ECOLOGICAL GENETICS OF ADAPTIVE LIFE-HISTORY PHENOTYPES IN THE CRICKET *ALLONEMOBIUS SOCIUS*

by

DIANA LEA HUESTIS

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2008

ACKNOWLEDGEMENTS

This research would not have been possible without the advice and assistance of many individuals and the monetary support of several agencies. First, I owe many thanks to Dr. Jeremy L. Marshall, who has advised me for the past five years, first as a M.S. student and now as a Ph.D. student. Jeremy's patience and guidance throughout my graduate career has enabled me to grow as a scientist and as a colleague. Second, Dr. Esther Betrán has also worked with me for the past five years, and filled in as my advisor at UTA in recent years. Next, I thank Dr. Michael J. Wade for allowing me to work in his lab at Indiana University in Bloomington for the past three years, providing incubator and desk space as well as a collegial lab environment and scientific input. Finally, I would like to thank my other doctoral committee members, Dr. Andre Pires da Silva, Dr. James V. Robinson, and Dr. Daniel F. Formanowicz.

This work has also been supported monetarily by several associations. I was suppored by a US EPA STAR GRO doctoral fellowship from September 2005 through August 2008, which provided tuition, research monies, and a living stipend, allowing the majority of this work to be conducted. I also received a Thelma Howell Memorial Fellowship from the Highlands Biological Foundation to collect crickets throughout the southeastern US in July 2006. The UTA Phi Sigma chapter, Department of Biology, and College of Science also provided small research grants. Finally, funding for molecular work conducted at Kansas State University was provided by the Kansas Agricultural Experimental Station to Dr. Jeremy L. Marshall.

I would also like to thank the many family members and friends for their continued support along the way. This work is dedicated to Virginia Huestis and to the memory of K. Frederick Huestis and Francis S. Wood.

August 11, 2008

ABSTRACT

ECOLOGICAL GENETICS OF ADAPTIVE LIFE-HISTORY PHENOTYPES IN THE CRICKET *ALLONEMOBIUS SOCIUS*

Diana Lea Huestis, Ph.D.

The University of Texas at Arlington, 2008

Supervising Professor: Esther Betrán

As many organisms inhabit unstable environments, the ability to respond to changing environmental conditions is highly adaptive, and results in the evolution of phenotypic plasticity. Thus, identifying the genetic basis and evolutionary origins of adaptive life-history phenotypes, such as phenotypic plasticity, is an active and important research area within the field of evolutionary biology. In the last two decades, gene-expression techniques have been developed which allow genes to be identified and gene-expression levels measured in a range of species, reducing dependence on classic genetic model systems and allowing an array of species with interesting ecological phenotypes to be used as models to study the evolution of adaptive traits.

Here, I first provide an overview of gene-expression techniques, both classic and more modern, which can be used to identify genes and their expression levels in a range of species, including both classical genetic model systems and any species of interest. This chapter provides details of the use and specific applications for many common (and some less common) gene-expression techniques, provides information about the specific laboratory methodology used to perform each technique, whether a given technique can be used in any system or is restricted to

genetic model systems, and finally what follow-up experiments are needed to confirm differential gene expression. This chapter is intended to be a resource for any biologist interested in learning more about what gene-expression techniques may be useful in their particular system to address a given biological question.

Second, I use one of these gene-expression techniques, suppression subtractive hybridization (SSH), to identify gene-expression differences between developmental phenotypes in the cricket *Allonemobius socius*. Specifically, I looked at expression differences between diapausing and direct-developing embryos. The ability to break diapause in response to changing environmental conditions is an adaptive trait, yet little is known about its genetic basis in egg-diapausing species. Several genes were identified as being upregulated in eggs which had broken diapause, relative to their diapausing counterparts, and other genes were identified as expressed in diapausing eggs but not in direct-developing eggs.

Next, I explore population-level variation in the plasticity of the diapause response within twelve geographically-diverse populations of *A. socius*. I also use the amplified fragment-length polymorphism (AFLP) technique to look for genetic correlates of mean diapause frequency and plasticity among these twelve populations. I found significant variation within and among populations for both mean and plasticity of diapause. Additionally, different components of genetic variance appear to underlie the mean and plasticity of the diapause response, indicating that these two components may be able to evolve independently in this species.

Finally, I explore the possible evolutionary origins of a cline in allele frequencies at the isocitrate dehydrogenase-1 (*Idh*-1) locus in *A. socius*. Differences in enzyme kinetic performance appear to underlie the natural geographic distributions of two alleles at this locus, indicating that natural selection may have shaped the distribution of *Idh*-1 alleles in this species. This chapter serves as a case-study for the type of additional follow-up experiments which can be performed once a gene of interest associated with a particular adaptive phenotype is identified using a combination of phenotypic screens and genetic techniques.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1

FROM GENE EXPRESSION TO PHENOTYPE IN INSECTS: APPROACHES FOR THE MASSES

1.1 Abstract

Transcripts and their expression levels link an organism's genotype and phenotype; thus understanding this relationship is critical to our understanding of phenotypic evolution. The emerging field of functional genomics links the study of allelic and gene-expression variation to observable, biologically-relevant phenotypes. Insects are particularly well-studied in this area, because they are both good laboratory systems and of interest because of their incredible biodiversity and agricultural importance. New technology now allows for gene expression studies in any insect system, advancing the fields of functional and ecological genomics beyond traditional genetic model systems. Here, I provide an overview of commonly-used gene expression techniques in insect systems and review several empirical studies utilizing each technique. I also discuss RNA interference (RNAi) as a means to test the link between gene expression and phenotype for candidate loci. I end with a discussion of how new sequencing methods are advancing the field of functional genomics.

1.2 Introduction

Insects are at the forefront of the burgeoning field of genomics, as more than a quarter of the animal genomes that have been or are currently being sequenced are from insects (i.e., 40 of 147 animal genomes; www.ncbi.nlm.nih.gov/genomes/static/gpstat.html). This focus on insects is easy to understand given their incredible biodiversity and role as human and agricultural pests. Advances in high-throughput sequencing technologies like pyrosequencing (Margulies et al. 2005, Emrich et al. 2007) will enable researchers to sequence their insect

genome of choice, at a minimal cost (<$100K USD), within the next five years or so. Therefore, entomologists and other researchers utilizing insects will soon have all the genomic information they need to tackle questions ranging from “what genes control speciation?” to “what genes should be targeted for pest management?”. However, just sequencing large numbers of genomes is not enough to answer these types of questions; researchers must also be able to link genotype to phenotype. This link, which constitutes the area of functional genomics, will be one of the most active areas of study over the next half century.

An organism’s genome is composed not only of genes and their regulatory regions, but also pseudogenes, repetitive elements, and other non-coding sequence elements. Additionally, all genes in the genome are not expressed in every tissue or at all times in an organism’s ontogeny. Therefore, even having an entire sequenced genome is not enough to link genes and phenotype. Within cells, messenger RNA (mRNA) is transcribed from the genome, and these transcripts are then translated into proteins; mRNA, therefore, is the critical bridge between the genotype and the phenotype.

Whole-tissue or cell-specific studies of gene expression focus on identifying and quantifying the mRNA transcripts that are present at a particular time and place (Donson et al. 2002, Gracey et al. 2003, Tittiger 2004). These mRNA transcripts are a direct by-product of gene transcription and encode the information to be translated into proteins; thus mRNA transcripts provide a way to assess genome products as well as a genome’s response to environmental cues (whether those cues concern mating, foraging, temperature regulation, photoperiod, etc.). The entire pool of mRNA transcripts in a given cell- or tissue-type is generally referred to as a transcriptome. For example, if we were interested in aphid feeding behavior, then we might want to study the salivary-gland transcriptome, as these would be all the salivary gland mRNA transcripts that might be involved in successful foraging.

The first transcriptome-wide studies in insects were published in 1995. Since these first publications, transcriptome-wide studies have usually focused on one of two goals: (1)

identifying all the genes that are expressed in a given cell or tissue or (2) identifying the differences in gene expression that are associated with particular phenotypes or experimental treatments. In addressing the former, researchers have typically employed one method: sequencing expressed sequence tags (ESTs) from a cDNA library made from the tissue or cell-type of interest. This straightforward technique consists of RNA isolation, cDNA synthesis, and cloning these cDNA fragments into a bacterial vector, resulting in a cDNA library. This cDNA library is then plated and clones are picked for sequencing. The resulting sequences, typically 200-600bp in length, are known as expressed sequence tags (ESTs) and represent fragments of genes expressed in the cells or tissue of interest. EST libraries have been generated for many species of insects, including crickets (Andrés et al. 2006, Braswell et al. 2006), locusts (Kang et al. 2004), silkworms (Mita et al. 2003), and termites (Scharf et al. 2003).

The second major area that can be addressed with transcriptome studies is to identify gene expression differences associated with a particular phenotype or treatment. Gene expression differences between insect castes (Toth et al. 2007), life-history strategies (Chen et al. 2005), and insecticide-treated vs. untreated individuals (Guerrero et al. 2007) are just a few examples of phenotypes that have been studied using gene expression techniques in insects. Identifying the genes that underlie a particular phenotype or environmental response is of broad interest to most biologists, and it is here that I focus the rest of the chapter.

Researchers interested in identifying genes associated with a particular phenotype or response to experimental manipulation must address several questions before starting experiments, such as “how do we conduct this type of research?”, “is a genome for my system required?”, “what techniques are available?”, and “how much will it cost?”. The general flow for designing and carrying out such an experiment is diagramed in Figure 1.1, and my review will detail each step from start to finish. The experimental question to be addressed will determine

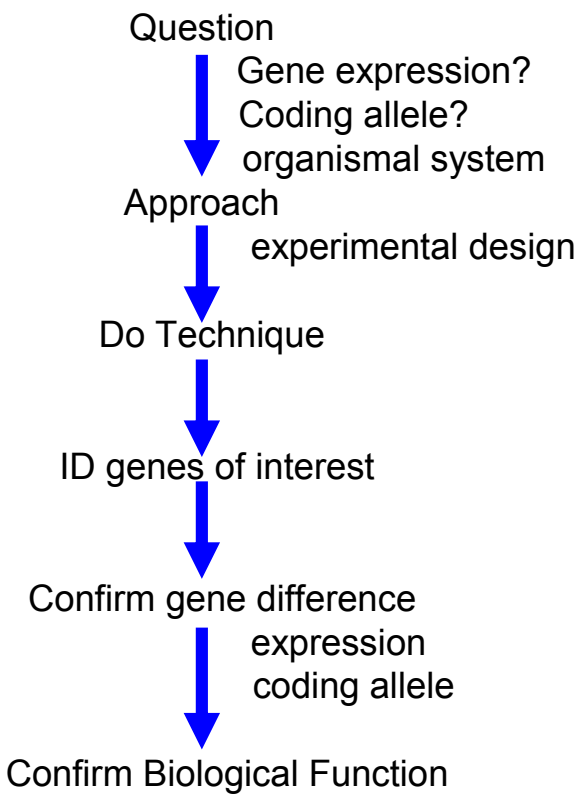

Figure 1.1. A general approach to conducting a gene expression study.

what technique should be used, and each technique has its own set of advantages and drawbacks. Once a technique is chosen, a proper experiment must be designed and samples collected (cells or tissues from which RNA will be isolated). These samples are then used to conduct the gene expression technique chosen. Fortunately for researchers working in non-genetic-model systems, many techniques do not require a sequenced genome and costs are lowering while quality is increasing. From the resulting gene expression data, candidate genes involved in the phenotype or response of interest are identified, which must then be confirmed with follow-up experiments, such as quantitative real time PCR (QRT-PCR). Once gene

expression differences are confirmed, the phenotypic consequences of the differences can be tested with the gene-silencing technology RNA interference (RNAi) – which has been a widely successful approach to knockdown expression of specific transcripts in larval and adult insects. I end with a discussion of new technology which may change and improve gene expression studies over the next few decades.

1.3 Choosing an Experimental Technique and Design

To begin, it always important to consider the set of questions that follows below. First, what is our question of interest? More specifically, what phenotype are we interested in and what is the best way to study that phenotype? Gene expression studies are not always the best approach, as quantitative genetic mapping (i.e., quantitative trait locus [QTL]) or proteomic approaches may be better suited for many questions. Second, what tissue type should be used to test our question of interest? If we are interested in insect foraging behavior, then we would likely want to study the brain and salivary glands. However, the choice of tissue type is not always clear, as many phenotypes are complex and involve many different tissue types. Therefore, it is always critical to consider carefully the biology of the phenotype of interest as well as the appropriateness of the genetic technique being used. Third, are we interested in coding allele (i.e., structural mutations that result in different amino acid sequences between treatments/samples) or gene-expression differences between treatments/samples, or both? Many of the techniques I outline below can address both types of genetic variation; however they are focused primarily on identifying gene expression differences and I discuss them in this light.

Once these basic questions are addressed and a gene-expression approach is determined to be the best course of action, the next step is to decide which technique to use in your study. When choosing a gene expression technique, the first consideration is the number of samples that need to be compared. Certain techniques identify gene expression difference between two samples (such as subtractive hybridization and microarrays) whereas others can

be used to compare expression levels between any number of samples (including PCR-based techniques and comparative ESTs). The second consideration is how much genome or transcript sequence information is available for the system of interest. Serial analysis of gene expression (SAGE), for example, requires a substantial amount of sequence data to successfully identify genes; others are more effective with, but do not require, sequence data (PCR-based techniques, subtractive hybridization); while others do not require prior transcript information because sequence data are generated by performing the technique (comparative ESTs). Cost may also be an important factor when choosing a technique, as some methods cost more to implement than others. For example, the comparative EST method will be expensive if ESTs have not yet been generated, whereas PCR-based techniques are relatively inexpensive since only PCR and sequencing reagents are required. Importantly, the techniques we describe here are not limited to any particular trait, experiment, or phenotype and most can be conducted on any organismal system of interest.

Once a technique is chosen, proper experimental design must be used to generate the samples for analysis. The design will depend on the question being asked and what level of variation the researcher is interested in (i.e., between individuals, populations, or species). Biological replicates are needed to reduce experimental bias but must be generated with the question in mind. For example, if the researcher is examining gene expression difference due to treatment effects, it may be appropriate for biological replicates to be different populations of the species of interest; on the other hand, if we are looking at differences between populations or species, then the level of replication should be individuals within a population. Technical replicates may be used to identify errors or bias while performing the molecular portion of the work, and are useful for some techniques (PCR-based techniques, subtractive hybridization) but would be costly or impractical for others (comparative ESTs, microarrays). Once a technique is chosen and the proper experimental manipulations carried out to generate samples, the next

step is to perform the gene expression technique of choice and start identifying candidate genes associated with the phenotype or treatment effects.

1.4 Techniques for Comparing Gene Expression

1.4.1 PCR-based Differential Display Methods

There are two closely-related comparative transcriptomics methods utilizing the polymerase chain reaction (PCR) to amplify differentially-expressed mRNAs: mRNA differential display PCR (dd-PCR) and cDNA amplified fragment length polymorphism (cDNA-AFLPs). Each method includes the isolation of RNA from two or more samples to be compared, cDNA synthesis, and some form of PCR amplification. The end result of both is the visualization of differentially-amplified products, typically on a polyacrylamide gel; the resulting bands must then be excised, re-amplified, and sequenced to yield gene identification. The key difference between these techniques is the method used to amplify the cDNAs to be compared between samples (see box 1 for detailed methods).

1.4.1.1 dd-PCR

One very popular method for detecting differences in gene expression is differential display PCR (dd-PCR; Liang and Pardee 1992). To begin, single-stranded cDNA is generated for each sample using a poly-T primer anchored with one or two bases on the 3' end (box 1, panel a). Next, PCR on each cDNA pool is performed, using the appropriate poly-T primer and a random upstream primer for amplification. Typically, dozens or even hundreds of primer combinations are used to cover the entire transcriptome (Liang and Pardee 1992, Liao and Freedman 2002). PCR products from different samples amplified with the same primer combination are then electrophoresed side-by-side (box 1, panel b), typically on a polyacrylamide gel (Liang and Pardee 1992, Liang et al. 1995), though agarose methods have also been developed (Rompf and Kahl 1997, Zeppa et al. 2002). Bands which appear to be differentially expressed between the samples of interested are then excised from the gel, re-

amplified, cloned, and sequenced. Differential expression of genes of interest may then be confirmed by quantitative real-time PCR (QRT-PCR) or Northern blots.

