes identified in our
ANOVA in JMP/G. We then used Virtual Plant to generate gene interaction networks
using previously published data from the model plant A. thaliana. To accomplish this, we
only used genes in S. alterniflora that had homologs in A. thaliana. Of the 15,907
transcripts included in our ANOVA, we found 14,670 S. alterniflora contigs with
homology to genes in A. thaliana, corresponding to 7,606 unique gene hits (due to
multiple S. alterniflora contigs corresponding to the same A. thaliana homolog). We
used these homologs to generate gene interaction networks, made of genes (nodes) and interactions (edges), by building edges between 1,211 highly differentially expressed gene homologs (P<0.001). Edges were parameterized using previously generated data from *A. thaliana*, including information on micro RNA binding populations, protein-to-protein interactions, transcriptional regulation (which includes transcription factors, enhancers, and repressors), and transport interactions. Computationally generated data from the metabolic interaction databases Aracyc and KEGG created additional edges, and some edges were further generated using published literature interactions. We visualized the resulting network using Cytoscape and counted the number of connections using Virtual Plant. Using our gene interaction networks, we identified highly connected genes as targets for our technical confirmations.

*Target genes and technical confirmation*

Using the same RNA as used in the microarray hybridization, we reverse transcribed total RNA using RetroScript kits (Ambion). We used the resulting cDNA as template for quantitative PCR (qPCR) confirmation of the microarray. From our list of highly connected oil-responsive target genes, we selected 32 genes to create primers for confirmation of the microarray. These included several additional genes that were highly responsive to oil but not represented in our list of highly connected genes. These additional genes were identified from literature and annotation information as encoding epigenetic regulatory mechanisms, whose interactions were not quantified as part of our network construction but may exert important broad effects on phenotype (Cortijo et al.)
Primers were generated for these target genes as well as for $a$-tubulin, which has been validated as an endogenous control (Baisakh *et al.* 2008, RamanaRao *et al.* 2012). Reactions were run in duplicate on a 96 well plate, using template from three individuals from each population (for a total subsample of 18). Out of the 32 primer pairs generated, we were able to quantify differential expression for three genes using the delta-delta Ct method with corrections for primer efficiency (Schmittgen and Livak 2008). The other 29 showed poor amplification and could not be reliably scored.

**Results**

*Oil-contaminated populations were differentiated from uncontaminated controls*

Despite the potential for environmental variation in field studies as well as for possible variation introduced during the pooling process, our technical replicate pools were highly correlated with each other ($r=0.995$ for $n=2$ one oiled, one non-oiled population). In our PVCA, principal component (PC) 1 explained 37.8% of the total variation, with the highest loading by oil contamination (60%). PC1 separated the uncontaminated Louisiana populations from the contaminated Louisiana populations and both populations in Mississippi (Figure 2.2). Principal components 2 and 3 explained 16.8% and 9.8% of the variance, respectively, and were both loaded primarily by population-level variance. However, our uncontaminated population in Bay Saint Louis was differentiated from other populations not by PC1 or PC2, but by PC3, which was primarily loaded by population-level variance (Figure 2.2 – PVCA). This grouping of our uncontaminated Mississippi population with the contaminated populations may be because of unique population structure in our sampled Mississippi populations,
historical contamination, or because oil from the *Deepwater Horizon* had actually made landfall at this location and we were unable to visually confirm this.

Using a mixed-model ANOVA, we found expression of 6,495 genes significantly differed by population, and another 7,614 genes significantly differed by oil exposure (FDR q< 0.05, Figure 2.3). Of these 3,622 genes were significantly differentially expressed exclusively due to oil exposure (Figure 2.3). We found high levels of differentiation between populations, with over 50% of expression variation explained by population.

*Gene Ontology and network analysis identified genes and categories of interest*

To understand the function of genes that responded to oil stress, we used a gene set enrichment analysis, which identifies GO categories that appear more often than they would due to chance (JMP/G). We found four overrepresented categories: “chloroplast photosystem I”, “mitochondrial proton-transporting ATP synthase complex, catalytic core F(1)”, “RNA polyadenylation”, and “response to far red light”. Enrichment of these functional categories, particularly those involved in photosynthesis, may reflect physiological processes taking place during the response to oil stress, including the reduced overall biomass that oil-exposed plants exhibit in the field (Lin and Mendelssohn 2012).

To understand how the interaction of genes and gene products may contribute to crude oil response, we used highly oil-responsive genes (FDR q < 0.001) genes from our ANOVA to construct gene interaction networks (GIN). We used a total of 1,211 highly differentially expressed genes with homologs in *Arabidopsis thaliana* to visualize
their interactions in Virtual Plant (Katari et al. 2010), and create a gene interaction network (Figure 2.4). We selected highly connected genes from the GIN, along with epigenetic regulators of interest, as targets of interest for our downstream qPCR assay (Table 2.1). Of the 32 total primers generated, we were able to generate PCR product for three of them, which were up- or downregulated in the same direction as they were in our microarray (Figure 2.5-2.6). The concurrence between our qPCR results and our microarray confirm the precision of the custom microarray.

**Discussion**

Anthropogenic stressors, such as crude oil spills, can be leveraged to create natural “treatment and control” designs to understand molecular function in ecologically relevant settings (see Chapman et al. 2011, Whitehead et al. 2012). Crude oil is a common anthropogenic stressor in Gulf of Mexico salt marsh communities due to oil exploitation activities of the oil industry (Gedan et al. 2009). Crude oil is composed primarily of hydrocarbons and a mixture of heavy metals, including arsenic, mercury, cadmium, and vanadium (Gohlke et al. 2011), and induces toxicity in plants (Mendelssohn et al. 2012) and animals (Anderson et al. 1974, Gulec et al. 1997, Whitehead et al. 2011). Crude oil exerts its toxic effects by both its chemical properties and by physically coating leaves and roots (Pezeshki et al. 2000). We expected to see a large number of genes across multiple GO categories involved in the response to crude oil that may not have been previously annotated as such for two reasons, primarily because of the potential for neofunctionalization in *S. alterniflora* (Fortune et al. 2007) and the power of complex, natural settings to elicit novel expression patterns.
Our study found 3,622 genes that were differentially expressed due to crude oil exposure alone in natural populations of *S. alterniflora*. Because this design studies the effects of a complex stressor in a natural setting, our microarray revealed context-specific expression variation and novel transcript behavior that may not have been visible under controlled conditions (Dalziel *et al.* 2009, Colbourne *et al.* 2011). For example, genes annotated as being responsive to far-red stimulus were overrepresented in our microarray data. This behavior, to our knowledge, has not been documented outside of this study. Additionally, several oil responsive genes, such as HD1, were not significantly differentially expressed in previous stress response experiments (Baisakh *et al.* 2008). Thus, our oil-responsive genes may be considered “ecologically annotated” (Landry and Aubin-Horth 2009, Pavey *et al.* 2012), an important step in the development of molecular resources for *S. alterniflora*.

One advantage of whole-transcriptome assays is the ability to assess the activity of genes working in concert with each other, rather than a few chosen loci working independently (Alvarez *et al.* 2015). We generated gene interaction networks to explore the architecture of crude oil response and identified highly connected genes based on computational predictions and empirical data from *A. thaliana* that may be important contributors to the response to crude oil. High connectivity suggests that these genes are essential components of the larger molecular response to our oil stress, beyond their relationships in localized pathways that they are a part of (Lee *et al.* 2008). Many of these gene targets individually play a role in cellular phenotype. AtRNR1, for instance, is a ribonucleotide reductase involved in the production of dNTP for DNA
replication and repair (Tang et al. 2012). AtOST1 is a protein kinase, and mutations in this gene disrupted stomatal closure and opening in A. thaliana (Imes et al. 2013). AtOST1 plays a role in drought and freezing resistance, affecting organismal phenotype beyond cellular phenotype (Ding et al. 2015).

We were also able to identify other GO categories that were significantly overrepresented in our differentially expressed genes. Enrichment of these functional categories, particularly those involved in photosynthesis like “chloroplast photosystem I”, may reflect physiological processes contributing to the response to oil stress, such as the reduced overall biomass that oil-exposed plants exhibit in the field (Lin and Mendelssohn 2012).

We found a number of S. alterniflora genes that were differentially expressed due to oil and could be annotated as such, but we relied heavily on information from Arabidopsis thaliana for network construction. Although S. alterniflora and A. thaliana are not closely related, model organisms, such as Arabidopsis thaliana, have high-resolution genomic maps, detailed genome annotations, comparatively well-understood metabolic pathways, and a number of analysis tools and databases (e.g AraCyc, Rhee et al. 2005; Virtual Plant, Katari et al. 2010). In contrast, the non-model S. alterniflora, and non-model organisms in general, have comparatively few genetic resources available. In particular, the interaction data used to create expression networks is only available in A. thaliana, and we were able to use only homologous genes to create these networks. Leveraging the substantial genetic resources of model organisms provides a “first look” into genome function during oil stress in S. alterniflora without incurring substantial resource investment. Nevertheless, there is evolutionary
divergence between *A. thaliana* and *S. alterniflora*, and it is very likely that novel genes present in *S. alterniflora* contribute to the phenotypic response to crude oil (Colbourne *et al.* 2011). Even in *S. alterniflora* genes with substantial homology to *A. thaliana*, the functions of those genes may vary through neofunctionalization of paralogs (Wagner *et al.* 2000, Larracuente *et al.* 2008). The *Spartina* genus has a particularly high potential for neofunctionalization of paralogous genes because of historical genome duplication that resulted in multiple copies of genes (Fortune *et al.* 2007). Genome duplication creates the opportunity for both novel coding regions and novel function (Fortune *et al.* 2007, Flagel and Wendel 2009). Fortune and colleagues (2007) examined the evolution of *Waxy* gene paralogs throughout the *Spartina* genus and found that all members of the *Spartina* genus contained between 1 and 3 copies of two paralogs, *WaxyA* and *WaxyB*, creating the potential for novel function of these divergent paralogs as both are retained (Fortune *et al.* 2007). Interrogation of gene function for target genes will be crucial to confirm the impacts of each of these genes on phenotype.