Due to its technical simplicity, lack of genomic information needed, and relatively low cost, dd-PCR is perhaps one of the most widely-used differential expression techniques (Kuhn 2001, Liao and Freedman 2002). Dd-PCR has been applied widely to mammals, plants, and insects. For example, Graff et al. (2007) used dd-PCR to identify differentially-expressed genes between queens and workers of the ant *Lasius niger*. Northern blots and QRT-PCR were used to confirm 16 dd-PCR gene fragments as being differentially expressed between these two castes, providing candidate genes involved in caste differentiation. In another experiment, dd-PCR was used to characterize gene expression following viral infection in the midge *Culicoides sonorensis* (Campbell and Wilson 2002). Out of 29 transcripts initially identified with dd-PCR, 13 were confirmed with differential hybridization; of these, 7 were confirmed by QRT-PCR and a follow-up experiment detailing expression profiles over time after infection (Campbell and Wilson 2002).

The traditional dd-PCR technique can be modified to include the use of restriction enzymes, known as restriction fragment differential display (RFDD-PCR). RFDD-PCR reduces the 3' bias of traditional dd-PCR and may be more reproducible (Masinde et al. 2005). Because it does not require a poly-A tail for cDNA synthesis, RFDD-PCR has been the preferred differential display method for prokaryote systems, but has not been widely used in insects. In one application to insect systems, this method was used to compare gene expression between iron-treated and control Colorado potato beetles (*Leptinotarsa decemlineata*) and resulted in the identification of a ferritin protein with significant amino acid similarity to that of other insects (Qiu et al. 2005).

Drawbacks of the dd-PCR technique include a high false-positive rate (5-50%; Martin and Pardee 2000), bias toward the 3' end of cDNA (though this can be reduced by using arbitrary primers for both reverse transcription and PCR; Welsh et al. 1992), and redundancy of

amplified bands (as one differentially-expressed gene may by amplified with several primer combinations). However, the ease of this technique makes it applicable to almost any system, making it a popular choice for researchers (~2,100 SCI citations by August 2007).

1.4.1.2 cDNA-AFLPs

Another method for amplifying differentially-expressed mRNAs is based on the cutting of cDNA by restriction enzymes, ligation of adaptors to cut sites, and PCR amplification from these adaptor sequences (box 1). As in dd-PCR, products are then electrophoresed side-by-side and banding patterns compared; bands of interest may then be excised and sequenced. Read lengths typically fall between 100-500 bp, though choice of restriction enzyme can increase or decrease length (Bachem et al. 1996, Habu et al. 1997, Kuhn 2001) and improve specificity to the transcriptome of the desired system (e.g. Reineke et al. 2003). Typically, 256 primer combinations are required to cover the majority of the transcriptome (Habu et al. 1997, Kuhn 2001), though in practice many fewer are used. The cDNA-AFLP method was originally developed in plants (Bachem et al. 1996, 1998), but has since been applied to invertebrates (e.g. Reineke et al. 2003, Yang et al. 2006) – though it is underutilized in animals as an alternative to microarrays (Bensch and Akesson 2005).

The AFLP technique was originally developed for genomic DNA (Vos et al. 1995), but was quickly applied to cDNA (Bachem et al. 1996). Though developed and most often used in plants, cDNA-AFLPs have been successfully applied to several insect systems. For example, Reineke and Löbmann (2005) identified 59 transcripts that were differentially expressed between caterpillars (*Ephestia kuehniella*) parasitized and unparasitized by the endoparasitic wasp *Venturia canescens*. Of these transcripts, 27 were successfully excised, cloned, and sequenced, and 13 of these were confirmed with Northern blots and QRT-PCR, all of which corresponded to transcripts suppressed in parasitized caterpillars relative to unparasitized caterpillars (Reineke and Löbmann 2005). Similarly, cDNA-AFLP was used to detect changes in gene expression between brown planthoppers (*Nilaparvata lugens*) feeding on resistant and

susceptible strains of rice (Yang et al. 2006). Using 30 primer combinations, 61 differentially-expressed fragments were identified, cloned and sequenced. Thirteen bands had sequence similarities to known genes, with functions including from detoxification, stress response, and signaling. Of these, 4 were chosen as genes of interest for further characterization, all of which were confirmed with Northern blots (Yang et al. 2006).

Due to stringent amplification conditions for PCR, the cDNA-AFLP method has higher reproducibility and a lower false-positive rate than dd-PCR (Bachem et al. 1996, 1998, Habu et al. 1997, Donson et al. 2002), as well as lower redundancy (about 2%; Bachem et al. 1998). Use of fluorescent labeling and capillary electrophoresis allows for high-throughput analysis (Donson et al. 2002), but limits the ability to identify differentially-expressed transcripts. Future advances using pyrosequencing of cDNA-AFLP products could eliminate this shortcoming (see section 1.4.5).

1.4.2 Suppression Subtractive Hybridization

Another PCR-based method is suppression subtractive hybridization (SSH) of two cDNA samples. In SSH, cDNA library subtractions are performed on samples of interest to identify genes expressed (or upregulated) in one sample but not in the other; as a result, this technique is especially useful for comparing two closely-related samples (Diatchenko et al. 1996, 1999). As each subtraction only identifies genes expressed in one sample (the “tester”) relative to the other (the “driver”; see box 2 for detailed methods), forward and reverse subtractions must be performed to identify expression differences in both directions. Once genes expressed in both samples are subtracted out, PCR amplification and electrophoresis produces fragments which represent genes expressed in the “tester” but not in the “driver,” which may be extracted from the gel, purified, and sequenced for identification.

Though first developed in mammals, SSH can be performed in any system with mRNA. In insects, SSH has been used to study immunity, ecological, and behavioral traits. For example, Zhu et al. (2003) used SSH to identify genes upregulated in response to bacterial

infection in the tobacco hornworm, *Manduca sexta*. Over 230 differentially expressed genes were identified, half of which were identified as immune-response genes after sequencing. Genes of interest were confirmed with a combination of Northern blots, QRT-PCR, and a follow-up experiment involving 2-D protein electrophoresis (Zhu et al. 2003). In another recent study, gene expression changes in response to cadmium exposure were studied in springtails (Roelofs et al. 2007). Subtractions were performed in both directions to identify genes both up- and downregulated after exposure; expression between tolerant and susceptible populations was also studied. This study resulted in the confirmation of 7 genes, confirmed by QRT-PCR as upregulated in response to cadmium exposure, which are candidate genes for further study of heavy-metal response and tolerance (Roelofs et al. 2007).

Since SSH incorporates suppression PCR, it is able to both suppress equally-expressed fragments and preferentially amplify those which significantly differ in copy number (Lisitsyn et al. 1993). This method therefore reduces the number of false-positives and fragment redundancy relative to dd-PCR (Diatchenko et al. 1996). SSH has also been reported to be effective for detecting gene-expression differences between closely-related samples, identifying low copy-number fragments, and probing cDNA libraries to confirm differential expression (Diatchenko et al. 1999). Two disadvantages to this method are that SSH requires a large amount of mRNA (although amplification techniques may be used; Diatchenko et al. 1996) and that expression levels must be very great for a fragment to be identified (i.e., expressed vs. not expressed, or expressed greater than 4-fold higher; Diatchenko et al. 1999) due to the driver cDNA being added in excess (see box 2). However, the lack of starting genomic information needed combined with the availability of a commercial kit (which can produce a full library of differentially-expressed genes in about 7-10 days) makes SSH a good choice for many applications comparing 2 or 3 treatments/samples.

1.4.3 Comparative ESTs

Other methods for comparing expression levels between samples of interest are computational, rather than PCR-based. One such method involves the comparison of sequence abundance between two or more cDNA libraries, either already generated by other experiments or created for the purpose of comparing expression levels. Once the cDNA libraries are generated, ESTs are sequenced and the resulting sequences are grouped together in clusters of similar sequences (called contigs; see box 3 for detailed methods). The number of times a particular EST is found in a ontigs can then be calculated for each library, and compared statistically between the samples using one of several methods (e.g. Audic and Claverie 1997, Stekel et al. 2000, Ribeiro 2003).

This method was used to analyze expression differences between solitary and gregarious morphs of the migratory locust, *Locusta migratoria*, an economically-important agricultural pest (Kang et al. 2004). Out of over 12,000 ESTs analyzed, 532 were identified as potentially involved in morph differentiation; only a few unigenes were selected for QRT-PCR, but each was subsequently confirmed (Kang et al. 2004). This study not only provided the first EST data for a hemimetabolous insect, but shows the potential power of the comparative EST technique. This technique was also used to identify expression differences between the embryonic and larval stages of the silkworm *Bombyx mori*. Of the 69 ESTs identified as potentially differentially-expressed by comparing copy number, 33 were confirmed with semi-quantitative RT-PCR (Oh et al. 2006). As in the previous example, this study provided both new genome/transcriptome sequence data and candidate genes involved in differentiation of life-history stages.

One drawback to this method is that cDNA libraries must be generated and sequenced for the samples of interest if these data are not already available. This can be a time-consuming and expensive project; however, once completed, researchers have a valuable genomic resource for the system of interest. Additionally, given gene expression differences are based

on computational methods and dependent upon the total number of ESTs initially sequenced, each gene must be confirmed with QRT-PCR or other quantitative methods; though the false-positive rate decreases with the number of times a particular EST appears in the combined libraries (Audic and Claverie 1997). In summary, the comparative EST approach is a fairly straightforward approach to identify differentially-expressed genes if appropriate EST libraries already exist; if libraries must be generated, doing so will require extra work but provide an important database of genetic data. Additional advances, including the use of pyrosequencing in generating ESTs, may greatly increase the ease of this method (see section 1.4.5).

1.4.4 Serial Analysis of Gene Expression

Serial analysis of gene expression (SAGE) is a restriction enzyme-based, high-throughput method for determining gene expression in eukaryotes, first described in 1995 (Velculescu et al. 1995). Briefly, double-stranded cDNA is synthesized from the sample(s) of interest, cut into small segments with restriction enzymes, and short linker segments ligated onto the cleaved ends. These small segments with linkers attached, known as tags, are enzymatically joined end-to-end and the resulting fragments, each containing 10-50 tags, are cloned and sequenced (see box 4 for detailed methods). The expression of a given EST is determined by the number of times its tag appears in the resulting sequences (Kuhn 2001). Because the resulting sequence tags are very small (~15 bp), SAGE requires initial genomic and/or transcriptomic data for the tags to be matched to (Pleasance et al. 2003). Thus, SAGE is restricted to systems with a sequenced genome or EST library (Tittiger 2004), unless the researcher also constructs an EST library.

In insects, SAGE experiments are usually conducted in model organisms with a sequenced genome or EST library available, including *Drosophila* (Jasper et al. 2001, Metta et al. 2006), *Bombyx mori* (Huang et al. 2005, Funaguma et al. 2007), and *Anopheles gambiae* (Rosinski-Chupin et al. 2007). One recent study using SAGE in a non-model insect system compared gene expression in pesticide-resistance and susceptible strains of the cattle tick

Rhipicephalus microplus (Guerrero et al. 2007). Expression differences of many tags were detected, though many could not be unambiguously identified due to lack of genomic data for this species (Guerrero et al. 2007). However, several genes of interest were identified and confirmed as being differentially-expressed with QRT-PCR, providing candidate genes for future research on pesticide susceptibility. In systems with abundant genetic data, however, SAGE is quite effective. For example, 57 genes associated with *Plasmodium* infection were identified in *A. gambiae*, and expression levels detected with SAGE were significantly correlated with QRT-PCR results (Rosinski-Chupin et al. 2007), indicating that SAGE is a powerful technique for systems where it can be used.

One of the benefits of SAGE is that it is high-throughput, as the expression of 20-100 ESTs is detected with each sequence read (Velculescu et al. 1995, Kuhn 2001). Therefore, large amounts of expression data can be generated with relatively little sequencing, provided that the sequence tags can be matched to an appropriate database. One obvious drawback to the SAGE technique is that extensive genomic and/or transcriptomic data for the study system is required, and must be generated if not already available. Also, because the sequence tags are small (~15bp) and biased toward the 3'UTR, some tags cannot be unambiguously assigned to a particular gene; additionally, the expression of genes which lack the appropriate restriction enzyme recognition sites cannot be detected by SAGE (Kuhn 2001, Pleasance et al. 2003). Estimates of genes which cannot be analyzed with SAGE range from 8% in *Drosophila melanogaster* to 15% in *Caenorhabditis elegans* (Pleasance et al. 2003). For genetic model systems, it has been suggested that SAGE has comparable sensitivity to microarrays, but has the benefits of requiring no special equipment and being lower in cost (Kuhn 2001). In summary, although SAGE is typically restricted to a few genetic model systems, its use could increase as sequence data become available, because the high-throughput nature of this technique has the potential to provide large amounts of expression data with less effort and at a lower cost than some other techniques.

1.4.5 Pyrosequencing and Other New Technologies

Pyrosequencing is a recently-developed, high-throughput method of DNA sequencing (reviewed in Ronaghi 2001) that has the potential to greatly facilitate both genome sequencing and gene identification when used in conjunction with one of the gene expression techniques described above. Briefly, the DNA to be sequenced is randomly sheered and then loaded onto a chip containing all reagents necessary for PCR amplification and sequencing (Margulies et al. 2005, Hudson 2007). Sequencing is performed by washing fluorescently-labeled nucleotides over the wells, and a fluorescent pulse is given off when a nucleotide is incorporated.

The most widely-available platform for pyrosequencing is from 454 Life Sciences™, now marketed by Roche Applied Science (http://www.454.com/). Current 454 pyrosequencing technology generates 400,000 sequence reads, averaging 250bp in length, from a single sequencing run (Hudson 2007; http://www.454.com/enabling-technology/the-system.asp). One chip may be divided between 2-16 samples. Although this technology was only described in 2005 (Margulies et al.), 70+ publications to date have utilized this technology (Hudson 2007), including several transcriptome studies (Emrich et al. 2007, Hudson 2007, Toth et al. 2007, Weber et al. 2007). Relative to traditional Sanger sequencing methods, 454 pyrosequencing has the advantages of a 10-fold reduction in cost/base, faster, high-throughput results (one <7hr run yields up to 100mb), and an error rate of less than 0.5% (Hudson 2007, Huse 2007). Thus, utilizing pyrosequencing along with another gene expression technique, such as PCR-based methods, SSH, and comparative EST's, may increase the amount of sequence data able to be generated, thereby increasing the power of the chosen technique.

Additionally, 454 pyrosequencing alone, if used correctly, also has the potential to be applied as a stand-alone gene expression technique. During a pyrosequencing run, template DNA is randomly sheered and dispersed into sequencing wells. It is therefore predicted that highly-expressed transcripts will produce more sequence reads than less-expressed transcripts, provided the sample is not "normalized" during cDNA preparation to equalize transcript

abundance. If this assumption is true, it would be possible to divide a single chip between two (or more) cDNA samples generated from tissues of interest and generate ESTs by direct pyrosequencing. The number of ESTs for each gene would then be calculated for each sample and compared between the samples. Thus, due to its high-throughput capability, 454 pyrosequencing technology has the potential to provide both sequence and expression data with one procedure. Candidate genes which appear to be differentially-expressed in EST abundance should later be confirmed with quantitative real-time PCR (QRT-PCR). Other high-throughput sequencing technologies (e.g., SOLiD and Solexa) also have the above potential and research over the next decade will determine which of these platforms positively impact the study of gene expression in non-traditional genetic models.

1.5 Techniques for Confirmation of Gene Expression Differences

Once candidate genes are identified using one of the above gene expression techniques, differential expression must be confirmed; typically, a PCR-based technique is used, which can be either semiquantitative or quantitative. The candidate gene sequence is used to design primers resulting in a small (~300bp) PCR fragment of that gene. First-strand cDNA products from the samples of interest are used as template DNA for PCR reactions. PCR is conducted simultaneously with the candidate gene-specific product and a similar-sized fragment of a housekeeping gene that should be equally expressed in all samples.

1.5.1 Semiquantitative PCR

Semiquantitative PCR was the first quantitative method developed to compare relative amounts of starting cDNA and uses relative brightness of bands on an agarose gel (Wiesner et al. 1992, Tsai and Wiltbank 1996). In semiquantitative PCR, reactions are performed in a standard thermalcycler under standard PCR conditions, except that the number of cycles should be low (~15-18) – to ensure the reaction is in the exponential phase and not the saturation/plateau phase of the PCR reaction. Resulting products are then electrophoresed side-by-side on a standard gel and visualized (Figure 1.2a). A brighter band in one sample

relative to another is indicative of a higher amount of template cDNA (i.e., more mRNA transcripts). The relative brightness of one band relative to another can be analyzed with imaging software, such as NIH's ImageJ, providing a semiquantitative measure of relative gene expression.