Transcriptome assays have enabled unprecedented access to the patterns that underlie phenotypic response to environmental stimulus, and it is crucial to link molecular variation to ecology and evolution by correlating gene expression patterns to fitness (Alvarez *et al.* 2015). The data generated in this study are correlative, therefore detailed reverse genetic screens of oil-responsive genes found in this study, using gene knockdowns, knockouts, or overexpression mutants, may be insightful to confirm both the importance of these genes in regulating their respective networks and their effects on phenotype. Further, as the *Deepwater Horizon* oil spill caused extensive mortality as well as stress, changes in genetic variation in response to selection may have shaped
the population-level gene expression patterns observed in our study (Robertson et al. unpublished). Selection on both coding and non-coding regions can impact gene expression, which is a heritable trait over evolutionary time (Oleksiak et al. 2002). Alternatively, variation in gene expression between populations may be modulated by epigenetic modification, which may vary in absence of genetic variation (Kilvitis et al. 2014, Robertson and Richards 2015).

Although it has been well characterized in the ecological literature (e.g Pennings and Bertness 2001, Hughes et al. 2008, Hughes and Lotterhos 2014), genomic resources to characterize the molecular basis of S. alterniflora traits have only recently been developed (de Carvalho et al. 2013). Our study adds to the existing body of literature on S. alterniflora by providing a population-level scan of gene expression in oil-exposed natural populations. Similar to other studies of genetic diversity in S. alterniflora (Richards et al. 2004, Foust et al. 2016), we found high levels of differentiation between populations, and over 50% of expression variation was explained by population, while only 25% of the variation in expression was explained by contamination. Our custom microarray serves as a novel, validated resource for the Spartina genus, which can be applied to populations around the world. As a foundation species comprising the majority of the biomass in native salt marshes, S. alterniflora plays a vital role in mediating the resilience of the salt marsh to continued anthropogenic impact, through oil spills or other exploitation (Pennings and Bertness 2001, Silliman et al. 2012). Understanding the mechanisms of its remarkable resilience is vital to understanding the evolutionary fate of oil-exposed populations of native S. alterniflora in the salt marshes of the Gulf of Mexico.
References


Lin, Qianxin, and Irving A. Mendelssohn. "Impacts and recovery of the Deepwater Horizon oil spill on vegetation structure and function of coastal salt marshes in the northern Gulf of Mexico." Environmental science & technology 46.7 (2012): 3737-3743.


Figure 2.1. Map of Gulf of Mexico study sites. Green markers represent sites with no visible oil and red dots represent sites with visible oil in or on sediment.
**Figure 2.2.** Principal variance components analysis. Figure were generated using JMP Genomics.
Figure 2.3. Results of ANOVA, using oil, population (which is nested within oil) and state.
**Figure 2.4.** Interaction network of all genes highly responsive to oil (P<0.001), visualized in Cytoscape. Nodes in light blue represent protein-coding genes, while nodes in dark blue represent all other genes. Orange circular nodes represent metabolites, and purple triangles represent transcription factors.
Table 2.1. Highly oil-responsive target genes and their respective homologs. Target Rationale and Description outlines whether genes were chosen because of number of connections in the network analysis or because of other interest.

<table>
<thead>
<tr>
<th>Spartina alterniflora contig ID</th>
<th>Arabidopsis annotation</th>
<th>Target Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_alt_contig08070</td>
<td>At4g38130</td>
<td>negative regulation of transcription, histone deacetylase 1, ATHD1</td>
</tr>
<tr>
<td>S_alt_contig00846_RC</td>
<td>At2g25170</td>
<td>negative regulation of transcription, chromatin remodeling factor, CHD3</td>
</tr>
<tr>
<td>S_alt_contig03288</td>
<td>At2g22540</td>
<td>negative regulation of transcription, K-box region and MADS-box transcription factor, AGL22</td>
</tr>
<tr>
<td>S_5sp_contig45211</td>
<td>At2g26640</td>
<td>metabolic process, involved in the biosynthesis of VLCFA, 107 connections</td>
</tr>
<tr>
<td>S_alt_contig12988</td>
<td>At3g62870</td>
<td>RNA methylation, ribosome biogenesis, translation, 64 connections</td>
</tr>
<tr>
<td>S_alt_contig04438_RC</td>
<td>At5g23960</td>
<td>terpene synthase, ATTPS21, 44 connections</td>
</tr>
<tr>
<td>S_mar_contig02490_RC</td>
<td>At2g35160</td>
<td>negative regulation of transcription, histone modification, methylation, histone methyltransferase, SU(VAR)3-9 homolog 5</td>
</tr>
<tr>
<td>S_alt_contig08149</td>
<td>At4g38630</td>
<td>proteasome, 57 connections cellular polysaccharide biosynthetic process, encodes a mitochondrial-localized CMP-KDO synthetase, activating KDO as a nucleotide sugar, ATCKS</td>
</tr>
<tr>
<td>S_alt_contig03666_RC</td>
<td>At1g53000</td>
<td>negative regulation of transcription, histone deacetylase 1, ATHD1</td>
</tr>
<tr>
<td>S_alt_contig11442</td>
<td>At4g38130</td>
<td>cytoskeleton organization, expressed during reproductive development, ACT11</td>
</tr>
<tr>
<td>S_5sp_contig44561</td>
<td>At3g12110</td>
<td>cytoskeleton organization, expressed during reproductive development, ACT11</td>
</tr>
<tr>
<td>S_alt_contig04730_RC</td>
<td>At3g12110</td>
<td>cytoskeleton organization, expressed during reproductive development, ACT11</td>
</tr>
<tr>
<td>Spartina alterniflora contig ID</td>
<td>Arabidopsis annotation</td>
<td>Target Rationale</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>S_5sp_contig30510</td>
<td>At1g09700</td>
<td>cellular response to organic substance, involved in mRNA cleavage, dsRNA-binding, DRB1</td>
</tr>
<tr>
<td>S_alt_contig05500</td>
<td>At5g63840</td>
<td>cellular polysaccharide biosynthetic process, cell morphogenesis, glycosyl hydrolases family 31 protein, mutant impaired in cellulose production, PSL5</td>
</tr>
<tr>
<td>S_alt_contig07745</td>
<td>At1g29900</td>
<td>cellular nitrogen compound biosynthetic process, encodes carbamoyl phosphate synthetase (CPS) large subunit, CARB</td>
</tr>
<tr>
<td>S_5sp_contig17182</td>
<td>At3g48560</td>
<td>cellular nitrogen compound biosynthetic process, chlorsulfuron/imidazolinone resistant 1, catalyzes the formation of acetolactate from pyruvate, AHAS</td>
</tr>
<tr>
<td>S_alt_contig10959</td>
<td>At4g39400</td>
<td>cell morphogenesis, involved in brassinosteroid signal transduction (makes plant bigger), ATBRI1</td>
</tr>
<tr>
<td>S_alt_contig00645_RC</td>
<td>At2g33150</td>
<td>carboxylic acid metabolic process, involved in fatty acid b-oxidation during germination and subsequent seedling growth, KAT2, 54 connections</td>
</tr>
<tr>
<td>S_5sp_contig13752_RC</td>
<td>At3g06860</td>
<td>carboxylic acid metabolic process, multifunctional protein, involved in peroxisomal fatty acid oxidation, ATMFP2, 45 connections</td>
</tr>
<tr>
<td>S_alt_contig05152</td>
<td>At3g51840</td>
<td>carboxylic acid metabolic process, acyl-CoA oxidase, first step of peroxisomal fatty acid beta-oxidation during early, post-germinative growth, ACX4, 30 connections</td>
</tr>
<tr>
<td>S_alt_contig00939_RC</td>
<td>At1g01040</td>
<td>cellular response to organic substance, Dicer-homolog, ASU1</td>
</tr>
<tr>
<td>Spartina alterniflora contig ID</td>
<td>Arabidopsis annotation</td>
<td>Target Rationale</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>S_alt_contig06344</td>
<td>At2g35630</td>
<td>cytoskeleton organization, ARM repeat superfamily protein, required to establish interphase arrays of cortical microtubules, GEM1</td>
</tr>
<tr>
<td>S_alt_contig00370_RC</td>
<td>At5g14620</td>
<td>chromatin modification, putative DNA methyltransferase, DMT7</td>
</tr>
<tr>
<td>S_5sp_contig22738</td>
<td>At2g36770</td>
<td>metabolic process, UDP-Glycosyltransferase, 216 connections</td>
</tr>
<tr>
<td>S_5sp_contig20309</td>
<td>At3g48750</td>
<td>regulates the mitosis-to-endocycle transition during leaf development, CDC2, 37 connections</td>
</tr>
<tr>
<td>S_5sp_contig40889</td>
<td>At1g11890</td>
<td>SEC22 Gene Family, ATSEC22, 35 connections</td>
</tr>
<tr>
<td>S_alt_contig13004_RC</td>
<td>At1g53750</td>
<td>RPT1A, 34 connections</td>
</tr>
</tbody>
</table>
Figure 2.5. Average fold change in expression relative to control (expressed as 1) in qPCR target genes
Figure 2.6. Normalized change in expression in microarray due to oil stress in target genes.
Chapter Three:

Oil responsive genes in S. alterniflora modulate phenotype in Brachypodium distachyon with applications in Spartina alterniflora

Abstract

Experiments in natural environments may reveal novel genetic patterns and functions that are otherwise cryptic in controlled settings. However, molecular ecology is often hampered by the high costs of developing resources for non-model species to confirm correlative data obtained in natural populations. To confirm the function of candidate genes involved in resilience to oil stress in the foundation salt marsh species Spartina alterniflora, we used T-DNA insertion genotypes of the emerging model grass species Brachypodium distachyon. We chose lines that are disrupted in one of eight genes whose expression were correlated to contamination by the BP Deepwater Horizon oil spill in natural populations of S. alterniflora. We identified four genes, which differed significantly from wild-type in their measured phenotypes, but we did not detect a response to oil treatment in either the wild type or any of the mutant lines.
Introduction

Molecular ecology is a burgeoning discipline that integrates molecular techniques with ecologically relevant systems and settings (Ungerer et al. 2008, Andrew et al. 2013). In particular, ecological genomics has been instrumental in testing and generating new hypotheses in landscape ecology, speciation, phylogeography, and adaptation at the level of the genome (Andrew et al. 2013). Of note are experiments in natural environments that have revealed novel genetic patterns and functions that are otherwise cryptic in controlled settings (Colbourne et al. 2011). However, molecular ecology is often hampered by the high costs of developing resources for non-model species, despite the often-rich history of ecological work in these important species (Alvarez et al. 2015). Therefore, a comparative genomics approach may provide an avenue to the early exploration of genome function without incurring substantial financial and resource investment. Well-studied model organisms, such as Arabidopsis thaliana, Drosophila melanogaster, and Mus musculus have high-resolution genomic maps, detailed genome annotations, comparatively well-understood metabolic pathways, and a number of analysis tools and databases (e.g AraCyc, Rhee et al. 2005; Virtual Plant, Katari et al. 2010) that are already being leveraged for ecological studies. For example, Kobayashi and colleagues (2013) used annotations from A. thaliana to identify differentially expressed flowering time genes in the tropical tree Shorea beccariana before a mass flowering event. In addition, transgenic lines of A. thaliana confirmed the function of two of these genes: a floral pathway integrator (SbFT), and a floral repressor (SbSVP). This study leveraged the genomic resources of a model
organism to conduct functional genomics in a non-model species (Kobayashi et al. 2013). Although ecological subjects of interest are often substantially diverged from model organisms, the genetic resources and annotations available in model organisms may provide a “first look” into genome function and a method to identify genes of interest in non-model species.

*Spartina alterniflora*, or smooth cordgrass, is a foundational species in the Gulf of Mexico salt marshes, providing refuge for invertebrates (Silliman and Bertness 2002), nurseries for birds and fish (Mendelssohn et al. 2012), and habitat-building ecosystem services (Pennings and Bertness 2001). *Spartina alterniflora* is also highly resilient, surviving and maintaining biomass in response to natural stressors such as the large salinity gradient of the salt marsh (Pennings and Bertness 2001), climatic stressors such as heat (Baisakh and Subudhi 2009, Subudhi and Baisakh 2011), and anthropogenic stressors such as crude oil (Gedan et al. 2009, Silliman et al. 2012).

Because of extensive extraction and refining of oil in the Gulf of Mexico, *S. alterniflora* populations are frequently exposed to crude oil (Gedan et al. 2009). The physiological results of crude oil stress to *S. alterniflora* are well characterized and include changes in biomass and photosynthetic rate (Lin and Mendelssohn 2012). However, despite several investigations, the genomic mechanisms of *S. alterniflora*’s resilience to stress, and particularly crude oil stress, remain understudied (but see Baisakh and Subudhi 2008, Baisakh et al. 2008, RamanaRao et al. 2012).

In 2010, the *Deepwater Horizon* oil spill released an estimated 4.9 million barrels of oil into the Gulf of Mexico (National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling 2011). This oil eventually made landfall in the estuarine
ecosystems of Louisiana, Mississippi, and other Gulf Coast states, and of the 1,773 kilometers of coastline oiled by the Deepwater Horizon disaster, an estimated 44.9% was salt marsh habitat (Michel et al. 2013). Despite the loss of some habitat to mortality and subsequent erosion, marshes dominated by Spartina alterniflora showed up to 100% recovery of above-ground biomass even to heavy doses of crude oil (Lin and Mendelssohn 2012, Silliman et al. 2012). Although tragic, the patchwork of salt marshes that were affected by Deepwater Horizon oil spill represents a unique natural experiment to understand the mechanisms of resilience to oil stress in S. alterniflora. To study the response of this foundation species to an infrequent but recurrent stress, we created a custom microarray built from EST assemblies from several Spartina species, and identified 3,622 differentially expressed genes between oil-affected and –unaffected populations. We used the 1,211 most highly differentially expressed (FDR Q<0.001) genes to build a gene network, and identify highly connected genes in the program Virtual Plant (Katari et al. 2010). We considered eight highly connected genes involved in biosynthesis, reproductive development, volatile production, and transcriptional regulation as targets likely to be involved in regulating the phenotypic response to oil stress. Ideally, we would confirm the function of individual genes through reverse genetic screens, but this approach is not feasible in S. alterniflora.

Despite being a relevant ecological system for studying oil resilience, we lack detailed genome annotation and dedicated analysis pipelines for S. alterniflora like those that have been constructed for model organisms. We do not yet have a complete reference genome sequenced for S. alterniflora, although a full transcriptome was recently published (de Carvalho et al. 2013). However, Brachypodium distachyon, a
short-statured grass, is an emerging model species to test hypotheses about the functional genomics of grasses, such as *S. alterniflora*. *Brachypodium distachyon* has a number of genomic resources and is similar in gene content to other members of the grass family (Brklajací et al. 2011). In addition, a collection of 23,000 T-DNA insertion lines has been developed for *B. distachyon*, including knockout and activation tagging (overexpression) lines (Bragg et al. 2012). *Brachypodium distachyon* is easily raised in space-limited environments, making it efficient to culture T-DNA insertion lines in high-density under controlled conditions. To confirm the function of oil resilience candidate genes in *S. alterniflora*, we used T-DNA insertion genotypes of *B. distachyon* that are disrupted in one of eight genes whose expression were correlated to the response to hydrocarbon stress in *S. alterniflora*. All eight *Brachypodium distachyon* T-DNA insertion lines, along with a wild-type control, were exposed to crude oil over several months to determine the phenotypic effects of these genes of interest. Because crude oil stress manifests in phenotype through a variety of traits, we expected that our epigenetic and regulatory gene targets would impact various aspects of phenotype in *B. distachyon*. Further, we expected that knockout and overexpression of some target genes would be involved in modifying the response of *B. distachyon* to oil stress, but that these interactions might be cryptic until exposed by oil stress.

**Methods**

**Generating gene targets**

We identified 3622 genes that were differentially expressed in response to exposure to the *DWH* oil spill alone across populations of *S. alterniflora* (FDR Q < 0.05;
Alvarez et al. unpub). From the 3622 genes exclusively responding to oil stress (and not also population differences), we segregated highly oil-responsive genes (FDR Q < 0.001) to reduce the number and complexity of the interactions in the network construction resulting in 1,561 genes. Of these, we used only the 1,211 genes with homologs in the model plant Arabidopsis thaliana to generate gene interaction networks made of highly differentially expressed genes (nodes) and interactions (edges). Edges were parameterized between nodes using previously generated data from A. thaliana. These data included information on micro RNA binding sites, protein-to-protein interactions, transcriptional regulation (which includes transcription factors, enhancers, and repressors), and transport interactions. Additional edges were created from computationally generated data from the metabolic interaction databases Aracyc and KEGG, and from previously published literature interactions. We visualized the resulting gene interaction network using Cytoscape and identified the number of connections that each node had using the built-in network analysis function in Virtual Plant. We chose genes that were highly connected, based on the number of edges in the resulting network, for further study, with the rationale that highly connected and highly interactive genes were more likely to be orchestrating the response to hydrocarbon stress. We also chose several differentially expressed genes from S. alterniflora that were likely to have broad regulatory effects during the response to crude oil (Table 3.1).

**Gene targets**

Eight highly connected genes, including genes involved in biosynthesis, reproductive development, volatile production, and transcriptional regulation (Table 3.1)
were identified as targets for follow up validation using *B. distachyon* T-DNA insertion genotypes. The T-DNA insertion lines show either overexpression or no expression of these highly connected genes, depending on the vector used, but impacts on phenotype have not been confirmed. These genes were identified in our survey of natural *S. alterniflora* that had survived contamination from the BP *Deepwater Horizon* oil spill, and are thus suspected to be involved in response to crude oil.