1.5.2 Quantitative PCR

Quantitative real-time PCR (QRT-PCR) is a more powerful and precise method to measure relative gene expression between two samples (reviewed in Freeman et al. 1999). The basic method is similar to that described above, except that the primers are tagged with a fluorescent label to track incorporation of the primers into a cDNA fragment (Gibson et al. 1996, Heid et al. 1996). PCR reactions proceed exponentially before leveling off, and it is the number of cycles that it takes to reach the exponential growth phase that is measured (Figure 1.2b). Reactions which reach this earlier have a higher amount of starting template cDNA (in this case, cDNA corresponding to the candidate gene of interest). Real-time PCR has been used extensively to quantify gene expression differences in insect systems, with at least 321 studies utilizing this technology [SCI citations using the following keywords: real time PCR AND (insect or drosophila or tribolium or anopheles or apis)].

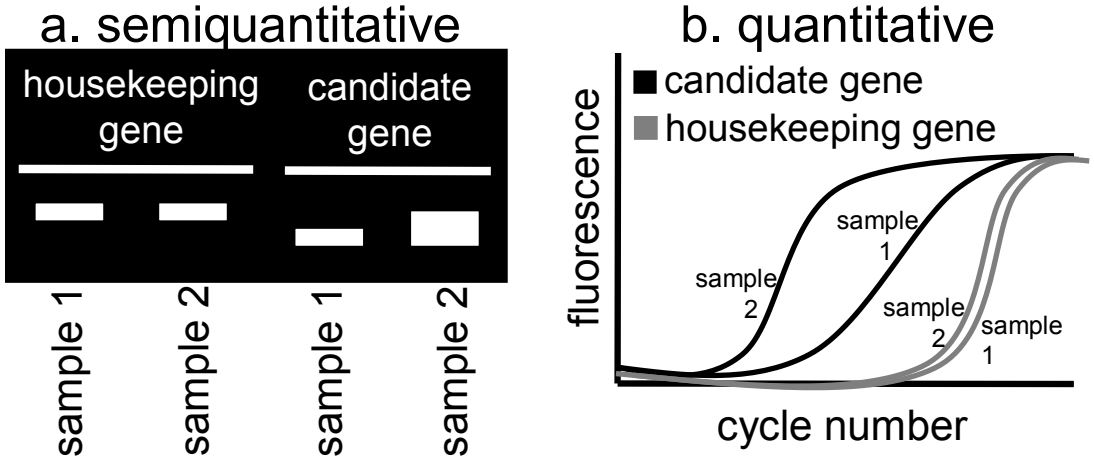

Figure 1.2. Diagrams of a) semiquantitative and b) quantitative PCR. In a, the brightness of the housekeeping fragment is equal in both samples, while the candidate gene has a brighter band in sample 2 than in sample 1. In b, the reaction reaches the exponential phase earlier for sample 2 than in sample 1 for the candidate gene, while the housekeeping genes are approximately equal.

1.6 Phenotypic Confirmations Using RNA Interference

Once gene expression differences between the treatments/samples of interest have been identified and confirmed, the next step is to test the biological importance of that difference on the phenotype that originally defined your treatments/samples. There are several approaches that can be used to do this, including injections of the protein product that the transcript produces (e.g., Cayre et al. 1995), genetic deficiency mapping (e.g., Pasyukova et al. 2000) and RNAi (see below). However, the former two can be technically challenging and may not be appropriate for many insect systems. RNA interference (RNAi), on the other hand, is relatively easy and will likely work in any insect system (review of RNAi in Gheysen and Vanholme 2007). Specifically, injection of double stranded RNA (dsRNA; see flyrnai.org for a detailed description of methods to generate dsRNA), derived from a specific gene, into the abdomen or body cavity of larval or adult insects can knockdown mRNA expression of the target gene. Developed by Fire et al. in 1998, this Nobel Prize winning technology (Nobel Prize in Medicine in 2006) has revolutionized the study of gene expression effects on phenotypes. Indeed, injection of dsRNA into larval and adult insects has been a successful strategy to knockdown gene transcripts in a diverse array of insect taxa, including aphids (e.g., *Acyrthosiphon pisum*; Mutti et al. 2006), beetles (e.g., *Tribolium castaneum*; Arakane et al. 2005), cockroaches (e.g., *Blattella germanica*; Maestro and Bellés 2006), crickets (e.g., *Gryllus bimaculatus*; Meyering-Vos et al. 2006), fruit flies (e.g., *Drosophila melanogaster*; Dzitoyeva et al. 2003), honeybees (e.g., *Apis mellifera*; Nelson et al. 2007), mosquitoes (e.g., *Anopheles gambiae*; Lycett et al. 2006), moths (e.g., *Spodoptera frugiperda*; Meyering-Vos et al. 2006), and termites (e.g., *Reticulitermes flavipes*; Zhou et al. 2006).

For example, if a candidate gene has been confirmed as up-regulated in individuals displaying the phenotype or treatment effect of interest, knocking down that gene may confirm that the gene is involved in that pathway. If successful knockdown of the candidate gene removes or reduces the expected phenotype, there is evidence supporting the role of that gene

in that particular phenotype. Alternatively, if a candidate gene is found to be down-regulated when associated with a particular phenotype, knockdown of that gene in wild-type individuals should produce the phenotype of interest, again supporting the role of that candidate gene.

1.7 The Future of Gene Expression Research in Insects

Recently-developed technologies, like pyrosequencing, could change the way we conduct gene expression studies in insects and other species. Pyrosequencing has several advantages, as described above, over current expression techniques but only time will tell if these new technologies help identify candidate genes for biologically important phenotypes. In our opinion, the real future of gene expression studies lies not in the rapid identification of possible candidates, but in the confirmation of gene expression differences and more importantly in testing the consequences of these confirmed gene expression differences on phenotypic differences. Relatively few studies have tested for a direct link between differential gene expression and phenotype, compared with the thousands of studies that simply look at gene expression differences. It is the phenotype that usually drives our interest in gene expression and it is understanding the link between gene expression and phenotype that will likely consume our collective attention in the decades to come.

1.8 Boxes with Detailed Gene-Expression Technique Methods

1.8.1 Box 1: Detailed Methods: PCR-based Differential Display (dd-PCR and cDNA-AFLP)

Both dd-PCR and cDNA-AFLP's begin with the isolation of RNA from the samples of interest to be compared (e.g., different tissue types, species, experimental conditions, etc.). **First-strand cDNA synthesis** is then performed on each sample, using a reverse transcriptase and an anchored oligo-dT primer, complementing the poly-A tails of mRNA (panel a, top).

In the case of dd-PCR, several reverse transcriptase reactions are performed on each sample, using oligo-dT primers anchored with either one or two bases on the 3' end, resulting in 3 or 12 pools of cDNA, respectively, to cover the entire transcriptome. This single-stranded cDNA is used directly as the template for PCR amplification without further modification (panel

a, right side). PCR is then conducted in parallel on each sample with the oligo-dT primer used for first-strand synthesis and an arbitrary upstream primer, typically 6-13 bp in length. In general, several random primers are used with each oligo-dT cDNA pool to provide greater coverage of the transcriptome.

For cDNA-AFLPs, **second-strand synthesis** is performed on the first-strand cDNA product. Double-stranded cDNA is then cut with two restriction enzymes and adaptor sequences ligated to the cleaved ends (panel a, left side). Non-selective amplification using adaptor primers follows, generating a large amount of starting material for further amplification. Numerous PCR reactions using one adaptor primer and one arbitrary primer are then performed.

For both techniques, the resulting PCR products for each primer combination and sample to be compared are then electrophoresed side-by-side, and the resulting banding patterns compared (panel b). For both dd-PCR and cDNA-AFLP, there are 3 possible comparisons between two samples: 1) a fragment can be equally amplified between the two samples, indicating that the expression of this gene is equal across treatments, 2) a fragment can be amplified in one sample but not in the other, indicating that either (i) the gene is expressed only under those particular treatment conditions or (ii) there is unique variation in the coding sequence between the samples, or 3) a fragment can be amplified more in one treatment than the other, indicating it is **upregulated** (more expressed) or **downregulated** (less expressed) relative to the other treatment. Panel b provides a schematic representation of these banding patterns (top) and an example of experimental dd-PCR data (bottom; data from JL Marshall et al., unpublished). Fragments exhibiting expression patterns 2 or 3 may be chosen for further characterization, typically by excision from the gel, sequencing, and confirmation with either QRT-PCR or Northern blotting.

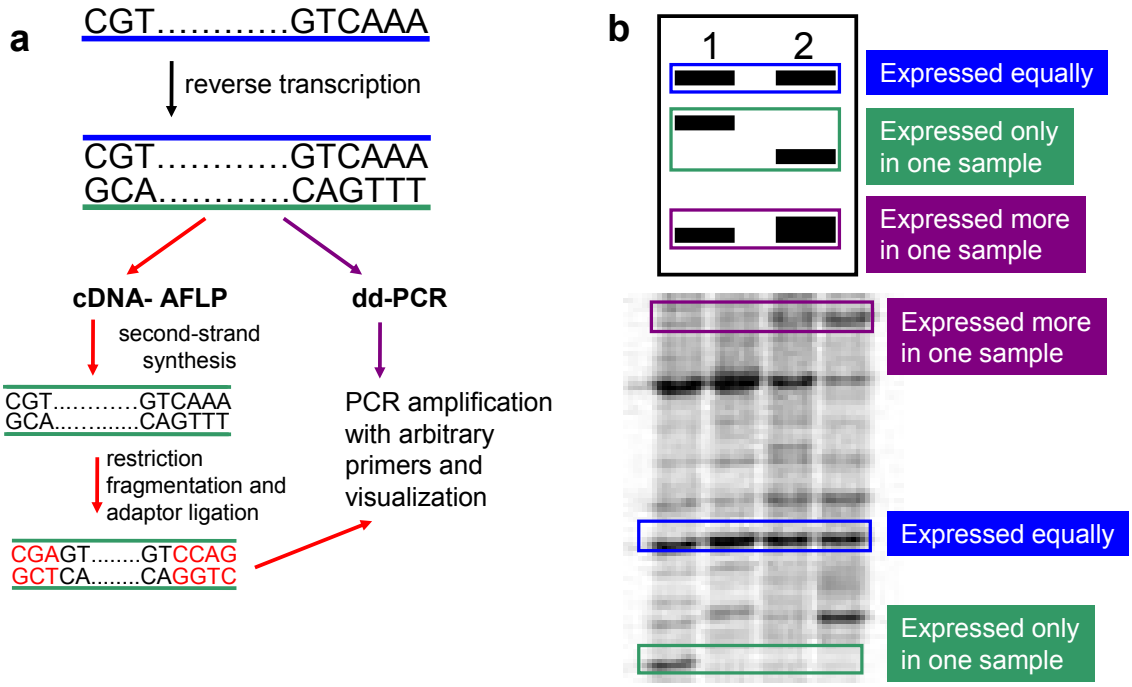

Figure 1.3. Pictoral diagrams of the cDNA AFLP and dd-PCR techniques.

1.8.2 Box 2: Detailed Methods: Suppression Subtractive Hybridization (SSH)

SSH identifies genes which are expressed in one sample but not in the other; thus it is a unidirectional comparison. For each hybridization, one sample is designated as the "**tester**," the sample in which genes will be identified, and one is designated as the "**driver**," which will be subtracted from the tester library. A reverse subtraction is often performed in parallel, allowing gene expression differences in both directions to be identified. Here we briefly detail the methods for SSH as first described (Diatchenko et al. 1996, 1999) and easily implemented using a commercially-available kit (PCR-Select™ cDNA Subtraction Kit; Clontech).

To begin, RNA is isolated from the samples of interest (panel a, top). Next, high-quality cDNA is generated separately for each sample and then cleaved with a restriction enzyme (panel a, middle). The driver cDNA is now ready for subtraction; however, the tester cDNA is split into two samples for ligation of two different adaptors (panel a, bottom). These adaptors include primer complementation sites to be used for PCR.

After ligation is complete, two rounds of subtraction are performed. For each subtraction, driver cDNA is added in excess to tester cDNA, the samples are heat-denatured, then allowed to re-anneal by cooling; this process generates tester-tester, tester-driver, and driver-driver hybrids (panel b, top). Once the subtractions are performed, the tester-tester cDNA fragments, representing those genes expressed in the tester but not in the driver, are amplified using primers specific to the adaptor sequences ligated onto the fragments before subtraction (panel b, bottom). Tester-driver and driver-driver hybrids will not amplify, as driver fragments lack the adaptor fragments containing the primer-matching sequence (panel b, bottom). The resulting PCR product, which may contain dozens or even hundreds of amplified fragments, thus represents genes expressed in the tester but not in the driver.

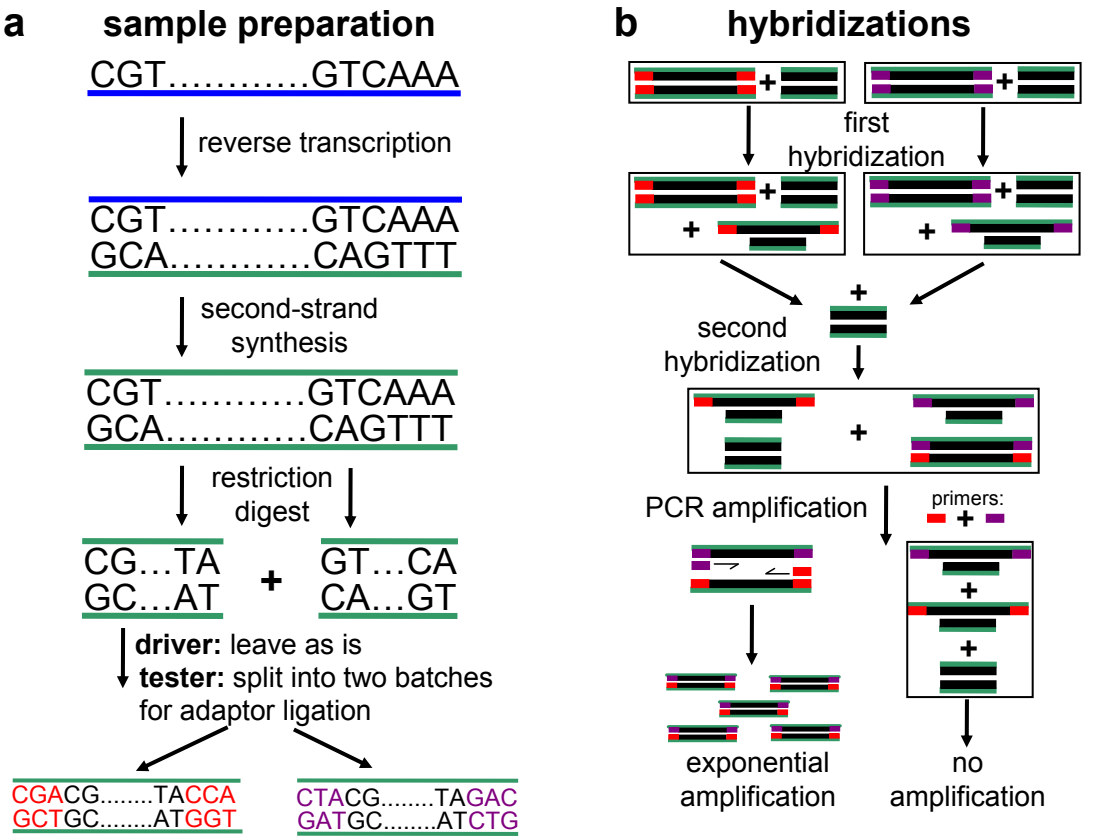

Figure 1.4. Pictoral diagram of the suppression subtractive hybridization technique.

Electrophoresis of the amplified product follows, and bands may be excised, reamplified, cloned, and sequenced. Unlike dd-PCR and cDNA-AFLP, any band which appears on an SSH gel is potentially of interest, as each band represents a gene expressed (or expressed more) in one sample than the other. After sequencing, a gene of interest may be confirmed as differentially expressed using QRT-PCR.

1.8.3 Box 3: Detailed Methods: EST Comparisons

Comparing **expressed sequence tag** (EST) libraries is a straightforward yet powerful way to generate a list of genes which may be differentially expressed between two or more samples. However, EST libraries with extensive sequence data must be available for the samples of interest, and must be generated if they do not already exist. Here we briefly describe the methods for making an EST library and define the key terms associated with this method.

As with other methods described above, RNA is first extracted and then reverse-transcribed into cDNA from the samples of interest. The double-stranded cDNA pools may first be enriched for full-length sequences, or inserted directly into an appropriate bacterial vector. A **cDNA library**, therefore, is defined as a large pool of bacteria, with each clone containing an inserted cDNA fragment. The cDNA library is plated, grown to produce numerous colonies, and individual clones picked for single-read sequencing (panel a).