Some of the target genes may influence phenotype and oil response by altering the ability of plants to divert resources to germplasm investment, such as Bradi1g72150. Bradi1g72150 is homologous to At2G22540, or AGL22, which regulates floral transition (Li et al. 2008). *Arabidopsis thaliana* mutants that are deficient in AGL22 display an early-flowering phenotype, making AGL22 a repressor of floral development. Other genes may influence the production of seeds by altering the available resources of a plant, such as Bradi2g16710 or Bradi1g62540. Bradi2g16710, which is a homolog of At1g53000, or KDSB, is involved in pollen tube elongation (Delmas et al. 2008). In *A. thaliana*, KDSB-deficient mutants displayed abnormally large mitochondria (Duncan et al. 2011), which may alter the amount and distribution of energy to produce seeds. Bradi1g62540 is a homolog of ATTPS21, or At5g23960, which is a terpene synthase gene involved in the production of volatiles in flower petals in *A. thaliana* (Liu et al. 2015). ATTPS21 is also differentially expressed in response to herbivory (Broekgaarden et al. 2007), which may provide olfactory cues for pollinators (Chen et al. 2003). Although ATTPS21 function may differ in *B. distachyon* as compared to *A. thaliana*, variation in ATTPS21 expression may modulate volatile production, which may require additional resources that stunt *B. distachyon* growth while under stress.
Alternatively, genes such as Bradi3g08060 may exert more wide-ranging effects through the regulation of other loci and not participation in a particular pathway. Bradi3g08060, for example, corresponds to At4G38130, or HD1, a histone deacetylase that is expressed in a wide range of tissues and regulates gene expression (Kagale and Rozwadowski 2011) and seed maturation (Zhou et al. 2013). HD1 mutants produce variable expression of seed maturation genes, suggesting a role for HD1 in regulating embryogenesis. As an epigenetic regulator of gene expression, HD1 has also been shown to participate, primarily through transcriptional repression and regulation, in the organismal response to a number of stresses in in A. thaliana, including phosphate starvation (Chen et al. 2015), drought (Song et al. 2005), and pathogen defense (Choi et al. 2012). We predicted that the involvement of this gene in the regulation of stress response may be conserved in B. distachyon, and S. alterniflora. The specific function of each of these genes in the response to crude oil stress has not previously been unexplored.

Oil exposure experiment

We obtained T-DNA insertion genotypes from the Western Regional Resource Center (WRRC; Bragg et al. 2012), and stored seeds in ambient conditions before vernalizing them for two weeks. We then sowed two replicates of the eight T-DNA insertion lines, and a wild-type line (B21-3) in each of two oil treated trays, and two untreated trays, which were all grown in a single growth chamber. Because treatment is applied at the level of trays, this design is a split plot (Richards et al. 2008). Each treatment tray received 1000ul of 2.5% crude oil in tap water every other day, which is a
sub-lethal concentration that we found induced phenotypic response in *S. alterniflora*. Untreated trays received 1000ul of only tap water every other day. Seeds were grown for 72 days until a majority of plants had flowered and senesced. On the 72\textsuperscript{nd} day, we scored individuals as either dead (having no green tissue) or alive, and harvested seeds and above-ground biomass from each plant. After drying for 2 days at 60C, we weighed total above-ground biomass. Additionally, we weighed inflorescences separately to generate an inflorescence biomass to the total above-ground biomass ratio (which included the inflorescence, hereafter referred to as percent inflorescence), which may capture variation in growth strategy and resource allocation. Finally, we counted the total number of seeds per plant to quantify a measure of fitness.

**Analysis**

Our experiment was a split-plot design. When analyzed as an ANOVA, the model is: 

\[
\text{Response} \sim \text{Treatment} + \text{Genotype} + (\text{Treatment} \times \text{Genotype}) + \text{Error(Block*Treatment)},
\]

which treats block as a random effect. However, because we only had two blocks (trays), we used block as a fixed effect (Kéry *et al.* 2010). To understand effects of oil exposure on the phenotype of *B. distachyon*, we fit the modified split-plot as a linear model to each phenotype and T-DNA insertion line separately (biomass, inflorescence weight, percent inflorescence, and number of seeds). To compare each T-DNA insertion line to the wild-type line, we used the contrasts function to set wild-type data as the baseline 

\[
\text{Phenotype} \sim \text{Treatment*Genotype} + \text{Tray X Treatment},
\]

where Treatment*Genotype represent both the main effects and interactions of oil treatment and genotype (T-DNA insertion line).
All models were linear except in the case of seeds, which was fit as a generalized linear model with a poisson distribution, as the seeds represented count data.

Results

_T-DNA insertion genotypes differ from wild-type, but do not alter phenotypic response to oil stress_

We found a significant effect of T-DNA insertion in at least one phenotype for 4 out of 8 of our lines: Bradi3g08060, Bradi1g72150, Bradi3g35330, and Bradi1g68290 (Tables 2-5). Bradi3g08060 knockout lines showed increased seed production over wild-type strains, while Bradi1g72150 and Bradi3g35330 overexpression lines showed reductions in both inflorescence weight and percent inflorescence. Bradi1g68290 knockout lines showed reduction in inflorescence weight. However, in each model, we did not find significant effects of oil exposure (Tables 3.2-3.5, Figures 3.1-3.4).

Discussion

By using resources from model plants, like the T-DNA insertion genotypes of _B. distachyon_ and annotation data from _A. thaliana_, we were able to make inferences about genes that contribute to aspects of phenotype in the non-model species _S. alterniflora_. Across both treatment and control trays, we found significant differences in seed production between wild-type and the Bradi3g08060 (HD1 homolog) knockout lines (Table 3.3). HD1 is an epigenetic regulator of embryogenesis, and loss of function in this gene dramatically reduced the number of seeds produced. We also found decreased inflorescence mass and percent inflorescence in Bradi3g35330
overexpression lines, a SUVH5 homolog and another epigenetic regulator. SUVH5 regulates gene expression by modulating the transcription of genes through methylation in the CHG context, where H is any nucleotide (Stroud et al. 2015). In A. thaliana, SUVH5 is upregulated during sperm development, perhaps reflecting the complex and flexible role of DNA methylation in development (Borges et al. 2008).

We found significantly reduced inflorescence weight and percent inflorescence in Bradi1g72150 (AGL22 homolog) overexpression lines as compared to wild-type, potentially reflecting a reduced investment in reproduction versus growth. The floral transition pathway is a complex pathway with a number of gene interactions (Balanza et al. 2014), and AGL22 may affect phenotype via epistatic effects beyond its immediate gene product. This speculation is supported by transcriptome profiling of S. alterniflora which shows a number of highly differentially expressed genes with homology to genes in the floral transition pathway, such as AGL16 and AGL8 (Alvarez et al. unpublished), which are known to interact with AGL22 (Balanza et al. 2014). Thus, variation in the size of each inflorescence relative to the total biomass in knockouts may reflect a reduced ability to regulate the floral development pathway and reapportion resources toward the production of inflorescences. We also found a significant effect of Bradi1g68290 overexpression on inflorescence weight. Bradi1g68290, which is a homolog of ATMCB1 in A. thaliana, is a proteasome involved in the degradation of a number of proteins in the A. thaliana (Jain et al. 2008). ATMCB1 plays a role in oxidative stress tolerance, and loss of ATMCB1 function causes modulations in cell size and number in shoots, and the gene also plays a role in (Kurepa et al. 2007, Kurepa et al. 2009). ATMCB1 is also
differentially expressed during pollen tube development, suggesting a role in reproduction (Wang et al. 2008).

In the remaining four T-DNA insertion lines, we did not find significant phenotypic divergence from wild type. While this may reflect a lack of power within our design to detect phenotypic variation between wild-type and T-DNA insertion lines, work in \textit{Saccharomyces cerevisiae} has shown that many genes can be modulated or knocked out without any phenotypic effect (reviewed in Giaever and Nislow 2014). This may be due, in part, to the epistatic effects of other genes that buffer the overexpression or underexpression of the target gene (Segre et al. 2004). Further, the plants did not respond to our treatment, and the importance of these genes may be elicited in a more effective oil dosage. However, the significant variation between four of our eight T-DNA lines and wild-type confirmed the effect of these genes on the phenotypes measured.

All T-DNA insertion lines were chosen because they possessed modified genes that were homologous to genes that were differentially expressed in oil-impacted \textit{S. alterniflora} populations. Using annotation data from \textit{A. thaliana}, we parameterized functional networks of genes to identify highly connected genes, which may be orchestrating the phenotypic response to oil stress. In each phenotype, we found no significant effect of either oil exposure or the interaction of oil exposure and line. The lack of oil-induced effect may be because of a lack of power in our study, but these grasses may also be resilient to oil. In \textit{S. alterniflora}, individuals may recover up to 100% over 7 months (Lin and Mendelssohn 2012), and our lengthy oil exposure may have given time for affected individuals to recover. However, it is important to note that while \textit{B. distachyon} and \textit{S. alterniflora} are in the same family (Poaceae), it is one of the
largest plant families, and there is substantial divergence time between the two species (Grass Phylogeny Working Group 2011). As mutations accumulate over evolutionary time, selection and drift alter the function of genes. Additionally, gene networks may themselves be the targets of selection and drift, creating novel function and cryptic variation even when sequencing divergence in a single component is minor (Cork and Purugganan 2004). This problem may be particularly acute in the hexaploid S. alterniflora, as genome duplication and copy number variation can complicate our understanding of the role of specific genes and increase the proportion of false positive matches (Primmer et al. 2013). Divergent functionality may explain the lack of treatment effect that we observed in our experiment, and future studies with either larger sample sizes or a more closely related study species could alleviate these problems. Despite these limitations to inference, T-DNA insertion lines still proved to be a useful method in annotation the gene function of homologs in S. alterniflora, even without a phenotypic response to crude oil stress.