To obtain a complete EST library, thousands of clones must be picked and sequenced. Next, the thousands of single-pass sequences are imported into sequence-analysis software and aligned. Alignment algorithms group and align sequences based on sequence homology; each group of sequences is called a **ontigs** (panel b). The combined end-to-end sequence for each ontigs is known as the **consensus sequence**, and forms a **unigene**, a transcript identified by numerous ESTs. An EST that doesn't align with any other sequence is known as a **singleton**.

Once all of the unigenes and singletons are identified in each cDNA library, the number of ESTs with which each unigene is composed can be counted, and used as a proxy for amount

of gene expression; a gene with a relatively higher copy number should appear more often in a cDNA library. Next, the consensus unigenes between each library are compared, identifying which genes are expressed in both libraries. For those which appear in both libraries, the numbers of ESTs comprising that unigene in each library are compared (panel c). Genes which are expressed equally should have similar copy numbers (panel c, left side), while genes that are differentially expressed will have unequal copy numbers (panel c, right side) and may not appear in one of the EST libraries if it is not expressed in that sample. Several computational methods for statistical analysis of EST copy number have been described (e.g. Audic and Claverie 1997, Stekel et al. 2000, Ribeiro 2003) and will not be detailed here.

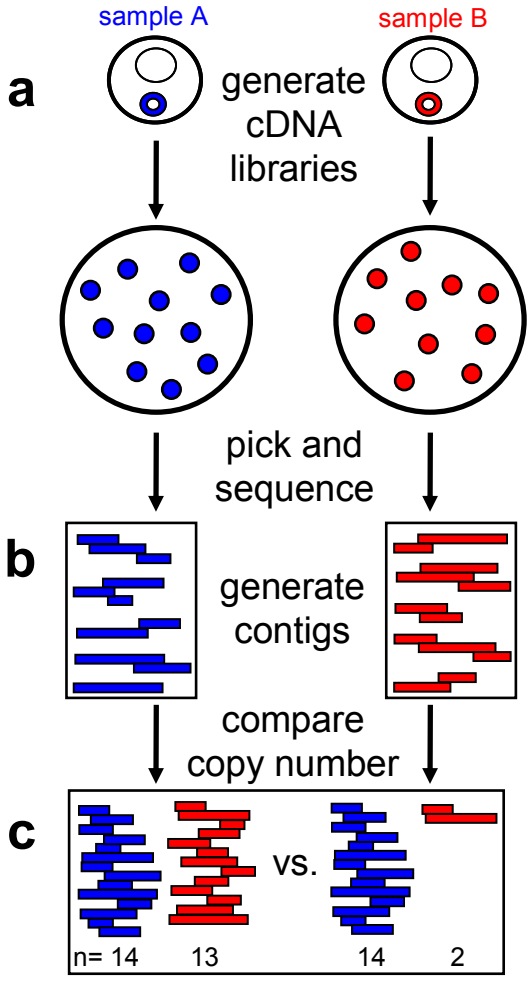

Figure 1.5. Pictoral diagram of the comparative EST technique.

1.8.4 Box 4: Detailed Methods: Serial Analysis of Gene Expression (SAGE)

SAGE is restriction enzyme-based method for simultaneously cloning, sequencing, and measuring gene expression. As in other methods, RNA is first isolated from samples of interest and first- and second-strand cDNA synthesized (panel a, top). Next, double-stranded cDNA is digested with the restriction enzyme ***Nla*III**, which is a 4-base cutter which cuts DNA frequently. The 3' ends of the cDNA fragments are collected with **streptavidin beads** (which bind poly-A tails), and a **linker** fragment ligated to the *Nla*III splice site (panel a, middle). The linker fragments contain a ***Fok*I** cleavage site immediately adjacent to the *Nla*III site. Fragments are then digested with *Fok*I, which cleaves exactly 13bp downstream from the recognition site (panel a, bottom). These 14bp fragments are known as **tags**, and theoretically are unique for each gene (Velculescu et al. 1995).

Blunt-end ligation of cleaved fragments follows, resulting in the formation of **ditags**, a fragment with two tags linked tail-to-tail (panel b, top). These ditags are then amplified using primer sequences matching those of the linker sequence; the resulting PCR product is then cleaved with *Nla*III to remove the linker fragments (panel b, middle). This cleavage results in ditag fragments with 4-base overhanging ends, which are then ligated together in a process termed **concatenation**, to generate fragments with 10-50 ditags each (panel b, middle). These concatenated fragments are cloned, sequenced, and sequenced tags matched to existing sequence databases (panel b, bottom), using software designed for this purpose (e.g., SAGEmap from NCBI; http://www.ncbi.nlm.nih.gov/projects/SAGE/). SAGE software packages analyze existing sequence data for the species of interest to identify cDNA fragments expected to result from restriction digests with *Nla*III and *Fok*I (Lash et al. 2000, Pleasance et al. 2003).

Once gene identification is made for all tags for which identification is possible (~90% on average; Lash et al. 2000), the number of times each gene's tag is found is calculated. For a comparative study, these values are then statistically compared between samples of interest

(e.g. Audic and Claverie 1997). Genes identified as being potentially differentially expressed with SAGE are then usually confirmed with QRT-PCR and/or Northern blots.

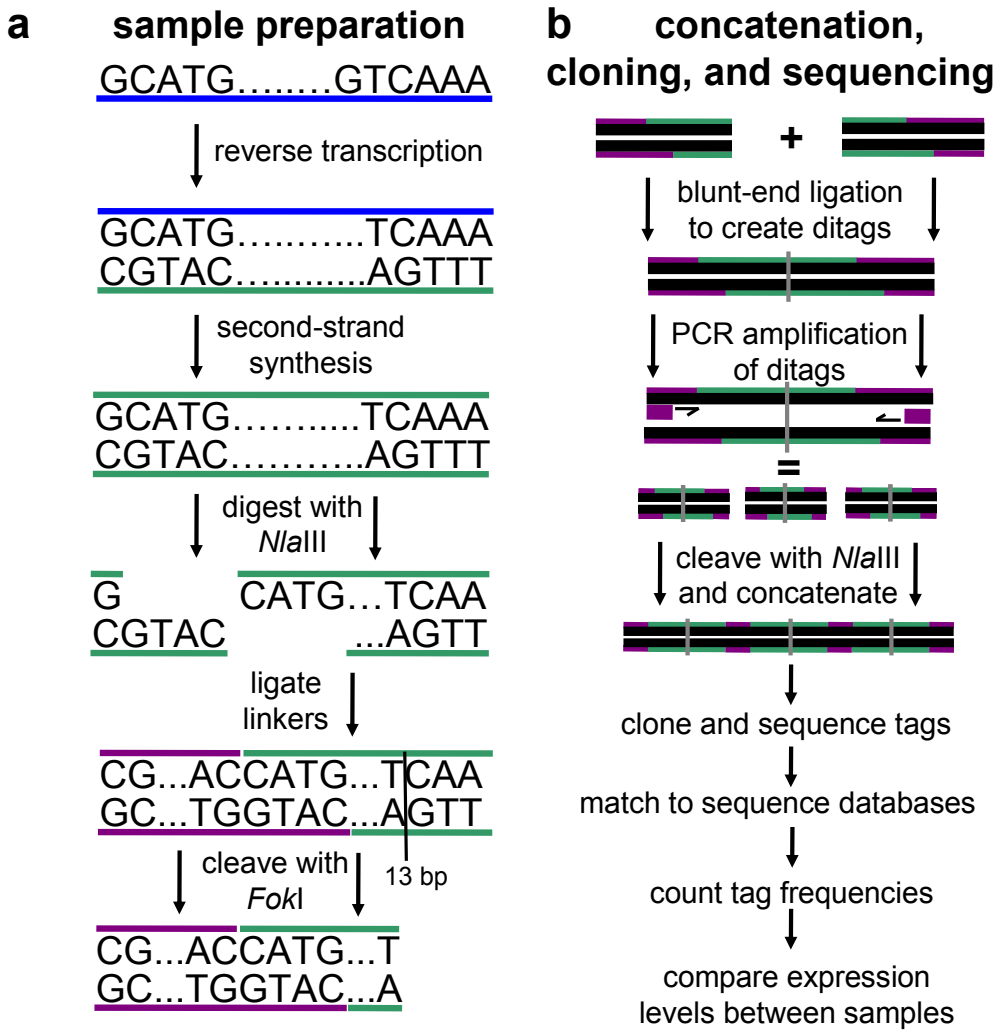

Figure 1.6. Pictoral diagram of the SAGE technique.

CHAPTER 2

GENE-EXPRESSION DIFFERENCES BETWEEN DIAPAUSING AND DIRECT-DEVELOPING EMBRYOS OF THE STRIPED GROUND CRICKET, *ALLONEMOBIUS SOCIUS*: IDENTIFICATION AND GENE ONTOLOGY

2.1 Abstract

Variation in environmental conditions can lead to the evolution of phenotypic plasticity, as individuals within a population able to respond to changing conditions will have relatively higher fitness than those that cannot. Environmentally-induced phenotypic plasticity has been documented in a wide range of species, and identifying the mechanisms which underlie the evolution and maintenance of plasticity is an active area within evolutionary biology. Finding the genetic basis of adaptive plasticity is of particular interest to researchers, and is now becoming increasingly possible as gene-expression techniques continue to be developed. Furthermore, many gene-expression techniques now allow genes to be identified and expression levels measured in any system, reducing dependency on traditional genetic models. The striped ground cricket, *Allonemobius socius*, provides an ideal model system in evolutionary ecology, as it has a wide geographic range with documented natural variation in many life-history traits which correlate to environmental variation across the species' distribution. I used suppression subtractive hybridization (SSH) to identify genes which are differentially-expressed between diapausing and direct-developing eggs in *A. socius*. Here, I report gene ontology findings for both types of eggs, describe differences in the type of genes which are expressed in each developmental type, and relate my findings to previous studies identifying genes associated with diapause in other insect species.

2.2 Introduction

Many organisms are able to adjust to changes in their environment within their lifetime, often through plastic developmental responses including morphological, behavioral, and life-history changes (Stearns 1992; Schlichting and Pigliucci 1998; Roff 2002). Because these changes often have a genetic basis (West-Eberhard 2003; Carroll et al. 2005), they may be studied using a gene expression approach. Environmental conditions experienced during development can alter the expression level of certain genes to yield an environment-specific phenotype (Berrigan and Charnov 1994; Schlichting and Pigliucci 1998). Thus, environmentally-contingent gene expression is the likely cause of much of the adaptive plasticity observed in natural systems.

In insects, the environmental induction of embryonic diapause instead of direct development is an example of adaptive plasticity: under unfavorable (cold or dry) conditions, it may be best to initiate diapause and delay hatching until conditions have become more favorable. However, if current conditions are good (warm and wet), direct-development allows exploitation of these favorable conditions, resulting in increased fitness via shortened generation time and/or additional generations (Via and Lande 1985; Philippi and Seger 1989; Mousseau and Fox 1998). Diapause can occur in the egg, larval, pupal, or adult stage of the life-cycle, but a given species can typically undergo diapause in only one of these stages (Danks 1987; Masaki 1996; Rossiter 1996). In species that experience egg diapause, the initiation of diapause is believed to be under maternal control (Tanaka 1986a; Mousseau and Dingle 1991; Hockham et al. 2001), although environmental conditions may directly influence embryonic development (Shiga and Numata 1997; Olvido et al. 1998; Huestis and Marshall 2006a). Determining which environmental conditions lead to diapause initiation has been the focus of dozens of studies (e.g., Mousseau and Dingle 1991; Mousseau and Fox 1998), and the mechanism of diapause induction has also been the subject of much recent research, although it remains unknown for most species (Denlinger 2002).

Recently, several genes have been identified as up- or down-regulated during diapause, but many of these studies have focused on species which undergo diapause late in development, such as the pupal or adult stage (Denlinger 2002). Examples of known up-regulated genes include antifungal peptides (Daibo et al. 2001), storage proteins (Denlinger 2002), stress proteins (Denlinger et al. 1995, 2001; Flannagan et al. 1998; Moribe et al. 2001), diapause hormone (Su et al. 1994; Yamashita 1996; Yamashita et al. 2001), transcription factors (Suzuki et al. 1999), lipids (Michaud and Denlinger 2006), and several unidentified proteins (Yocum 2003). Genes down-regulated during diapause include actin (Denlinger et al. 1995) and elastin-like proteins (Flannagan et al. 1998). One gene recently found to be specifically associated with the breaking of diapause in pupal sweet potato hornworms is cytochrome *c* oxidase subunit I (Uno et al. 2004). Most of the genetic research on diapause has been conducted in species with larval, pupal, or adult diapause, and little is known about the genetic basis of embryonic diapause (Denlinger 2002), making it an ideal area for further study.

The induction of embryonic diapause in the striped ground cricket, *Allonemobius socius*, has been well-studied. For example, previous research has shown that diapause in the *A. socius* complex is influenced by maternal effects, egg-incubation temperature, photoperiod, and moisture (Tanaka 1984, 1986a,b, 1987, 1992; Mousseau and Roff 1989a; Mousseau 1991; Bradford and Roff 1993, 1995, 1997; Olvido and Mousseau 1998; Olvido et al. 1998; Roff and Bradford 2000; Huestis and Marshall 2006a). However, as in many species in which diapause has been studied at the phenotypic level (Denlinger 2002), the molecular genetic basis has not yet been characterized.

The breaking of embryonic diapause in response to direct environmental cues can be viewed as an adaptive mechanism which allows individuals to exploit favorable changes in environmental conditions. To explore this area further, I studied how one key environmental variable, egg-incubation temperature, affected embryonic gene expression and developmental phenotype (i.e., diapausing or direct-developing) utilizing the cricket *Allonemobius socius* as a model system. Specifically, I used suppression subtractive hybridization (SSH) to identify genes

which are differentially expressed during the adaptive breaking of diapause as a direct result of changing environmental conditions. Here, I report results for 28 candidate genes identified as differentially-expressed between diapausing and non-diapausing eggs using SSH, describe the gene ontology (GO, *sensu* Ashburner 2000; Harris et al. 2004) of these two samples, and relate my findings to previous research on the genetics of diapause in other insect species.

2.3 Methods

2.3.1 Experimental System

Striped ground crickets (*Allonemobius socius* complex) have a wide distribution in North America, with much geographic and environmental variation across this range, and thus provide an ideal system for studying geographic variation in life-history traits. Their ease of capture in the field and maintenance in the laboratory also makes crickets in this complex well-suited for laboratory manipulation, and they have been used as a model for research on many topics within evolutionary biology and ecology for over 70 years (e.g., Fulton 1931; Howard 1983, 1986; Britch et al. 2001; Fedorka and Mousseau 2004). For example, they have been used to study speciation and hybrid-zone dynamics (e.g., Howard 1986; Gregory and Howard 1994; Britch et al. 2001), sexual conflict (e.g., Fedorka and Mousseau 2002a,b, 2004), geographic variation in morphology (e.g., Mousseau and Roff 1989a,b, 1995; Bradford and Roff 1993; Olvido et al. 2003; Fedorka et al. 2007), and environmental effects on life-history traits (e.g., Tanaka 1986; Olvido et al. 1998; Huestis and Marshall 2006a,b).

In particular, *Allonemobius* have been used as a model system for studying environmental and maternal effects on diapause for over 2 decades (e.g., Tanaka 1984, 1986, 1987, 1992; Olvido et al. 1998; Huestis and Marshall 2006a; Winterhalter and Mousseau 2007). Previous research has shown that diapause occurrence is clinal in *Allonemobius*, with individuals from univoltine, northern populations producing larger proportions of diapausing eggs (Tanaka 1986a; Mousseau and Roff 1989a; Mousseau 1991; Winterhalter and Mousseau 2007). There is also natural variation in the number of generations per year, as there is a shift from univoltinism in the northern part of their range, to bivoltinism in the middle of the range, to

multivoltinism at the southern extent of the range. Bivoltine populations display individual plasticity (Winterhalter and Mousseau 2007), as individuals can adopt either a univoltine or multivoltine strategy depending on the environmental conditions they experience (Mousseau and Roff 1989a; Bradford and Roff 1997; Winterhalter and Mousseau 2007). However, oviposition environment appears to play only a minor role in diapause occurrence relative to embryonic incubation environment (Olvido et al. 1998; Huestis and Marshall 2006a). Other factors including maternal age, photoperiod, and moisture availability have also been shown to affect diapause in laboratory experiments (Tanaka 1984, 1992; Mousseau 1991; Bradford and Roff 1995). Although much is known about the environmental and maternal facts that affect diapause in *A. socius*, the genetic basis of this trait is not yet known, making it an ideal avenue for further study. The genetic basis of diapause is not yet known for most species, especially for species which diapause in the egg stage (Denlinger 2002); this fact, combined with much previous knowledge about diapause in this species, makes *Allonemobius* an ideal system for this research.