Rapidly falling sequencing costs, new publically available genomes, and new informatics software have contributed to the democratization of molecular biology, giving ecologists access to understanding the molecular pattern and process that underlies environmental interactions (Ekblom and Galindo 2011, Alvarez et al. 2015). In turn, application of these technologies by ecologists can provide valuable annotation of gene function in heterogeneous environmental conditions (Pavey et al. 2012). These data benefit molecular biologists working in model organisms and evolutionary biologists attempting to describe the action of genes on phenotype, since genomes evolved in real complex environments and gene function is often hidden in controlled, standardized
environments (Colbourne et al. 2011). Gene function is particularly conserved in slower-evolving essential genes and regions, as highly connected genes commonly are (Lee et al. 2008). Although molecular ecology is rapidly accelerating (Alvarez et al. 2015), the lack of other model resources, such as high-quality genomes, inbred lines, and mutant lines, has limited the ability to perform functional genomics in ecologically relevant settings. This hurdle is currently a major limitation in making inferences about the function of genetic variation in non-model systems (Pavey et al. 2012, Andrew et al. 2013). Despite the lack of treatment effect observed in this study, linking the ecology of a foundation species like S. alterniflora with the genetic resources of B. distachyon and A. thaliana, may still be a useful methodology to validate ecological and molecular annotations in genes of interest. Although we were not able to validate the ecological annotations of S. alterniflora genes, we found 4 genes that affected phenotype when disrupted by T-DNA insertion, providing useful confirmation of molecular annotation and functional divergence. More relevant treatment conditions and more powerful designs in follow-up experiments may expose variation that underlies the response to crude oil stress.

The expanded use of molecular tools in ecology allows for examination of phenotype and function in ecologically relevant settings (Alvarez et al. 2015). In particular, large-scale events like oil spills are complex and occur on a scale that is difficult to replicate. Oil spills are frequent yet unpredictable stressors that threaten coastal ecosystems both because of their toxicity (Lin and Mendelssohn 2012) and because of the difficulty in containing them (Lin and Mendelssohn 1998, Silliman et al. 2012). Both the chemical effects, such as polyaromatic hydrocarbon (PAH) toxicity, and
the physical effects, such as coating the foliage, mediate the stress potential of crude oil (Pezeshki et al. 2000). Through these mechanisms, crude oil imposes a complex immediate and long-term stress to the organisms with which it comes in contact (reviewed in Pezeshki et al. 2000). The 2010 Deepwater Horizon oil spill caused extensive die-back and sloughing of S. alterniflora, which may cause long-term disturbance to both S. alterniflora populations and the greater salt marsh community (Silliman et al. 2012). Thus, further dissection of the molecular mechanisms of oil response in S. alterniflora may allow us to make inferences about the ecological consequences of stress response and resilience. Additionally, an understanding of functional genomic mechanisms of oil response in S. alterniflora and other grasses would provide a useful marker for conservation, restoration, and cleanup efforts in the Gulf of Mexico.

References


McCall, BD & Pennings, SC. Disturbance and recovery of salt marsh arthropod communities following BP Deepwater Horizon oil spill. PloS one (2012).


Zhou, Y et al. HISTONE DEACETYLASE19 interacts with HSL1 and participates in the repression of seed maturation genes in Arabidopsis seedlings. The Plant... (2013). doi:10.1105/tpc.112.096313
Figures and Tables

Table 3.1. T-DNA insertion lines are shown with their corresponding *S. alterniflora* contigs from a previously constructed microarray (Alvarez et al., unpublished). These contigs were converted to their closest *Arabidopsis thaliana* homolog, which is listed with its TAIR number as well as its alternate common name. T-DNA lines represent the catalog numbers from the WRRC. Modification type represents the effect of the T-DNA insertion, either knocking out the function of the particular gene or tagging the promoter region to induce overexpression. The function of each gene, as described by TAIR, is listed along with the justification for choosing each.

<table>
<thead>
<tr>
<th>contig</th>
<th>At annotation</th>
<th>Function</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_alt_contig08070</td>
<td>AT4G38130</td>
<td>negative regulation of transcription, histone deacetylase 1</td>
<td>Epigenetic gene of interest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>negative regulation of transcription, K-box region and MADS-box transcription factor</td>
<td>Regulative gene of interest</td>
</tr>
<tr>
<td>S_alt_contig03288</td>
<td>AT2G22540</td>
<td>metabolic process, involved in the biosynthesis of VLCFA</td>
<td>Highly connected gene in overall network (107 connections)</td>
</tr>
<tr>
<td>S_5sp_contig45211</td>
<td>At2g26640</td>
<td>RNA methylation, ribosome biogenesis, translation</td>
<td>Highly connected gene in overall network (64 connections), ribosomal gene of interest</td>
</tr>
<tr>
<td>S_alt_contig12988</td>
<td>At3g62870</td>
<td>terpene synthase</td>
<td>Highly connected gene in overall network (44 connections)</td>
</tr>
<tr>
<td>S_alt_contig04438_RC</td>
<td>At5g23960</td>
<td>negative regulation of transcription, histone modification, methylation, SU(VAR)3-9 homolog 5 regulatory particle non-ATPase subunit of the 26S proteasome</td>
<td>Epigenetic gene of interest</td>
</tr>
<tr>
<td>S_mar_contig02490_RC</td>
<td>AT2G35160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S_alt_contig08149</td>
<td>At4g38630</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3.1 (Continued)

<table>
<thead>
<tr>
<th>contig</th>
<th>At annotation</th>
<th>Function</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>S_alt_contig03666_RC</td>
<td>At1g53000</td>
<td>cellular polysaccharide biosynthetic process, pollen tube development, activating KDO as a nucleotide sugar</td>
<td>Locally connected gene (10 connections) involved in reproductive development</td>
</tr>
</tbody>
</table>
### Table 3.2. Linear model for biomass.

| Biomass                        | Estimate   | Std. Error  | t value | Pr(>|t|) |
|-------------------------------|------------|-------------|---------|----------|
| (Intercept)                   | 0.1768082  | 0.0923822   | 1.914   | 0.0617   |
| TreatmentOil                  | -0.1359348 | 0.1043893   | -1.302  | 0.1992   |
| Bradi3g08060                  | -0.0299255 | 0.0644884   | -0.464  | 0.6448   |
| Bradi1g72150                  | -0.0092255 | 0.0644884   | -0.143  | 0.8869   |
| Bradi3g07730                  | -0.0282255 | 0.0644884   | -0.438  | 0.6636   |
| Bradi5g24680                  | -0.0149255 | 0.0644884   | -0.231  | 0.818    |
| Bradi1g62540                  | -0.0378255 | 0.0770113   | -0.491  | 0.6256   |
| Bradi3g35330                  | -0.0054755 | 0.0644884   | -0.085  | 0.9327   |
| Bradi1g68290                  | 0.1289745  | 0.0644884   | 2       | 0.0513   |
| Bradi2g16710                  | -0.0296005 | 0.0644884   | -0.459  | 0.6483   |
| TreatmentOil X Bradi3g08060   | 0.0173908  | 0.0911977   | 0.191   | 0.8496   |
| TreatmentOil X Bradi1g72150   | 0.0135505  | 0.0877652   | 0.154   | 0.878    |
| TreatmentOil X Bradi3g07730   | 0.0188255  | 0.0877652   | 0.214   | 0.8311   |
| TreatmentOil X Bradi5g24680   | 0.0147005  | 0.0877652   | 0.167   | 0.8677   |
| TreatmentOil X Bradi1g62540   | 0.0224755  | 0.0973381   | 0.231   | 0.8184   |
| TreatmentOil X Bradi3g35330   | -0.0011745 | 0.0877652   | -0.013  | 0.9894   |
| TreatmentOil X Bradi1g68290   | -0.1465245 | 0.0877652   | -1.67   | 0.1017   |
| TreatmentOil X Bradi2g16710   | 0.0171018  | 0.0911977   | 0.188   | 0.8521   |
| TreatmentControl X Tray       | -0.0421531 | 0.0294604   | -1.431  | 0.1591   |
| TreatmentOil X Tray           | -0.0009693 | 0.0097214   | -0.1    | 0.921    |

F-statistic: 0.9124 on 19 and 47 DF, p-value: 0.571
Table 3.3. Linear model for inflorescence weight.