2.3.2 Experimental Animals

Crickets were collected from a wild population located in upstate South Carolina (Latitude: 34.4869°N; Longitude: 83.0229°W) in July 2006. The population (founded from ~50 wild-caught individuals) was maintained in the laboratory under standard rearing conditions for *Allonemobius* (27°C, 14:10 L:D; Huestis and Marshall 2006a,b) for 2 generations to remove any environmental or maternal effects from the field, prior to use in subsequent experiments. Laboratory stocks consisted of $\geq$ 50 individuals to maintain adequate population size and genetic diversity.

A split-brood design was used to generate batches of sibling eggs reared under two egg-incubation temperatures: 27 and 32°C. To generate these egg batches, crickets were raised in sex-specific cages as juveniles to prevent mating, under standard conditions as described above. After reaching adulthood, crickets were allowed to mature for 7-10 days, placed into a mating cage, and allowed to mate for 7 days. Females were then placed in

individual boxes at 27°C and provided strips of rolled cheesecloth in 60 mm Petri dish lids in which to lay eggs. Cheesecloth was collected each day, and new cheesecloth provided. Eggs were collected from females every day for 8-10 days, providing multiple batches for most females over her reproductive lifespan. The removed cheesecloth was examined immediately for the presence of eggs, and if $\geq$ 10 eggs were present, batches were randomly split between the original egg-incubation temperature of 27°C and an increased egg-incubation temperature, 32°C, designed to stimulate the breaking of diapause. Photoperiod was held constant at 14:10 L:D for all treatments. This design resulted in batches of eggs being split between the two egg-incubation temperature treatments within 24 hours of being laid.

Egg batches were maintained at their assigned egg-incubation temperature for an additional 4 days. At this point, batches with large numbers of eggs ($\geq$ 20 per batch) had some eggs carefully removed from the cheesecloth and frozen in RNA*later* (Ambion, Austin, TX) at -20°C, leaving at least 5 eggs per batch at the assigned egg-incubation conditions. All remaining eggs were maintained under their assigned treatments for 18-20 days, then scored for developmental phenotype. Eggs were scored as either hatched (hatchling cricket present), non-diapause (presence of eyespots, at a minimum), diapause (small and white), or rotten (less than 5% of all eggs show brown or black fungus; counts of rotten eggs not used). For each batch, the proportion of diapausing eggs at each temperature was calculated by dividing the number of diapause eggs by the total number of eggs in the batch. For those batches with eggs frozen for later genetic analyses, diapause proportions of sibling eggs were used as a proxy to score diapause phenotype for those frozen batches of eggs.

Three typical outcomes result from this split-brood design. In one outcome, eggs which are laid and maintained at 27°C will undergo diapause, while sibling eggs which are incubated at 32°C will break diapause and direct-develop (Figure 2.1A). This outcome shows that there is phenotypic plasticity in developmental response to temperature. Alternatively, some batches of eggs direct-develop at both 27 and 32°C (Figure 2.1B), indicating a genotypic or maternal component that drives direct-development at both of these temperatures. In contrast, certain

batches undergo diapause at both egg-incubation temperatures (Figure 2.1C), again indicating a genotypic and/or maternal effect to diapause regardless of egg-incubation temperature. There is variance around each of these three outcomes, and while batches will typically fall into one of these three general categories, most are not 100% of one phenotype or another. However, egg batches closest to desired phenotypic outcomes were chosen for gene-expression analyses (see below).

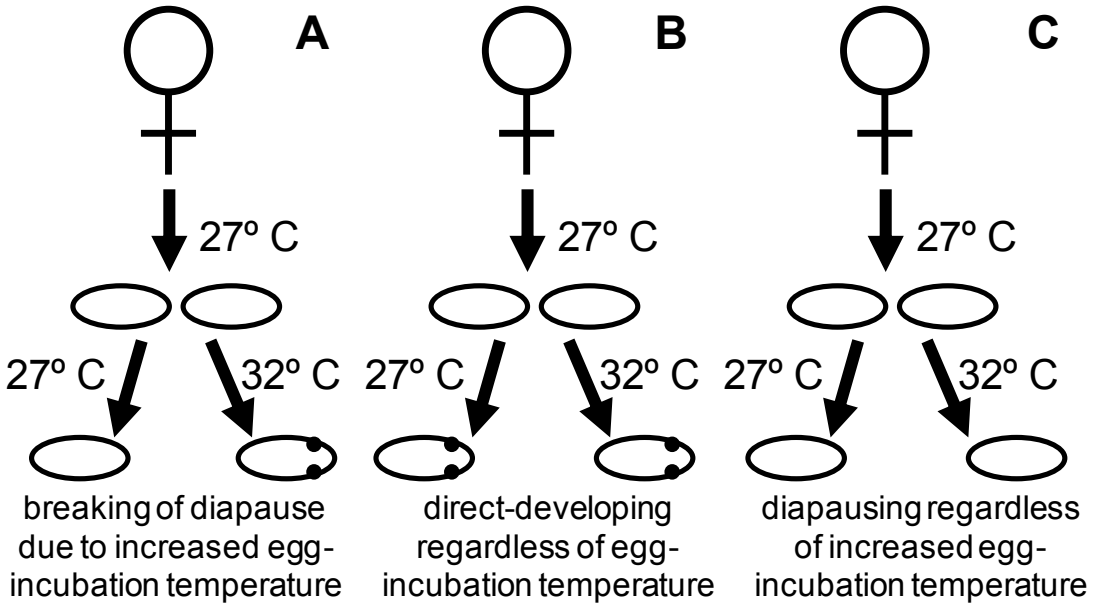

Figure 2.1. Three typical outcomes from split-brood design. In one common result, eggs which are laid and maintained at 27°C initiate diapause, while sibling eggs shifted to an egg-incubation temperature of 32°C break diapause and undergo direct development (A); this result indicates that eggs are directly affected by changes in egg-incubation temperature and can respond to favorable environmental conditions (an increase in temperature) by breaking diapause. In contrast, eggs which are laid at 27°C and maintained at 27°C, as well as their siblings moved to 32°C may both undergo direct development (B); this outcome may result from genetic or maternal effects on diapause and indicates that eggs do not change developmental phenotype based on egg-incubation temperature, at least across this range of temperatures. Finally, eggs laid at 27°C and incubated at either 27 or 32°C may both initiate diapause (C); as in B, this result may be due to genetic and/or maternal effects, and also indicates a lack of plastic response to egg-incubation temperature.

2.3.3 Suppression Subtractive Hybridization

To identify genes which are differentially-expressed between diapausing and non-diapausing eggs, I conducted subtractions between two specific pairs of samples. The first

subtraction is between a pair of sibling eggs which diapaused at 27°C but broke diapause at 32°C (see Figure 2.1A). This comparison reveals changes in gene-expression due to the different development phenotypes and also effects of temperature, while controlling for genetic and maternal effects. The second subtraction is between eggs which broke diapause at 32°C (the same sample as the first comparison describe above) and unrelated eggs which diapaused at 32°C (from a batch of eggs which diapaused at both temperatures; see Figure 2.1C). This comparison controls for effects of temperature, and reveals differences in expression due to developmental phenotype and genetic/maternal effects. Thus, genes which are found to be differentially-expressed in both subtractions provide the strongest candidates.

For each of these three samples, total RNA was isolated from the appropriate batches of eggs (*n*~25 per sample) using an RNAqueous-4PCR kit (Ambion) following manufacturer's recommended protocol, including DNase I treatment to remove contaminating genomic DNA. Total RNA was quantified using a NanoDrop-1000 spectrophotometer (Thermo Scientific, Wilmington, DE). 185 ng of total RNA was used to generate double-stranded complementary DNA (dscDNA) using a Super SMART cDNA synthesis kit (Clontech Laboratories, Mountain View, CA) following manufacturer's protocol with 23 PCR cycles for second-strand synthesis. The resulting double-stranded cDNA was then used for suppression subtractive hybridization (SSH) using a PCR-Select cDNA synthesis kit (Clontech Laboratories) following recommended protocol. For this technique, samples are labeled with adaptor fragments as either a "tester" or a "driver," and driver cDNA is subtracted from tester cDNA, followed by PCR to amplify those genes which are uniquely expressed in the tester sample (Diatchenko et al. 1996, 1999; see also Chapter 1 for additional details of this technique). A reverse subtraction is necessary to detect gene-expression differences in both directions, as each subtraction only reveals genes expressed in the tester, not in the driver.

Resulting PCR products from amplifying SSH products were electrophoresed on a 2% agarose gel with ethidium bromide in 1X TBE buffer for 4 hours to resolve banding patterns. For each subtraction, five brightly-appearing bands were carefully excised, gel-extracted using a

QIAquick gel extraction kit (Qiagen, Valencia, CA), and the resulting DNA concentration quantified using a NanoDrop-1000 spectrophotometer (Thermo Scientific). Cleaned PCR products were then cloned with using a TA cloning kit, pCR2.1 vector, and one-shot INVαF' chemically-competent cells (Invitrogen, Carlsbad, CA), following all manufacturer's recommended protocols. Transformed bacteria were grown overnight at 37°C and five positive colonies were picked for each original gel-extracted PCR product. Colonies were then homogenized in 50ul of nuclease-free water and subjected to PCR with standard M13 primers (M13F: 5' GTTTTCCCAGTCACGAC 3'; M13R: 5' CAGGAAACAGCTATGACC 3') under standard PCR conditions in 50ul total volume reactions. Resulting products were electrophoresed on a 1% agarose gel with ethidium bromide in 1X TBE and bands excised and gel extracted as above. These cleaned PCR products were then used as sequencing templates.

2.3.4 Sequence Analysis

DNA sequencing was performed at the University of Kentucky Advanced Genetic Technologies Center (Lexington, KY), in both forward and reverse directions using M13 primers as given above. Sequences were cleaned and vector sequences trimmed off using BioEdit 7.0.5.3 (Hall 1999). Edited sequences were imported into Sequencher 4.8 (GeneCodes, Ann Arbor, MI) and contigs generated using stringent assembly parameters (35 bp minimum overlap and 75% minimum identity). The resulting uniESTs were then analyzed for gene ontology (GO) using Blast2Go software (Conesa et al. 2005; Götz et al. 2008). Differences between the two developmental types in the numbers of sequences in each gene ontology category were analyzed using a Fisher's exact test within Blast2Go; the *P*-value was set at $\alpha \leq 0.1$ due to low power in detecting differences with small numbers of sequences in each category. Diversity in types of genes which were expressed within each of the 3 major GO categories was calculated using Shannon's diversity index and differences between the two developmental types were tested using a *t*-test following the method described by Magurran (1988).

2.4 Results

2.4.1 Developmental Phenotypes

Within the population used for this experiment, I observed batches of eggs falling in all three types of phenotypic outcomes (see Figure 2.1), and the majority of batches showed some developmental response to temperature (Figure 2.2). These results indicate that this population, like many southern populations of *A. socius*, displays large amounts of genetic variance for diapause mean and plasticity in response to egg- frozen incubation temperature (Figure 2.2). As a result, I was able to select egg samples from batches which displayed the desired outcomes: one batch which, when split, diapaused at nearly 100% frequency at 27°C but direct-developed at 100% frequency at 32°C (Figure 2.1A), and another batch in which all eggs incubated at both temperatures diapaused (Figure 2.1C). These three samples (sibling eggs as in Figure 2.1A and unrelated eggs which diapaused at 32C as in Figure 2.1C) were then used for suppression subtractive hybridization to identify candidate genes which were expressed in one sample but not in the other.

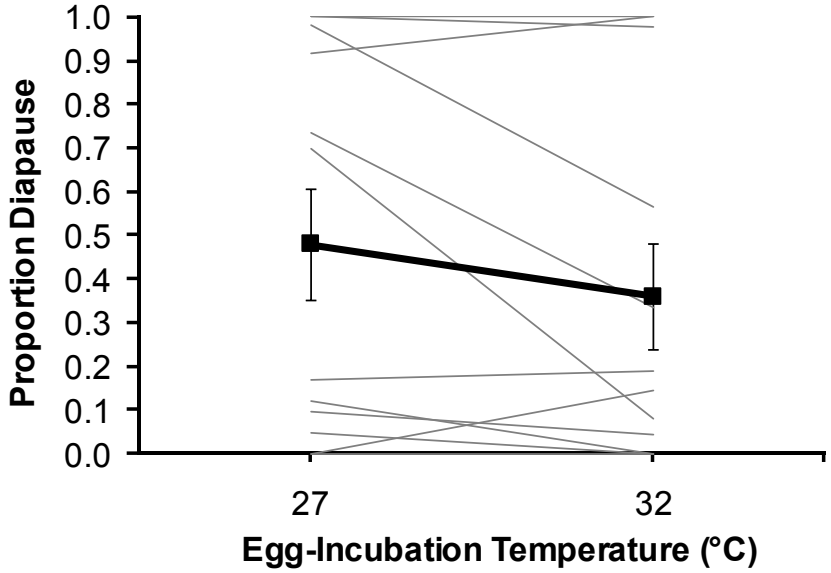

Figure 2.2. Combined results from all females in this population which produced eggs during this experiment. These data display the phenotypic plasticity of *Allonemobius socius* embryos in response to egg-incubation temperature, and also show the wide range of variance at the population-level for this phenotype. Single grey lines represent eggs produced by a single female, while the black squares, black line, and error bars represent the sample means and standard error.

2.4.2 Gene Ontology

A total of 44 unique genes (31 singletons and 13 contigs) resulted from cloning and sequencing 64 SSH fragments in both directions (Table 2.1). Of these, 28 (18 singletons and 10 contigs) could be identified using blastx and/or tblastx with a significance cutoff of e^{-05} (Table 2.1). The numbers of sequences which were identified as belonging to specific gene functions within each of the three main functional GO categories for both diapausing and non-diapausing eggs are shown in Figure 2.3.

Table 2.1. Distributions of numbers of genes resulting from cloning and sequencing subtractive hybridization PCR products, which compared diapausing and non-diapausing eggs of *Allonemobius socius*.

Total resulting genes		
	Diapause	Non-Diapause
Singletons	18	13
Contigs	4	9

Genes identified via BLAST		
	Diapause	Non-Diapause
Singletons	11	7
Contigs	4	6

In general, more types of genes are expressed in eggs which have broken diapause than in their diapausing counterparts (Figure 2.3). A Fisher's exact test can be used to compare the number of sequences between samples, a proxy for the amount of difference in expression; however, power to detect differences was extremely low with so few sequences, so I used a *P*-value of $\alpha < 0.1$ for this exploratory study. Within the cellular component ontology category, three classes of genes were significantly more represented in the eggs which broke diapause compared to diapausing eggs using this criterion (Figure 2.3A). However, there was not a significant difference in the diversity of types of sequences expressed in this GO category (Table 2.2).

In the biological process GO category, there are 5 classes of genes which are represented by one or two sequences in the non-diapause eggs but not represented in the diapausing samples (Figure 2.3B); this difference was not statistically significant using a Fisher's exact test. However there was a significant difference in the Shannon diversity indices between diapausing and non-diapausing eggs (Table 2.2), indicating that there is more diversity in the types of biological process genes expressed by eggs that have broken diapause relative to their diapausing counterparts. In the molecular function GO category, there was one type of gene expressed by diapausing eggs that was not detected in the direct-developing eggs and two types of genes found in direct-developing eggs but not in diapausing eggs (Figure 2.3C), but these differences were not significant with either a Fisher's exact test or by comparing Shannon diversity indices (Table 2.2).

Table 2.2. Shannon diversity indices for each developmental type (diapause vs. non-diapause) in each of the three major gene ontology (GO) categories. *H* is Shannon's diversity index, and a t-test was used to test for differences between the two developmental types. Variance, *H*, *t*, and df were calculated using the methods of Magurran (1988).