| Inflorescence Weight | F-statistic; 3.148 on 19 and 47 DF, p-value: 0.0007159 | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------------|------------------------------------------------------|----------|------------|---------|---------|
| (Intercept)          |                                                      | 0.0191755| 0.006858   | 2.796   | 0.00747 ** |
| TreatmentOil         |                                                      | -0.0085938| 0.0077494 | -1.109  | 0.27309 |
| Bradi3g08060         |                                                      | -0.0023735| 0.0047873 | -0.496  | 0.62236 |
| Bradi1g72150         |                                                      | -0.0136485| 0.0047873 | -2.851  | 0.00645 ** |
| Bradi3g07730         |                                                      | -0.0060985| 0.0047873 | -1.274  | 0.20897 |
| Bradi5g24680         |                                                      | -0.0037235| 0.0047873 | -0.778  | 0.4406  |
| Bradi1g62540         |                                                      | -0.0101235| 0.005717  | -1.771  | 0.08308 . |
| Bradi3g35330         |                                                      | -0.0113235| 0.0047873 | -2.365  | 0.02219 * |
| Bradi1g68290         |                                                      | -0.0126735| 0.0047873 | -2.647  | 0.01101 * |
| Bradi2g16710         |                                                      | -0.0086485| 0.0047873 | -1.807  | 0.07724 . |
| TreatmentOil X Bradi3g08060 |                                          | 0.0046838| 0.0067701 | 0.692   | 0.49244 |
| TreatmentOil X Bradi1g72150 |                                          | 0.0054485| 0.0065153 | 0.836   | 0.40724 |
| TreatmentOil X Bradi3g07730 |                                          | 0.0069485| 0.0065153 | 1.066   | 0.29165 |
| TreatmentOil X Bradi5g24680 |                                          | 0.0051735| 0.0065153 | 0.794   | 0.43116 |
| TreatmentOil X Bradi1g62540 |                                          | 0.0097735| 0.007229 | 1.353   | 0.18267 |
| TreatmentOil X Bradi3g35330 |                                          | 0.0067235| 0.0065153 | 1.032   | 0.30738 |
| TreatmentOil X Bradi1g68290 |                                          | 0.0111985| 0.0065153 | 1.719   | 0.09223 . |
| TreatmentOil X Bradi2g16710 |                                          | 0.0071548| 0.0067701 | 1.057   | 0.29599 |
| TreatmentControl X Tray |                                          | 0.0017592| 0.002187  | 0.804   | 0.42523 |
| TreatmentOil X Tray   |                                                      | -0.0007627| 0.0007217 | -1.057  | 0.296   |
Table 3.4. Linear model for percent inflorescence. * represents P<0.05, ** represents P<0.01, and *** represents P<0.001.

| Percent Inflorescence | Estimate | Std. Error | t value | Pr(>|t|) |
|-----------------------|----------|------------|---------|----------|
| (Intercept)           | 0.245086 | 0.119084   | 2.058   | 0.0452 * |
| TreatmentOil          | 0.010077 | 0.134562   | 0.075   | 0.9406   |
| Bradi3g08060          | 0.153642 | 0.083128   | 1.848   | 0.0709   |
| Bradi1g72150          | -0.205811| 0.083128   | -2.476  | 0.017 *  |
| Bradi3g07730          | 0.036801 | 0.083128   | 0.443   | 0.66     |
| Bradi5g24680          | -0.011638| 0.083128   | -0.14   | 0.8893   |
| Bradi1g62540          | 0.036644 | 0.09927    | 0.369   | 0.7137   |
| Bradi3g35330          | -0.189942| 0.083128   | -2.285  | 0.0269 * |
| Bradi1g68290          | -0.058477| 0.083128   | -0.703  | 0.4852   |
| Bradi2g16710          | 0.000945 | 0.083128   | 0.011   | 0.991    |
| TreatmentOil X Bradi3g08060 | 0.05249 | 0.117557   | 0.447   | 0.6573   |
| TreatmentOil X Bradi1g72150 | 0.001491 | 0.113132   | 0.013   | 0.9895   |
| TreatmentOil X Bradi3g07730 | 0.087777 | 0.113132   | 0.776   | 0.4417   |
| TreatmentOil X Bradi5g24680 | 0.064192 | 0.113132   | 0.567   | 0.5731   |
| TreatmentOil X Bradi1g62540 | 0.114203 | 0.125472   | 0.91    | 0.3674   |
| TreatmentOil X Bradi3g35330 | 0.077831 | 0.113132   | 0.688   | 0.4949   |
| TreatmentOil X Bradi1g68290 | 0.187237 | 0.113132   | 1.655   | 0.1046   |
| TreatmentOil X Bradi2g16710 | 0.080948 | 0.117557   | 0.689   | 0.4945   |
| TreatmentControl X Tray | 0.046963 | 0.037976   | 1.237   | 0.2224   |
| TreatmentOil X Tray   | -0.016786| 0.012531   | -1.339  | 0.1869   |
Table 3.5. Generalized linear model for seeds, using Poisson distribution. * represents $P<0.05$, ** represents $P<0.01$, and *** represents $P<0.001$.

<table>
<thead>
<tr>
<th>Seeds</th>
<th>AIC: 406.57</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>2.42044</td>
</tr>
<tr>
<td>TreatmentOil</td>
<td>-0.5734</td>
</tr>
<tr>
<td>Bradi3g08060</td>
<td>1.1382</td>
</tr>
<tr>
<td>Bradi1g72150</td>
<td>-0.56655</td>
</tr>
<tr>
<td>Bradi3g07730</td>
<td>-0.00693</td>
</tr>
<tr>
<td>Bradi5g24680</td>
<td>0.10128</td>
</tr>
<tr>
<td>Bradi1g62540</td>
<td>-0.09654</td>
</tr>
<tr>
<td>Bradi3g35330</td>
<td>-0.26644</td>
</tr>
<tr>
<td>Bradi1g68290</td>
<td>-0.16108</td>
</tr>
<tr>
<td>Bradi2g16710</td>
<td>0.10128</td>
</tr>
<tr>
<td>TreatmentOil X Bradi3g08060</td>
<td>0.18309</td>
</tr>
<tr>
<td>TreatmentOil X Bradi1g72150</td>
<td>-0.81975</td>
</tr>
<tr>
<td>TreatmentOil X Bradi3g07730</td>
<td>0.70008</td>
</tr>
<tr>
<td>TreatmentOil X Bradi5g24680</td>
<td>0.16108</td>
</tr>
<tr>
<td>TreatmentOil X Bradi1g62540</td>
<td>0.2363</td>
</tr>
<tr>
<td>TreatmentOil X Bradi3g35330</td>
<td>-0.16434</td>
</tr>
<tr>
<td>TreatmentOil X Bradi1g68290</td>
<td>0.30084</td>
</tr>
<tr>
<td>TreatmentOil X Bradi2g16710</td>
<td>0.33234</td>
</tr>
<tr>
<td>TreatmentControl X Tray</td>
<td>-0.09826</td>
</tr>
<tr>
<td>TreatmentOil X Tray</td>
<td>-0.09948</td>
</tr>
</tbody>
</table>
Figure 3.1. Means of percent inflorescence for each T-DNA knockout line in oil treatment and control. Error bars represent standard error.
Figure 3.2. Means of seed production for each T-DNA knockout line in oil treatment and control. Error bars represent standard error.
Figure 3.3. Means of inflorescence weight for each T-DNA knockout line in oil treatment and control. Error bars represent standard error.
Figure 3.4. Means of total biomass for each T-DNA knockout line in oil treatment and control. Error bars represent standard error.
Chapter Four:

Oil exposure induced effects in S. alterniflora, but not in a genotype-specific manner

Abstract

Although oil spills can be severe and unpredictable threats to coastal ecosystems, the foundational species Spartina alterniflora is exceptionally resilient to crude oil stress. This resilience may be the product of genotypic differentiation, phenotypic plasticity, or a combination of both. To identify the relative contributions of genotypic differentiation or phenotypic plasticity and to determine the extent of standing variation for oil response in S. alterniflora, we exposed oil-naïve populations of S. alterniflora gathered from Sapelo Island, GA to crude oil stress in a greenhouse experiment. In our experiment, we found main effects of oil and genotype on the number of leaves and leaves per ramets across the experiment. We also found no evidence of a genotype-by-environment or differences in mortality between genotypes. These results suggest that while there is variation amongst genotypes, there is no standing population-level variation in ability to respond to oil, creating the potential for selection on genotypes but not on plastic response to oil.
Introduction

Organisms living across broad environmental ranges must orchestrate a variety of processes on multiple biological levels to maintain homeostasis. These processes are often thought to culminate in either local adaptation (Clausen et al. 1948) or phenotypic plasticity, defined as ability of a single genotype to produce multiple phenotypes (reviewed in Pigliucci 2001, West-Eberhard 2003). However, these two processes are not mutually exclusive, as locally adapted species generally show some plasticity in traits. Additionally, plasticity that allows organisms to persist across broad ranges may be adaptive and under positive selection (Pigliucci et al. 2001). To understand the relative contribution of these processes in generating phenotypes of interests, researchers may rely on organisms that display a wide range of phenotypes, such as salt marsh organisms (Richards et al. 2005). In particular, salt marsh plants display unique intraspecific and community level patterns due to tidal and rainfall influences that create wide salinity gradients (Callaway et al. 1990, Pennings and Bertness 2001, Richards et al. 2005). As the interface between ocean and land, salt marshes experience frequent and often-unpredictable anthropogenic impacts, including land development, oil spills, and climate change effects (Kennish 2001, RamanaRao et al. 2012). Several salt marsh plant species display phenotypic variation that is correlated to these conditions, such as in the foundation species Spartina alterniflora, providing an excellent opportunity for studying the relative contribution of adaptation and plasticity in response to environmental stressors (Richards et al. 2005, 2010).

*Spartina alterniflora* is a clonally reproducing halophyte that is native to the salt marshes of the eastern United States (Pennings and Bertness 2001) and invasive
around the world (Callaway and Josselyn, 1992, An et al. 2007, Ainouche et al. 2009). *Spartina alterniflora* comprises up to 90% of the biomass in native habitats along the east coast of the United States, providing refuge for invertebrates (Silliman and Bertness 2002), nurseries for birds and fish (Mendelssohn et al. 2012), and habitat-building services (Pennings and Bertness 2001). In addition to tolerating natural environmental gradients, *Spartina alterniflora*-dominated salt marshes are also remarkably resilient to anthropogenic impacts, despite unpredictable disturbance and stress from exploitation, development, and pollutant releases like crude oil spills (Kennish 2001, Gedan et al. 2009, Silliman et al. 2012). Although it was thought that *S. alterniflora* tolerated the wide range of environmental conditions through large clones, *S. alterniflora* has high genetic diversity that is typical of outcrossing grasses (Richards et al. 2004, Foust et al. 2016). Despite this diversity, we found no consistent association of genotypes or alleles at specific loci with habitat (Richards et al. 2004).