Cellular Component					
Developmental Type	*H*	variance	*t*	df	*P*
Diapause	1.7041	0.00006	0.1338	37.66	0.894
Non-diapause	1.7151	0.00675			

Biological Process					
Developmental Type	*H*	variance	*t*	df	*P*
Diapause	1.2555	0.00828	6.595	33.81	< 0.0001
Non-diapause	2.0286	0.00546			

Molecular Function					
Developmental Type	*H*	variance	*t*	df	*P*
Diapause	1.1973	0.01314	0.319	28.92	0.752
Non-diapause	1.2622	0.02818			

2.4.3 Genes Identified

A total of 22 sequences (singletons and contigs combined) were identified as up-regulated in direct-developing eggs relative to diapausing eggs. Nine sequences were unable to

be identified on the basis of BLAST searches, while 13 sequences could be identified via BLAST similarity (Table 2.1). In contrast, 22 sequences were identified as up-regulated in diapausing eggs relative to direct-developing eggs, 15 of which were identified using BLAST (Table 2.1). A list of all genes which were identified using BLAST is given in Table 2.3. Genes found to be upregulated in direct-developing eggs relative to diapausing eggs included two members of the heat-shock protein 70 family (HSP70), apolipophorin III, a cytochrome P450-like protein, several ribosomal proteins, and several predicted and/or hypothetical proteins with unknown function (Table 2.3). Genes found to be expressed in diapausing eggs but not in eggs which had broken diapause included heat-shock protein 20.7, an arginine kinase, period, a pupal cuticle protein, several ribosomal proteins, and several predicted proteins (Table 2.3).

Table 2.3. Genes identified as being differentially-expressed between diapausing and non-diapausing eggs in the cricket *Allonemobius socius* using SSH. Developmental type is which type of egg the sequence was identified in, while sequence type denotes whether the sequence appeared a single time or was assembled into a contig. Length given is number of base pairs in the sequenced fragment, eValue is the significance of the BLAST match, and similarity is the percentage match between the SSH fragment and its best BLAST homolog.

Developmental Type	Sequence Type	Matching Sequence Description	Length (bp)	eValue	Similarity
Breaking	singleton	hypothetical protein	635	5.03E-36	70.7%
Breaking	singleton	similar to DEAD box polypeptide 5	273	1.70E-38	95.9%
Breaking	singleton	60S ribosomal protein L8	333	7.89E-36	92.4%
Breaking	singleton	similar to putative QM protein	646	6.44E-91	97.1%
Breaking	singleton	ribosomal protein S3Ae	346	1.68E-33	94.2%
Breaking	singleton	putative translationally controlled tumor protein	662	6.24E-52	86.3%
Breaking	singleton	similar to CG3195-PA, isoform A	570	1.00E-64	93.1%
Breaking	contig	hsp70 family member	958	7.81E-147	96.8%
Breaking	contig	hypothetical protein AaeL_AAEL003425	635	4.03E-106	97.3%
Breaking	contig	GF23497	599	1.05E-17	88.1%
Breaking	contig	cytochrome P450 like_TBP	1144	3.77E-24	86.6%
Breaking	contig	hsp70 family member	626	5.02E-29	93.0%
Breaking	contig	apolipophorin-III	917	4.28E-30	58.0%
Diapause	singleton	putative ribosomal protein S23e	515	2.43E-71	97.5%
Diapause	singleton	obstractor C isoform 2	431	2.70E-76	91.6%
Diapause	singleton	similar to ribosomal protein L34 isoform 2	397	8.29E-49	87.2%
Diapause	singleton	similar to ribosomal protein S27a	637	9.85E-60	95.9%
Diapause	singleton	heat shock protein 20.7	611	1.23E-45	68.8%
Diapause	singleton	ribosomal protein L31e	476	4.05E-56	89.6%
Diapause	singleton	similar to pupal cuticle protein 78E	339	2.54E-10	60.4%
Diapause	singleton	receptor for activated protein kinase C-like	644	2.58E-100	68.6%
Diapause	singleton	zinc finger protein	615	2.91E-50	56.8%
Diapause	singleton	similar to GA10081-PA	624	9.49E-20	66.3%
Diapause	singleton	period	517	3.80E-08	71.0%
Diapause	contig	similar to C-type lectin	959	7.40E-84	78.7%
Diapause	contig	hypothetical protein AaeL_AAEL014418	527	4.93E-14	61.2%
Diapause	contig	CSP precursor	486	6.19E-12	55.5%
Diapause	contig	arginine kinase	515	5.75E-89	95.6%

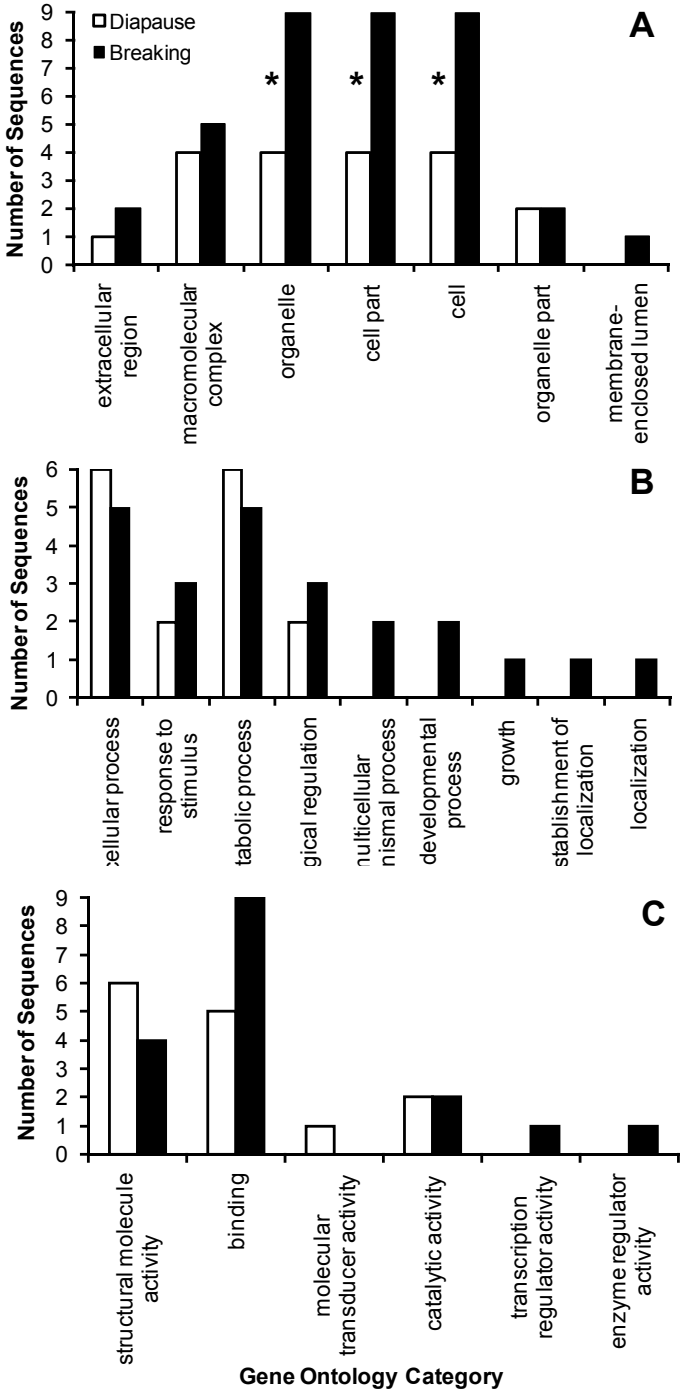

Figure 2.3. Gene ontology of genes which were identified as differentially-expressed between diapausing (open bars) and non-diapausing (solid bars) *A. socius* embryos. Three gene ontology categories are shown: cellular component (A), biological process (B), and molecular function (C). Asterisks signify differences in expression of genes in a given ontology category which were significant at $P<0.1$ using a Fisher's exact test (note that genes significant at a lower P-value are reported, due to low power to detect differences using only 28 sequences).

2.5 Discussion

In this study, I used suppression subtractive hybridization (SSH) to identify genes which were differentially-expressed between diapausing eggs and eggs which had broken diapause and initiated direct-development, using the cricket *Allonemobius socius* as a model for insect diapause. I was able to find potential gene identities for 28 out of 44 sequences (64%) generated using this method. Several of these are putative or predicted genes, while others were very strong matches to known sequences of known identity and function. Future work on *A. socius* and other insect species will likely allow easier gene identification in the next several years, as more sequence data are generated. Below, I detail several genes which I identified as having strong BLAST matches to genes of known function, and relate my findings to other studies on gene expression in insect diapause.

2.5.1 Genes Upregulated in Direct-Developing vs. Diapausing Embryos

Among those genes which could be identified with BLAST were two separate contigs with very high (~96%) sequence similarity to heat-shock protein 70 (HSP70) family members in the locust (*Locusta migratoria*; GenBank accession AAO21473) and silkworm (*Bombyx mori*; GenBank accession AB084922; Table 2.3). Proteins in the HSP70 family have been previously identified in studies of insect diapause, although they are generally found to be upregulated in diapausing individuals relative to non-diapausing individuals, in contrast to my findings here. For example, in the flesh fly *Sarcophaga crassipalpis*, HSP70 was found to be upregulated in diapausing pupae and early in the termination of diapause ($\leq$ 6 hours post-termination), but not later in development after breaking diapause (Rinehart et al. 2000). Additionally, HSP70 was found to be a critical component of cold-shock during diapause in both flesh flies and mosquitoes (genus *Culex*; Rinehart et al. 2006, 2007), and in the heat-shock response of diapausing maggots (*Rhagoletis mendax*; Teixiera and Polavarapu 2005). In contrast to other studies showing increased expression of HSP70 proteins during diapause, HSP70 expression was found to be identical and equally low in both diapause and non-diapause blow fly (*Lucilia sericata*) larvae (Tachibana et al. 2005), so its expression during either diapause or direct-

development is not universal. Unlike previous studies, I found that HSP70 expression was increased in direct-developing eggs incubated at 32°C, relative to diapausing eggs incubated at both 27 and 32°C. Because the SSH method removes genes which are expressed in both samples, it reduces the possibility that HSP70 expression resulted from the shift in temperature from 27 to 32°C. Specifically, if HSP70 expression was initiated in response to temperature alone, I would have expected to find it expressed only in the hybridization resulting from eggs incubated at 27 and 32°C, and not in the hybridization between diapausing eggs incubated at 32°C and eggs which had broken diapause when incubated at 32°C. Thus, the finding that it was expressed in both SSH samples may therefore provide strong evidence of its association with breaking diapause and/or the initiation of development in *A. socius* embryos, not simply a response to developmental temperature.

Another gene found to be upregulated in direct-developing embryos was apolipophorin-III (Table 2.3). Apolipophorin-III is one of the major hemolymph proteins in insects, and its primary function is to metabolize and transport fat (Sun et al. 1995; Kim et al. 2004). Therefore, it is not surprising that it was found to be upregulated in developing eggs, relative to diapausing eggs, and may indicate that the developing embryo is either metabolizing fat provided by the female in the yolk or has started developing hemolymph. However, to my knowledge, this is the first report of apolipophorin-III in a study comparing diapausing and direct-developing eggs, making it a unique and novel finding of this study.

Cytochrome P450 was also upregulated in eggs which had broken diapause, relative to diapausing eggs. There are many forms of cytochrome oxidase, and these perform an array of functions within cells. In insects, cytochrome P450 is involved in development, growth, molting, and immune function (Scott and Wen 2001; Niwa et al. 2004), so it is therefore also not surprising that this type of gene would be found in developing embryos. Indeed, previous studies in other insect species (e.g., Horike and Sonobe 1999; Maeda et al. 2008; Yang et al. 2008) and in the nematode *C. elegans* (Gerisch and Antebi 2004) have shown P450's to be weakly-expressed or not expressed in diapausing eggs while they are highly expressed during

development and associated with skipping diapause in *C. elegans*. Thus, as in several other insect species, the expression of cytochrome P450 appears to be associated with breaking diapause and/or initiating development in the cricket *A. socius*, making it an interesting candidate gene for further study.

2.5.2 Genes Upregulated in Diapausing vs. Direct-Developing Embryos

One gene found to be expressed in diapausing but not direct-developing eggs showed highly significant similarity (e^{-45}) to a small heat-shock protein, HSP20.7, in the locust (*Locusta migratoria*; GenBank accession ABC84494; Table 3.3) and other insect species (GenBank accession numbers ABM55532, ABE57137, and ABD98776). Small heat-shock proteins have been found to be upregulated during diapause in several insects, including *Culex pipiens* (Robich et al. 2007), flesh flies (Flannagan et al. 1998; Yocum et al. 1998; Rinehart et al. 2007), and the crustacean *Artemia franciscana* (Qiu et al. 2007). Heat-shock proteins protect other molecules inside of cells from damage during stress events, such as extreme heat or cold (Yocum et al. 1998; Rinehart et al. 2007). Thus, upregulation of small heat-shock proteins in diapausing *A. socius* may serve to protect other cellular components during temperature extremes, such as near-freezing, which may occur during diapause.

I also found an arginine kinase to be upregulated in diapausing eggs, relative to direct-developing eggs. In a protein-analysis study of the crustacean *Artemia sinica*, arginine kinsases were also found to be upregulated in diapausing embryos, known as cysts (Zhou et al. 2008). Arginine kinases have many functions in insects, including detoxification, metabolism, and movement (Gindling et al. 1995; Tanaka et al. 2007). Thus, metabolic proteins may still play an important role in diapausing arthropods, potentially in regulating the very low metabolic rate which must be sustained to maintain diapause.

Lastly, I found several ribosomal proteins to be upregulated in diapausing eggs, a finding also recently found using SSH on diapausing *C. pipiens* adults by Robich et al. (2007). Similarly, many genes identified by SSH in that study (Robich et al. 2007) were not able to be identified with either nucleotide BLAST or translated BLAST searches, and thus are candidate

genes for either diapause or direct-development which must be identified and confirmed using other methods.

2.5.3 Gene Ontology

Gene ontology (GO) is used to classify potential functions of genes identified in EST studies. GO can also be used to look for differences in the amount, type, and diversity of genes expressed when comparing two or more EST databases. In the present study, GO was applied to those sequences which could be identified via BLAST searching. I found that direct-developing eggs had significantly more sequences falling in three types of genes within the cellular component category, and a higher diversity of sequences expressed in the biological process category, relative to diapausing eggs. These data may indicate that a greater variety of genes are expressed in direct-developing eggs, but may also reflect bias of such a small sample size. Further work is needed to quantitatively determine the number of genes and levels of expression of those genes which are differentially-expressed between diapausing eggs and eggs which have broken diapause and initiated direct-development.

Identifying genes which underlie phenotypic plasticity, and other adaptive life-history traits, is important when exploring both how phenotypes may have evolved, and how new phenotypes may evolve under changing environmental conditions. Here I provide a relatively short list of genes which are upregulated in either diapausing eggs relative to direct-developing eggs, or direct-developing eggs relative to their diapausing counterparts. These genes likely represent downstream effects, as the likelihood of finding the exact gene underlying the switch between developmental types is probably quite small. However, further work on these and other genes will continue to provide insights into the evolution of the plastic diapause response in insects.

2.5.4 Future Directions

This study provides a first look into gene-expression differences between diapausing and direct-developing eggs in the cricket *A. socius*, and adds to the list of genes known to be associated with differences in these two developmental phenotypes, providing a list of

candidate genes to explore further. As this study only utilized one population, it will be important to confirm that these genes are differentially-expressed across the range of *A. socius*. Differential expression can be confirmed first with semi-quantitative PCR and then with quantitative real-time PCR (Weisner et al. 1992; Freeman et al. 1999; see also Chapter I). Additionally, once genes are confirmed in multiple populations, they can be sequenced in populations across the range of this species and in populations which vary in diapause occurrence, and tests of selection can be performed on the sequence data to see if these developmental genes are under positive or purifying selection. Lastly, to further confirm gene function and importance, RNAi knockdowns could be attempted for candidate genes, either by maternal RNAi or direct injections into freshly-laid eggs. RNAi has recently been developed for *A. socius* (Marshall et al., *in review*), although maternal and egg RNAi has not yet been attempted. Because RNAi reduces translation of the target gene into its protein product, it is a useful tool for testing gene function, as knockdown of a protein of interest could change the organism's phenotype based on its predicted function. By combining quantitative confirmations and RNAi, candidate genes identified here would be further confirmed and potentially provide avenues for further study on the genetics of diapause and the adaptive, plastic response to break diapause in response to temperature.