Oil spills threaten coastal ecosystems both because of their toxicity (Pezeshki et al. 2000, Lin and Mendelssohn 2012) and because of the difficulty in containing them (Lin and Mendelssohn 1998, Silliman et al. 2012). Both the chemical effects, such as polyaromatic hydrocarbon (PAH) toxicity, and the physical effects, such as coating the foliage, contribute to the stress potential of crude oil, imposing both an immediate and long-term stress on the organisms with which it comes in contact (reviewed in Pezeshki et al. 2000). For example, in 2010, the *Deepwater Horizon* oil spill released an estimated 4.9 million barrels of oil into the Gulf of Mexico (National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling 2011), impacting the shorelines of Louisiana, Mississippi, and Alabama (Thibodeaux et al. 2011, Mendelssohn et al. 2012).
Many S. alterniflora populations across the northern Gulf of Mexico were heavily affected by the Deepwater Horizon oil spill, during which heavy deposits of petroleum hydrocarbons caused extensive loss of above-ground biomass and habitat loss (Lin and Mendelssohn 2012, Silliman et al. 2012). Despite these severe impacts, Gulf Coast S. alterniflora populations have shown up to 100% recovery after 7 months and moderate recovery in as little as 2 months (Lin and Mendelssohn 2012). The physiological results of hydrocarbon stress in S. alterniflora are well characterized and include reduction in carbon fixation, transpiration, and aboveground biomass (Lin and Mendelssohn 2012, Silliman et al. 2012). However, our understanding of crude oil stress response in S. alterniflora comes from population-scale studies in natural conditions, which have not evaluated the importance of genotypic variation within populations.

To identify whether resilience to oil stress could be the result of selection of tolerant genotypes or general resilience of all genotypes through phenotypic plasticity, we exposed replicates of genotypes from oil-naïve populations of S. alterniflora gathered from Sapelo Island, GA to crude oil stress in a greenhouse experiment. Spartina alterniflora does not show genotypic differentiation that is correlated to microhabitat in the salt marsh, suggesting that the response to environment is instead mediated by phenotypic plasticity (Richards et al. 2004). We expected that S. alterniflora would also respond to crude oil stress through this same mechanism, suggested by the lack of significant genotype-by-environment interactions.
Methods

In May of 2010, we collected live *S. alterniflora* from the mid marsh of one oil-naïve population in the Sapelo Island National Estuarine Research Reserve in Georgia, USA. These individuals were spaced ten meters apart, maximizing the chance that individuals were of different genotypes. We acclimated these individuals to greenhouse conditions for a minimum of 3 years before beginning our experiments. We used rhizome cutting to generate 6 replicates each of ten *S. alterniflora* genotypes.

We distributed 3 replicates of each of the 10 genotypes in each of two tanks, for a total of 60 biological samples. One tank was filled with uncontaminated water, while the oil treatment tank was filled with 2.5% oil in 62 liters of water, and tides were simulated once per day by filling containment chambers with the water or water-oil mixture and allowing the fluid to drain into a catchment. We measured the number of living leaves and the number of living ramets when the experiment began, and again 7 days after crude oil was added. We inferred that plants with no living above-ground biomass were dead. To understand the effects of treatment and genotype on phenotype, we used analyses of variance (ANOVA) in R to analyze three phenotypes: the change in the number of leaves, ramets, and leaves per ramet over the course of the experiment. Each model was written as (Phenotype ~ Treatment + Genotype + Treatment*Genotype). Mortality data was analyzed via the same model by coding mortality data as 0 and 1 for dead and alive at the end of the experiment, respectively.
Results

We found a significant effect of oil exposure on both the number of leaves and the number of leaves per ramet, but not on the number of ramets (P<0.05, Table 4.1, Figures 4.1-4.2). We additionally found a significant effect of genotype on the number of leaves per ramet produced during short-term oil exposure (Table 4.1, Figure 4.3). Although we observed some mortality, we did not find a significant effect of genotype or treatment on mortality (Figure 4.4). For the three phenotypes and mortality, we did not find a significant genotype by environment (GxE) interaction in response (Table 4.1).

Discussion

Crude oil is a common anthropogenic stressor in *S. alterniflora*-dominated salt marshes, primarily due to exploitation of near shore and off shore drilling locations in close proximity to coastal salt marshes (Gedan *et al.* 2009). Crude oil stress is particularly relevant to *S. alterniflora* populations in the Gulf of Mexico, which is a site of frequent oil drilling and hydrocarbon release. However, *S. alterniflora* populations are highly resilient to oil stress, despite deleterious phenotypic effects (Lin and Mendelssohn 2012, RamanaRao *et al.* 2012, Silliman *et al.* 2012).

Similar to previous studies (Lin and Mendelssohn 2012, Silliman *et al.* 2012), we found substantial reduction in both the number of leaves and the number of leaves per ramet in response to crude oil exposure. Lin and Mendelssohn (2012) additionally found changes in photosynthetic rate, which we did not quantify in this experiment. Other studies reported minimal mortality in response to treatment with oil, but these studies have primarily been conducted using populations native to Louisiana (Lin and
Mendelssohn 2012, Silliman et al. 2012, reviewed in Pezeshki et al. 2000), where salt marshes are frequently exposed to crude oil through the exploitation and processing of crude oil in the area (Ko and Day 2004). We also found low mortality in our experiment: only 8 individuals out of 60 total experienced mortality, distributed evenly across treatments and genotypes. Since all of our genotypes originated from Sapelo Island, GA, which are not known to have been exposed to oil, our data suggest that response to crude oil stress may be a species-wide effect, and not exclusive to Gulf Coast populations of *S. alterniflora*. *Spartina alterniflora* shows resilience to other long-term complex natural stressors, such as nutrient and salinity stress (Pennings and Bertness 2001, Richards et al. 2004, 2005), and common mechanisms of resilience may be shared across the different stresses (Richards et al. 2012).

We also found a significant effect of genotype for the change in the number of leaves over the course of the experiment. These results suggest that there is standing variation amongst *S. alterniflora* genotypes in the traits measured. However, we did not observe an interaction of genotype and oil exposure, indicating that there is no variation in the ability to respond to oil stress for selection to act on (Pigliucci 2005). Previous studies have suggested that *S. alterniflora*, as well as other salt marsh plant species, may respond to salt stress through plasticity of ecologically relevant traits (Richards et al. 2005, Richards et al. 2010). We did not observe evidence for genotype-by-environment interactions, and thus variation in plasticity, in our study; however, we also observed high variance within genotypes, which may reduce our power to observe genotype and genotype-by-environment interactions.
As a foundation species comprising the majority of the biomass in native salt marshes, *S. alterniflora* plays a vital role in mediating the resilience of the salt marsh to continued anthropogenic impact, through oil spills or other exploitation (Pennings and Bertness 2001, Silliman et al. 2012). Although our data suggest that oil response and resilience could be species-wide traits of *S. alterniflora*, future studies should expand the number of phenotypes examined to resolve the contribution of individual variation to oil stress response. Comparative studies using other members of the *Spartina* genus may also help expose the evolutionary mechanisms behind oil stress resilience by comparing levels of plasticity and individual variation in oil stress response. Understanding the mechanisms of its remarkable resilience is vital to understanding the evolutionary fate of oil-exposed populations of native *S. alterniflora* in the salt marshes of the Gulf of Mexico.

**References**


Clausen, Jens, David Daniels Keck, and W. M. Hiesey. "Experimental studies on the nature of species. III. Environresponses of climatic races of Achillea."


## Tables and Figures

**Table 4.1.** ANOVA results of all phenotypes. * represents $P<0.05$, ** represents $P<0.01$, and *** represents $P<0.001$.