CHAPTER 3

THE GENETIC ARCHITECHTURE OF AN ADAPTIVE LIFE-HISTORY REACTION NORM IN A CRICKET

3.1 Abstract

The ability of organisms to respond to environmental cues during development is an important adaptation, particularly in variable or seasonal climates. As such, research identifying genetic pathways underlying phenotypic plasticity and their reaction norms is an active area of study within evolutionary biology. Specifically, genetic linkage of components which make up a reaction norm, such as trait mean and plasticity, may prevent or slow the evolution of new reaction norms. One example of an adaptive, plastic life-history strategy is insect diapause, which occurs in many insect species inhabiting temperate environments. Diapause is highly adaptive, as it allows individuals to respond to unfavorable environmental conditions by initiating a resting phase during their life-cycle for successful overwintering or to break diapause under favorable environmental conditions. Crickets of the *Allonemobius socius* complex are an ideal model system for studying maternal, genetic, and environmental effects on diapause, as there is much climatic variation over their wide geographic range, natural variation between populations in diapause occurrence, and genetic variation within and among populations. Here I survey twelve populations of *A. socius* from throughout the southeastern United States for initiation and plasticity of the diapause phenotype. Further, I used amplified fragment-length polymorphism (AFLP) markers to examine underlying genetic variation between populations and to look for genetic correlates with mean diapause and diapause plasticity. I found significant, independent axes of genetic variation which underlie mean occurrence and plasticity of diapause in *A. socius*, indicating that new reaction norms for this trait may be able to evolve quickly in this species.

3.2 Introduction

For any organism inhabiting an unstable and/or seasonal environment, the ability to detect and respond to cues indicating future environmental conditions is a highly advantageous adaptation. Thus, individuals of many species are able to adjust to changes in their environment within their lifetime, often through plastic developmental responses including morphological, behavioral, and life-history changes (Stearns 1992; Schlichting and Pigliucci 1998; Roff 2002). Phenotypic plasticity is the ability of a single genotype to produce a range of environmentally-induced phenotypes, and the shape of the relationship between environment experienced and phenotype produced by a genotype is known as a reaction norm (Schlichting and Pigliucci 1998). When different genotypes do not produce the same reaction norm, a genotype-by-environment interaction exists – permitting differential selection acting upon individuals with different genotypes.

Most life-history traits are quantitative in nature, and identifying genes which are expressed during plastic development and thus underlie an adaptive reaction norm is an active area of research within evolutionary biology (Bonin et al. 2006; Hoekstra and Coyne 2007). Two major types of approaches are typically used to identify genes associated with life-history phenotypes: gene expression approaches (i.e., differential display PCR, subtractive hybridization, and microarrays; Huestis and Marshall *in press*) and genomic association studies (i.e., genome scans or quantitative trait loci (QTL) mapping; Bonin et al. 2006). QTL approaches utilize molecular markers (e.g., microsatellites, SNPs, etc.) to identify regions of the genome associated with a particular trait of interest, and, for species with a sequenced genome, can be used to identify candidate genes which may then be examined further. For species without a sequenced genome, amplified fragment-length polymorphisms (AFLPs) can be used to generate markers associated with particular phenotypes, and this technique has been used successfully to identify markers associated with phenotypic traits in fish (Robison et al. 2001), honeybees (Rueppell et al. 2004), mosquitoes (Zhong et al. 2006), and crickets (Britch et al. 2007), among others.

Knowledge of the genetic architecture of the components which make up a reaction norm (i.e., shape of the curve, which is composed of the trait mean and slope of the plastic response) allows predictions to be made about the relative speed and ease with which new reaction norms can be produced. Specifically, if the mean and slope of a particular reaction norm are tightly linked, genetically, then the formation of new reaction norms may be slow relative to a reaction norm with unlinked shape components. As populations composed of individuals with genotypes that produce appropriate reaction norms for local environmental conditions will be successful, there may be selection pressure on populations to evolve reaction norms based on their habitat, leading to differentiation even between populations in close geographic proximity if habitats differ (Hoffmann et al. 2005; Liefting and Ellers 2008).

One example of adaptive plasticity is the environmental induction of either embryonic diapause or direct development in insects: under unfavorable (cold or dry) conditions, it may be best to initiate diapause and hatch later when conditions have become more favorable. However, if current conditions are good (warm and wet), direct-development allows exploitation of these favorable conditions, resulting in increased fitness via shortened generation time and/or additional generations (Via and Lande 1985; Philippi and Seger 1989; Mousseau and Fox 1998).

Diapause can occur in the egg, larval, pupal, or adult stage of the life-cycle, but a given species can typically undergo diapause in only one life-history stage (Danks 1987; Masaki 1996; Rossiter 1996). In species that diapause in the egg stage, the initiation of diapause is believed to be under maternal control (Tanaka 1986a; Mousseau and Dingle 1991; Hockham et al. 2001), although environmental conditions may directly influence embryonic development (Shiga and Numata 1997; Olvido et al. 1998; Huestis and Marshall 2006a). Determining which environmental conditions lead to diapause initiation and maintenance has been the focus of dozens of studies (e.g., Mousseau and Dingle 1991; Mousseau and Fox 1998), and the genetic mechanism of diapause induction has also been the subject of much recent research, although it remains unknown for most species (Denlinger 2002).

Crickets of the genus *Allonemobius* provide an ideal model system for many aspects of evolutionary ecology. The induction of embryonic diapause in the *Allonemobius socius* complex is particularly well-studied. For example, previous research has shown that diapause in the *A. socius* complex is influenced by maternal effects, egg-incubation temperature, photoperiod, and moisture (Tanaka 1984, 1986a,b, 1987, 1992; Mousseau and Roff 1989a; Mousseau 1991; Bradford and Roff 1993, 1995, 1997; Olvido and Mousseau 1998; Olvido et al. 1998; Roff and Bradford 2000; Huestis and Marshall 2006a). Diapause occurrence is clinal in *Allonemobius*, with individuals from northern populations producing larger proportions of diapausing eggs (Mousseau and Roff 1989a). This cline mirrors natural variation in the number of generations produced each year: a shift from univoltinism in the northern part of their range, to bivoltinism in the middle of the range, to multivoltinism at the southern extent of the range. Bivoltine populations display plasticity (Winterhalter and Mousseau 2007), and individuals may adopt either a univoltine or multivoltine strategy depending on environmental conditions (Mousseau and Roff 1989a; Bradford and Roff 1997; Winterhalter and Mousseau 2007). However, oviposition environment appears to play only a minor role in diapause occurrence relative to embryonic incubation environment in *A. socius* (Olvido et al. 1998; Huestis and Marshall 2006a), unlike many other species. Other factors including maternal age, photoperiod, and moisture availability have also been shown to affect diapause in laboratory experiments utilizing *Allonemobius* (Tanaka 1984, 1992; Mousseau 1991; Bradford and Roff 1995).

This study surveys variation among 12 populations from throughout the southeastern portion of the range of *A. socius*, where diapause is most plastic (Winterhalter and Mousseau 2007). Specifically, I used a split-brood design to assay the ability of eggs to break diapause in response to favorable environmental conditions (i.e., warmer temperature), and also to measure maternal-age effects on diapause plasticity. Additionally, I generated amplified fragment-length polymorphism (AFLP) markers and used principle coordinate analysis to look for associations between genetic variance and diapause phenotype. I found significant variation within and among populations in mean and plasticity of the diapause trait, both in response to egg-

incubation temperature and maternal age. I also found significant associations between these phenotypic traits and components of genetic variation at the population level. Further, I found that mean diapause and diapause plasticity are associated with different components of genetic variance, allowing for the possibility that they could evolve independently with new reaction norms forming relatively quickly.

3.3 Methods

3.3.1 Study System

Striped ground crickets (*Allonemobius socius* complex) provide an ideal model system for studying the evolution of life-history traits in natural populations due to their widespread distribution, abundance in appropriate habitat, and ease of laboratory maintenance. As such, they have been used to study many aspects of evolutionary biology, including speciation and hybrid zone dynamics (Howard and Waring 1991; Britch et al. 2001; Marshall 2004), life-history evolution (Mousseau 1991; Olvido et al. 1998; Huestis and Marshall 2006a), geographic variation in morphology (Mousseau and Roff 1989a,b, 1995; Fedorka and Mousseau 2002b; Olvido et al. 2003; Marshall 2007), and effects of *Wolbachia* on reproduction (Marshall 2004, 2007). Previous research has shown geographic variation in diapause (Mousseau and Roff 1989a; Winterhalter and Mousseau 2007) and genetic alleles (Howard and Furth 1986; Marshall 2004; Huestis and Marshall 2006b), making it an ideal system for identifying genetic associations with the diapause trait. I utilized 12 populations from throughout the range of this species to study geographic and genetic variation in diapause plasticity.

3.3.2 Experimental Animals

Crickets were collected from locations throughout the southeastern United States; collection trips were made in August 2005 and July 2006. Names and collection locations of the 12 populations used in this study are detailed in Table 3.1. Once collected, crickets were screened for malate dehydrogenase (*Mdh*-1) and isocitrate dehydrogenase (*Idh*-1) to ensure species status (following methods of Huestis and Marshall 2006b), as species in the genus *Allonemobius* are morphologically cryptic (Howard and Furth 1986; Britch et al. 2001).

Populations were maintained under standard laboratory conditions (27°C, 14:10 L:D; Huestis and Marshall 2006a,b) for a minimum of one generation prior to use in diapause plasticity experiments to remove maternal and environmental effects from the field.

3.3.3 Phenotypic Assays

Crickets used for diapause assays were raised under standard rearing conditions as described above, and separated as juveniles into sex-specific cages to prevent mating. Upon adult maturation, males and females of similar age (within 7 days) were placed into mating cages and allowed to mate for one week, ensuring females had ample opportunity to acquire sperm to fertilize their eggs. Females were then placed in individual boxes at 27°C and 14:10 L:D, and provided with rolled cheesecloth placed in a 60mm Petri dish lid. Petri dishes containing cheesecloth used for oviposition were collected each day and new cheesecloth provided. Eggs were collected from females each day for 10-12 days or until females died. 18-40 females from each of the 12 populations were initially used, although not all produced usable batches of eggs (see below).

Removed cheesecloth was immediately examined for the presence of eggs, and batches containing $\geq$10 eggs were randomly split between 27 and 32°C; photoperiod was held constant in both treatments at 14:10 L:D. After incubation at their respective temperatures for a period of 18-20 days, eggs were examined under a dissecting microscope and developmental status determined. Eggs were classified as either diapause (small and white), non-diapause (presence of eyespots, at a minimum), hatched (hatchling cricket present), or rotten (less than 5% of all eggs show brown or black fungus; counts of rotten eggs were not used in statistical analyses).

3.3.4 Genomic Analyses

Field-caught individuals from each population were used for AFLP analysis. DNA was isolated from 4 individuals per population using a DNeasy tissue kit (Qiagen; Valencia, CA) following manufacturer-recommended protocols. Restriction ligation, pre-amplification, and selective amplification reactions were performed with ABI reagents (Foster City, CA) using the

methods developed by Britch et al. (2007) for generating AFLP markers for *Allonemobius*. Two primer combinations were used: MseI-CTA with EcoRI-ACA and EcoRI-ACG; EcoRI primers were fluorescently-labeled with FAM and JOE, respectively. Resulting fragments were run on an ABI 3100 sequencer, and scored using GeneMarker software (SoftGenetics, State College, PA). Markers $\geq$ 50bp in size were included in analyses, and these two primer combinations yielded a total of 442 usable markers. Principle coordinate analysis was used to reduce and group these markers into independent axes of genetic variation with NTSYS 2.2 (Exeter Software, Setauket, NY) following the procedure outlined in Kropf et al. (2002).

3.3.5 Statistical Analyses

For each batch, the proportion of diapause eggs at each temperature was calculated by dividing the number of diapause eggs by the total number of eggs counted (excluding those scored as rotten). Diapause plasticity was calculated for each batch by subtracting the proportion of eggs diapausing at 32°C from the proportion diapausing at 27°C; this measure represents the proportion of eggs which broke diapause in response to the increased temperature treatment. To test for effects of maternal age, plasticity data was also compiled for individual females which produced eggs in each of 3 intervals: 1-3, 4-6, and 7+ days of egg-laying by combining data from all eggs produced in each of the 3 intervals. For each population, mean plasticity was calculated by averaging the diapause plasticity of eggs produced by each female in that population.

A repeated-measures analysis of variance (ANOVA) was used to test for effects of population (fixed-factor), temperature (repeated measure), and their interaction on the proportion of diapause eggs produced. Similarly, a repeated-measures ANOVA was used to test for effects of population (fixed-factor), maternal age (repeated measure), and their interaction on diapause plasticity. Post-hoc ANOVAs were used to determine which age intervals were significantly different. Stepwise multiple regression modeling was used to select which principle coordinates of genetic variance were significantly associated with population-

level mean diapause and diapause plasticity. All statistical analyses were performed with SAS Learning Edition 4.1 (SAS Institute, Cary, NC). Statistics were considered significant at α<0.05.

3.4 Results

3.4.1 Phenotypic Assays

A total of 187 females from 12 populations across the southeastern United States produced eggs (*n*=15,617) used to assay geographic variation in plasticity of the diapause response to temperature and maternal age (Table 3.1). The proportion of diapausing eggs produced by each egg-incubation temperature differed significantly (Table 3.2), such that diapause occurrence was generally lower at 32°C than at 27°C (Figure 3.1); thus, some eggs responded plastically to increased egg-incubation temperature by initiating direct-development. Additionally, there was a significant interaction effect between population and egg-incubation temperature on the proportion of diapausing eggs (Table 3.2, Figure 3.1), indicating that populations vary in their ability to respond to changing egg-incubation temperature. Because variation in diapause plasticity exists between populations (Table 3.3), genomic differences between populations may therefore be used to look for associations between genotype and phenotype (see below).

Table 3.1. Designations, collection localities, and number of females utilized for study populations.

Population	Letter	Degrees N Latitude	Degrees W Longitude	Elevation (ft)	number for plasticity	number for age
TX 30/146	A	33.1692	95.2431	479	21	7
NC 70/17S	B	35.1017	77.0915	23	12	7
GA 85/149	C	34.2572	83.4645	692	11	7
Piedmont	D	34.5667	83.5415	1404	8	1
SC 95/172	E	34.2808	79.6849	66	14	11
NC 40/298	F	35.7490	78.6497	295	13	11
SC 85/1	G	34.4869	83.0229	709	12	3
NC 11S	H	35.5155	77.4141	66	18	10
SC/RA2	I	34.6009	82.6538	685	35	4
Athens, GA	J	33.9257	83.3406	715	11	10
GA 441/19	K	34.6854	83.4262	1535	16	6
SC/28	L	34.7754	83.1004	961	16	15

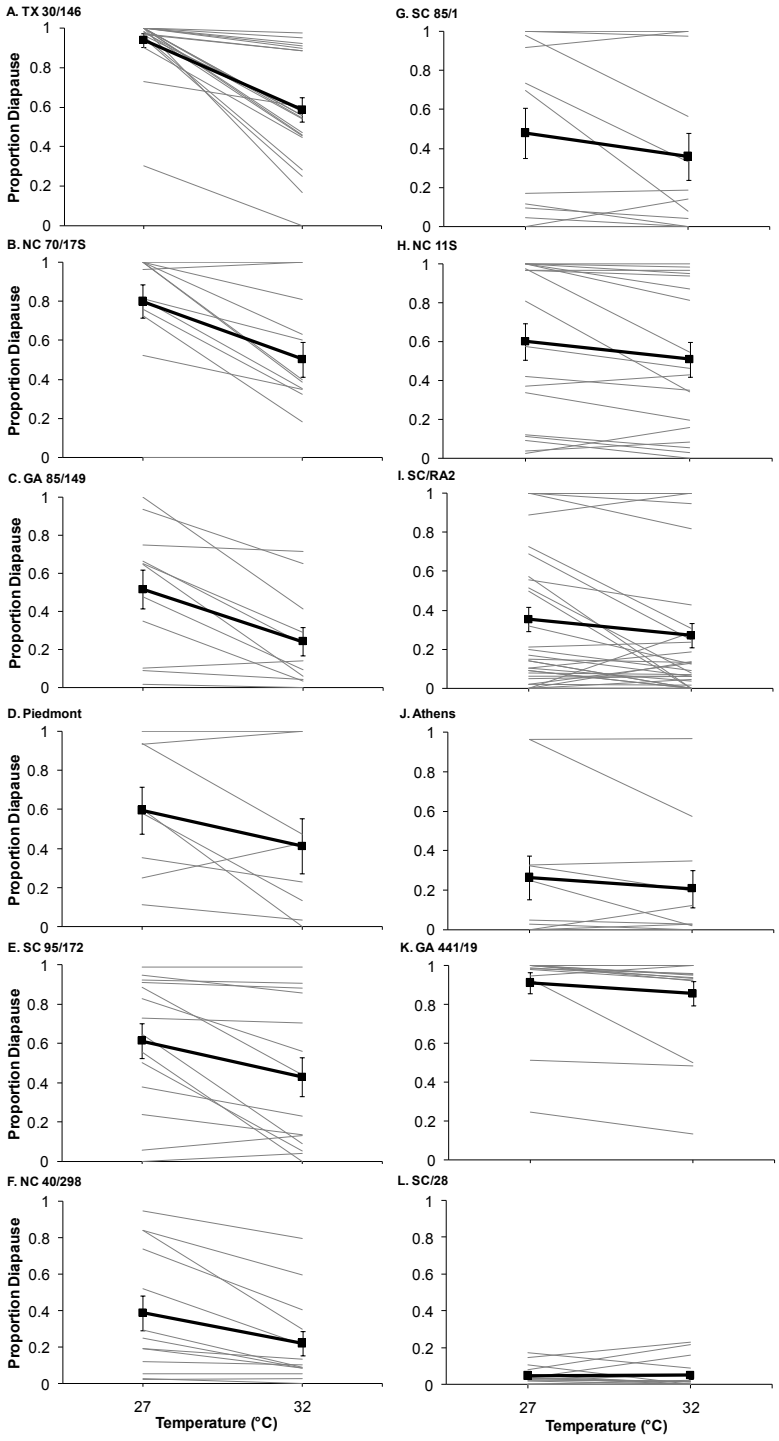

Figure 3.1. Variation in plasticity of diapause in response to egg-incubation temperature across twelve populations of the cricket *Allonemobius socius*. Females laid eggs at 27°C, and batches were split between 27 and 32°C within 24 hours post-laying. Thin grey lines represent diapause proportions produced by individual females (*n*=187), while black squares, lines, and error bars represent sample mean and standard error of each population.