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum of Squares</th>
<th>Mean Squares</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Leaves per Ramet</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>54.29</td>
<td>54.29</td>
<td>13.475</td>
<td>0.00067**</td>
</tr>
<tr>
<td>Genotype</td>
<td>8</td>
<td>32.76</td>
<td>4.09</td>
<td>1.016</td>
<td>0.43876</td>
</tr>
<tr>
<td>Treatment:Genotype</td>
<td>8</td>
<td>29.44</td>
<td>3.68</td>
<td>0.914</td>
<td>0.51483</td>
</tr>
<tr>
<td>Residuals</td>
<td>42</td>
<td>169.21</td>
<td>4.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Leaves</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>303.75</td>
<td>303.75</td>
<td>24.369$E$-01</td>
<td>1.31E-05**</td>
</tr>
<tr>
<td>Genotype</td>
<td>8</td>
<td>280.6</td>
<td>35.075</td>
<td>2.814</td>
<td>0.01361*</td>
</tr>
<tr>
<td>Treatment:Genotype</td>
<td>8</td>
<td>124.33</td>
<td>15.542</td>
<td>1.2469</td>
<td>0.29681</td>
</tr>
<tr>
<td>Residuals</td>
<td>42</td>
<td>523.5</td>
<td>12.464</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ramets</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>1.6667</td>
<td>1.66667</td>
<td>2.6582</td>
<td>0.1105</td>
</tr>
<tr>
<td>Genotype</td>
<td>8</td>
<td>7</td>
<td>0.875</td>
<td>1.3956</td>
<td>0.2266</td>
</tr>
<tr>
<td>Treatment:Genotype</td>
<td>8</td>
<td>7.3333</td>
<td>0.91667</td>
<td>1.462</td>
<td>0.2002</td>
</tr>
<tr>
<td>Residuals</td>
<td>42</td>
<td>26.3333</td>
<td>0.62698</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mortality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>0.0167</td>
<td>0.01666</td>
<td>0.168</td>
<td>0.684</td>
</tr>
<tr>
<td>Genotype</td>
<td>8</td>
<td>0.9333</td>
<td>0.11666</td>
<td>1.176</td>
<td>0.3361</td>
</tr>
<tr>
<td>Treatment:Genotype</td>
<td>8</td>
<td>1.0667</td>
<td>0.13333</td>
<td>1.344</td>
<td>0.2492</td>
</tr>
<tr>
<td>Residuals</td>
<td>42</td>
<td>4.1667</td>
<td>0.09920</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.1. Changes in the number of leaves of the course of the experiment.
Figure 4.2. Changes in the number of ramets over the course of the experiment.
Figure 4.3. Changes in the number of leaves per ramet over the course of the experiment.
Figure 4.4. Mortality over the course of the experiment.
Appendix 1
Primary Problems In Transcriptome Studies

Bias in Signals

DNA microarray and RNAseq data each display biases and distortions at different ends of the gene expression spectrum. RNAseq is biased toward highly transcribed genes (Łabaj et al. 2011; Malone & Oliver 2011). The genes that are more highly transcribed have more abundant transcripts and are more likely to be sequenced, leaving less highly transcribed genes with comparatively less sequencing coverage (Łabaj et al. 2011), which potentially reduces the resolution of more subtle patterns of gene expression. Rather than directly counting transcripts, microarrays depend on fluorescently labeled targets that hybridize to probes. Each gene has a measurement device (the probes) that saturates at high expression levels, but the probes will detect genes that are expressed at lower levels. Thus, microarrays may be a more appropriate choice for the detection of expression variation in low-abundance genes (Łabaj et al. 2011; Malone & Oliver 2011). However, due to the fluorescence-based quantification method, microarrays experience some compression at the higher end of expression. This reduces the ability of microarrays to quantify very highly expressed genes (Malone & Oliver 2011). A user should carefully consider which bias is more tolerable for the ecological question.
Heterologous Arrays

This problem only relates to microarray studies where microarrays of closely related species are used to characterize gene expression in a species with no genomic resources (so-called ‘heterologous arrays’). While heterologous arrays can be useful, they must be used with caution because of mis-hybridization between probes designed for one species and RNA extracted from a different species (Buckley 2007). The problem of probe mis-hybridization was made famous by a comparison that found that gene expression differences in human brains were much greater than those in any chimpanzee tissues (Enard et al. 2002). However, Hsieh, et al. (2003) showed that the use of short-oligonucleotide microarrays biased the results because some of the probes did not hybridize efficiently to the chimpanzee cDNA (Buckley 2007). A reanalysis with long oligonucleotide arrays, which are less sensitive to polymorphisms, revealed that the patterns in expression from the brain tissue were actually less divergent (Hsieh et al. 2003). In addition to highlighting some of the potential problems with microarray technology, this study provided a first glance at the complexity involved in studying divergence between two closely related species.

Polyploidy

Polyploid organisms may express many duplicate genes, and RNAseq-based transcriptomes that are assembled de novo (as non-model organisms often are) may align transcripts from different gene copies that have different function (Ilut et al. 2012). Further, when a polyploid organism is compared to a diploid reference genome, transcripts from duplicated genes may confound the relative expression of those genes.
Both of these issues may cause errors when inferring gene expression. Because RNAseq data provides information about polymorphisms, newer bioinformatics pipelines for sequence data, such as PolyCat (Page et al. 2013) and HomeoRoq (Akama et al. 2014), may alleviate these issues. PolyCat, developed for cotton, uses SNP information from related diploid species to accurately map sequencing reads from coresident genomes of allopolyploids (Page et al. 2013). HomeoRoq was developed specifically for RNAseq data and uses parental genomes to identify the ratio of gene expression from coresident allopolyploid genomes (Akama et al. 2014). These methods are some of the first attempts to decipher the relative contributions of duplicate genes in allopolyploids using genome-wide data. Because microarrays rely on hybridization and not direct sequencing, they provide a biologically relevant readout of the amount of gene transcript regardless of how many copies are contributing (with carefully designed probes). However, because similar sequences may co-hybridize with the same probe, microarrays are unable to discriminate between duplicated or highly similar genes, and they cannot describe the relative contributions of the hybridized transcripts.

**RNA Pooling**

Pooling RNA samples from multiple individuals before cDNA conversion allows multiple individuals to be screened on the same microarray or sequencing lane; therefore, population representation is increased without increasing cost (Pronk et al. 2011; e.g. Zhang et al. 2005; Richards et al. 2012). The utility of pooling relies on the concept of biological averaging, pooled transcript abundance represents an average of the expression states among pooled samples (Kendziorski et al. 2005, Zhang et al.)
2005). However, pooling introduces a number of artifacts into the data. First, overall expression variability is reduced (Kendziorski et al. 2005). While this may be advantageous for field studies that can be extremely variable, it may lower precision of detection of expression levels of some genes (Kendziorski et al. 2005). Second, the measured expression will be more attenuated. In a pooled design, genes are averaged twice – once biologically because of pooling, and once technically during data normalization, which results in nonlinear distortion (Kendziorski et al. 2005). Third, genes that are expressed at a lower level are more affected by the distortion introduced by pooling (Pronk et al. 2011), and differential expression of these genes may go undetected. Thus, RNA pooling may exacerbate the problem of minimally-expressed genes having larger effects on phenotype despite a small change in abundance (Oleksiak et al. 2005). Finally, because it is not possible to separate individuals from an RNA pool, pooling results in the loss of the ability to measure individual differences.

**Statistical Analysis**

Classic analysis of variance (ANOVA) has been adapted to the interpretation of gene expression and gene-specific modeling of microarray data by fitting a global normalization model incorporating all of the genes, and then running a separate ANOVA for each gene (Wolfinger et al. 2001; Aryoles & Gibson 2006). The analysis of RNAseq data is not as mature and a consensus does not yet exist for RNAseq on preprocessing, normalization, and inference methods. RNAseq data are generally described as an overdispersed Poisson distribution (Kvam et al. 2012; Wolf 2013), so familiar analyses and software that rely on normally distributed data, like the R package limma, may not
be suitable for RNAseq data (Kvam et al. 2012). Law et al. (2014) have proposed a newer methodology, which generates a precision weight for each observation. This weighting system allows users to analyze RNAseq data as normally distributed data, which would allow analysis to follow the methods previously-described for microarrays (Law et al. 2014). However, RNAseq facilitates analysis methodologies not available for microarray data. RNAseq detects transcript polymorphisms (Ekblom & Galindo 2011), which allows for the investigation of molecular evolution (Williams & Oleksiak 2008) and population genetics (Williams et al. 2010). This capability opens the door for integrating population genomics approaches into gene expression studies (Ekblom & Galindo 2011).

Unannotated Genes

One clear advantage of RNAseq over microarrays is the identification of previously uncharacterized transcripts. RNAseq directly screens transcripts and does not rely on the design of probes from previously identified targets. However, non-model organisms may have limited annotation information available for species-specific genes, and there may be many unidentified genes whose functional relevance cannot be determined (Pavey et al. 2012). Although this problem may be mitigated by annotating sequences from homologs in closely related species using the Basic Local Alignment Search Tool (BLAST; Altschul et al. 1990), caution should be used when annotating genes with homologs from more divergent species, as heterologous homologs may no longer possess the same function. A long-term solution to the problem of unannotated genes is the development of databases and repositories to collect ecological
annotations, which can alleviate the need for annotations derived from distantly-related model organism.
Appendix 2

Copyright Permissions

JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS

Jun 24, 2016

This Agreement between Mariano F Alvarez ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number: 3895520284823
License date: Jun 24, 2016
Licensed Content Publisher: John Wiley and Sons
Licensed Content Publication: Molecular Ecology
Licensed Content Title: Ten years of transcriptomics in wild populations: what have we learned about their ecology and evolution?
Licensed Content Author: Mariano Alvarez, Aaron W. Schrey, Christina L. Richards
Licensed Content Date: Jan 21, 2015
Licensed Content Pages: 16
Type of use: Dissertation/Thesis
Requestor type: Author of this Wiley article
Format: Print and electronic
Portion: Full article
Will you be translating? No
Title of your thesis / dissertation: Molecular response of Spartina alterniflora to the Deepwater Horizon oil spill
Expected completion date: Aug 2016
Expected size (number of pages): 115
Reuestor Location: Mariano F Alvarez
8517 Island Breeze Lane
Unit 202
TAMPA, FL 33637
United States
Attn: Mariano F Alvarez

Publisher Tax ID: EU826007151
Billing Type: Invoice
Billing Address: Mariano F Alvarez
8517 Island Breeze Lane
Unit 202
TAMPA, FL 33637
United States
Attn: Mariano F Alvarez

Total: 0.00 USD