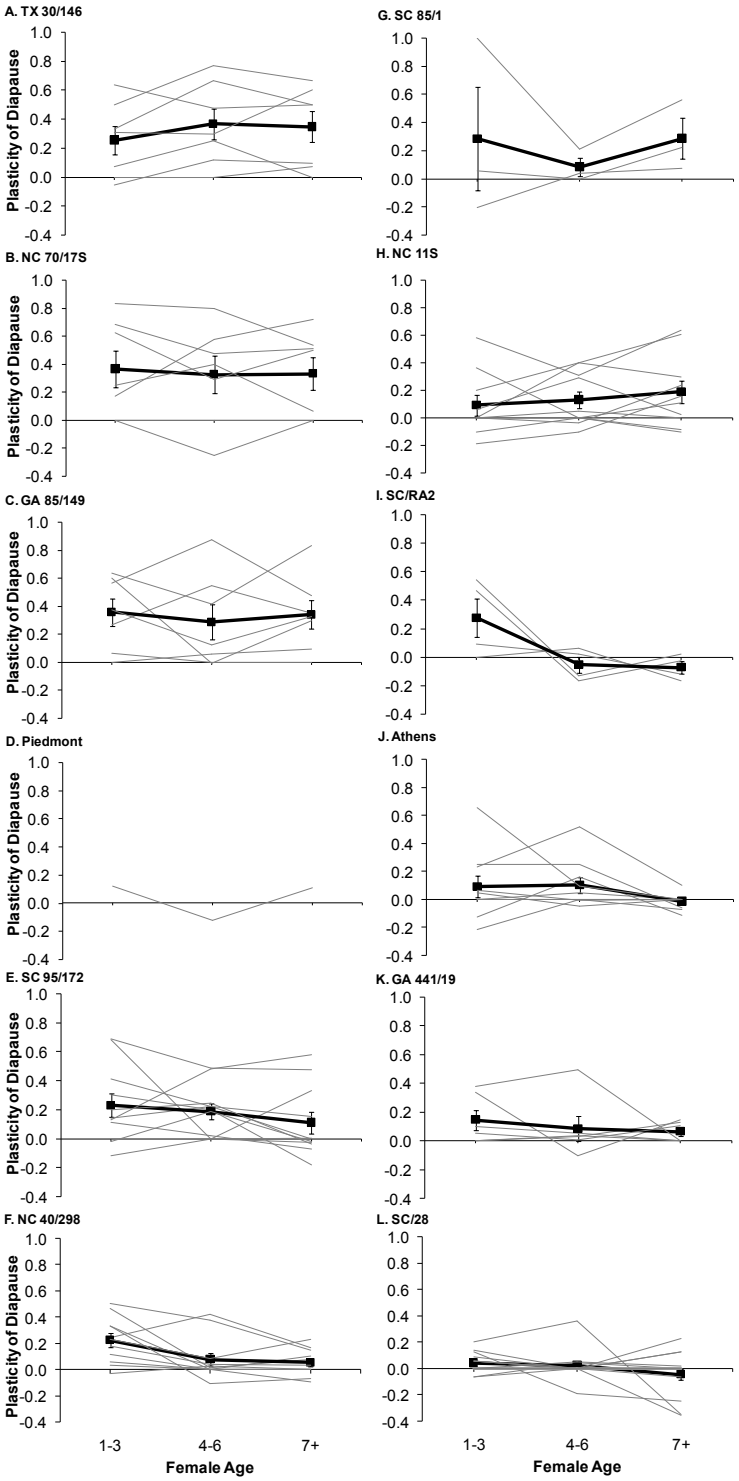

Figure 3.2. Variation in maternal age effects on diapause plasticity across twelve populations of *A. socius*. For females which produced eggs across all 3 age intervals (1-3, 4-6, and 7+ days; *n*=92), average diapause plasticity (measured as the difference in diapause between 27 and 32°C) for each interval is given by thin grey lines. Black squares, lines, and error bars represent sample mean and standard error of each population.

There was also a significant effect of female age on diapause plasticity (Table 3.3, Figure 3.2), but this effect did not vary among populations, as the interaction of female age and population was not significant at α<0.05 (Table 3.3). Specifically, plasticity was higher for eggs laid by younger females (days 1-3) relative to older females (days 4-6 and 7+), while the two older intervals were not significantly different from one another (Figure 3.3).

Table 3.2. Repeated-measures ANOVA on diapause occurrence, with population as fixed factor and temperature as repeated measure.

Between Subjects				
Source	df	MS	*F*	*P*
Population	11	1.79412	9.58	<0.0001
Error	175	0.18725		
Within Subjects				
Source	df	MS	*F*	*P*
Temperature	1	1.96249	105.56	<0.0001
Population x Temperature	11	0.09803	5.27	<0.0001
Error (Temperature)	175	0.01859		

Within-subject *P*-values are with Huynh-Feldt corrections.

Table 3.3. Repeated-measures ANOVA on diapause plasticity, with population as fixed factor and age as repeated measure.

Between Subjects				
Source	df	MS	*F*	*P*
Population	10	0.35195	3.62	0.0005
Error	80	0.09724		
Within Subjects				
Source	df	MS	*F*	*P*
Age	2	0.11810	4.80	0.0095
Population x Age	20	0.03642	1.48	0.0950
Error (Age)	160	0.02461		

Within-subject *P*-values are with Huynh-Feldt corrections.

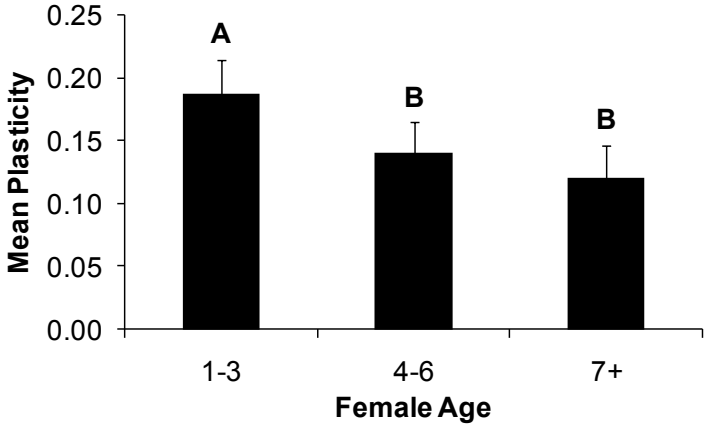

Figure 3.3. Effects of maternal age on diapause plasticity in *A. socius*. Bars represent mean plasticity (measured as the difference in diapause proportion between 27 and 32°C) and standard error, combining individuals from all populations (*n*=92). Letters denote significant differences based on post-hoc comparisons, and show plasticity was higher earlier for eggs laid earlier in the females' lifespan.

3.4.2 Genomic Analyses

A total of 442 AFLP markers were generated using two primer combinations on 48 individuals. Principle coordinate analysis resulted in 20 significant principle coordinates (PCs), which together accounted for ~90% of the genetic variance among populations (Table 3.4). These 20 PCs represent independent axes of genetic variance for these 12 populations and may therefore be used as independent components of genetic variance to look for associations with phenotypic variance in diapause.

Stepwise multiple regression modeling was used to identify which, if any, genetic PCs significantly predict phenotypic variance in both mean diapause and mean plasticity. For mean diapause, 5 PCs were selected (PCs 5, 10, 13, 14, and 16; Table 3.5), which together explained 94% of the variance among populations for mean diapause occurrence. The resulting equation was: population mean diapause = 0.4651 + 0.0217(PC5) – 0.0126(PC10) – 0.0398(PC13) + 0.0628(PC14) + 0.0227(PC16). The correlation between predicted and actual values of population mean diapause was highly significant (r = 0.982, P < 0.0001). As for diapause plasticity, the stepwise procedure included 3 PCs (PCs 6, 12, and 18; Table 3.6) able to explain 83% of the variance in this trait. The equation resulting from this procedure was: population

mean plasticity = 0.1547 – 0.0086(PC6) – 0.0102(PC12) – 0.0271(PC18). The correlation between observed and predicted values for mean plasticity was also highly significant (r = 0.936, P < 0.0001). Importantly, there is no overlap in which PCs predicted each trait, indicating that these two traits are controlled by independent components of genetic variance.

Table 3.4. Eigenvalues for the 20 principal coordinates identified.

PC	Eigenvalue	Percent	Cumulative
1	12982.22	12.36	12.36
2	9501.39	9.04	21.40
3	8043.68	7.66	29.05
4	7355.73	7.00	36.06
5	6232.24	5.93	41.99
6	6081.72	5.79	47.78
7	5067.91	4.82	52.60
8	4674.41	4.45	57.05
9	4249.23	4.04	61.09
10	4127.14	3.93	65.02
11	3636.82	3.46	68.48
12	3575.41	3.40	71.88
13	3132.69	2.98	74.87
14	2945.88	2.80	77.67
15	2673.21	2.54	80.21
16	2591.04	2.47	82.68
17	2356.16	2.24	84.92
18	2090.38	1.99	86.91
19	2038.49	1.94	88.85
20	1909.54	1.82	90.67

Table 3.5. Analysis of variance (ANOVA) and parameter estimates of stepwise regression model for principal coordinates of genetic variance on mean diapause.

Analysis of Variance					
Source	df	SS	MS	*F*	*P*
Model	5	0.5745	0.1149	33.93	0.0003
Error	6	0.0203	0.0034		
Corrected total	11	0.5948			
Parameter Estimates					
Variable	Estimate	SE	SS	*F*	*P*
Intercept	0.4652	0.0168	2.5964	766.77	< 0.0001
PC5	0.0217	0.0036	0.1238	36.56	0.0009
PC10	-0.0126	0.0043	0.0288	8.51	0.0267
PC13	-0.0398	0.0073	0.1013	29.92	0.0016
PC14	0.0628	0.0065	0.3177	93.82	< 0.0001
PC16	0.0227	0.0056	0.0550	16.23	0.0069

Table 3.6 Analysis of variance (ANOVA) and parameter estimates of stepwise regression model for principal coordinates of genetic variance on mean plasticity.

Analysis of Variance					
Source	df	SS	MS	*F*	*P*
Model	3	0.1140	0.0380	18.88	0.0005
Error	8	0.0161	0.0020		
Corrected total	11	0.1301			
Parameter Estimates					
Variable	Estimate	SE	SS	*F*	*P*
Intercept	0.1547	0.0130	0.2873	142.80	< 0.0001
PC6	-0.0086	0.0020	0.0390	19.36	0.0023
PC12	-0.0102	0.0031	0.0224	11.11	0.0103
PC18	-0.0271	0.0046	0.0708	35.21	0.0003

3.5 Discussion

Given climatic variation across a species' range, populations may develop distinct life-history strategies with only minimal genetic differentiation and without a speciation event (Hoffmann et al. 2005; Liefting and Ellers 2008). Only detailed surveys of naturally-occurring

populations throughout the range of a given species can reveal how common this pattern is, yet these surveys are often time-consuming. As such, insect species with short generation times provide a model system for studying this phenomenon. The ability of individuals to undergo diapause is a key life-history adaptation for insect species occurring in temperate or seasonal climates (Danks 1987; Mousseau and Dingle 1991; Bradford and Roff 1993; Denlinger 2002; Winterhalter and Mousseau 2007). Thus, the occurrence of diapause has been widely studied in many species, making it an ideal trait for the study of genetic and geographic variation within an insect species (Danks 1987; Denlinger 2002). Here, I surveyed variation in both mean diapause occurrence and diapause plasticity across 12 populations of *Allonemobius socius* from throughout the southeastern United States, and found significant population-level variation in both traits (Table 3.2; Figures 3.1 and 3.2). These results indicate that *A. socius* populations have evolved different life-history phenotypes across their naturally-occurring geographic range in a relatively short amount of time, given that all species in the *A. socius* complex are estimated to have diverged from a common ancestor within the last 3,000-30,000 years (Marshall 2004).

Additionally, I found a significant effect of female age on the plasticity of eggs produced, but this trait did not vary significantly among populations (Table 3.3). For the majority of females, regardless of population, plasticity was higher in batches of eggs laid in the first 3 days of the egg-laying trials than in later intervals (Figure 3.3). Thus, age effects may not be under different selective pressures in the populations we studied. Given that the length of season remaining typically decreases over a female's reproductive lifespan, this result is not surprising, as eggs laid later in the season will have a shorter time to complete development and therefore diapause may be a better strategy than direct-development (Danks 1987; Mousseau 1991; Mousseau and Dingle 1991). This finding could also result from low life-expectancy in the field (and therefore relaxed selection on life-history traits later in life), though this would not likely produce the same directional pattern in all populations. Geographic variation in maternal age effects has been previously reported in *A. socius*, but within bivoltine populations, clutch effects were similar to those I observed, with lower diapause incidence earlier in a female's

reproductive lifespan (Mousseau 1991). As my populations were sampled from primarily bivoltine locations, our results are in accordance with those previously reported for this species.

Many adaptive life-history traits are quantitative in nature (Stearns 1992; Schlichting and Pigliucci 1998; Roff 2002), and are therefore affected by many loci in the genome. I used AFLP markers to identify independent axes of genetic variation among the 12 populations surveyed, and were able to identify 20 significant principle coordinates of genetic variance. These were then used in multiple regression models to test for associations between coordinates and the population average values of mean diapause occurrence and diapause plasticity. These results found associations between both of these life-history traits with 5 and 3 independent axes of genetic variance, respectively (see Results above). Importantly, the PCs identified were different for each life-history trait, indicating genetic independence of these two critical traits. These results indicate that there is no significant genetic association between these traits, and that the mean and plasticity of the diapause reaction norm may evolve independently in this species. The independence of these components which define the reaction norm for diapause in this species would be critical for the rapid formation of life-history variation between these closely-related populations, as appears to have likely occurred given the above results and recent divergence of these populations (within the last 3,000-30,000 years; Marshall 2004). These results are significant, in that they have implications for the formation of new reaction norms and how locally adapted traits might arise and be maintained. Specifically, while genetic independence indicates that populations may be able to respond quickly to changes in environmental conditions, the spread of a particular advantageous combination will be slow. This is because a particular combination would be able to spread faster and be maintained easier with tight linkage between mean and plasticity of the diapause reaction norm.

CHAPTER 4

GEOGRAPHIC DISTRIBUTIONS OF *IDH*-1 ALLELES IN A CRICKET ARE LINKED TO DIFFERENTIAL ENZYME KINETIC PERFORMANCE ACROSS THERMAL ENVIRONMENTS

4.1 Abstract

Geographic clines within species are often interpreted as evidence of adaptation to varying environmental conditions. However, clines can also result from genetic drift, and these competing hypotheses must therefore be tested empirically. The striped ground cricket, *Allonemobius socius*, is widely-distributed in the eastern United States, and clines have been documented in both life-history traits and genetic alleles. One clinally-distributed locus, isocitrate dehydrogenase (*Idh*-1), has been shown previously to exhibit significant correlations between allele frequencies and environmental conditions (temperature and rainfall). Further, an empirical study revealed a significant genotype-by-environmental interaction (GxE) between *Idh*-1 genotype and temperature which affected fitness. Here, I used enzyme kinetics to further explore GxE between *Idh*-1 genotype and temperature.

I found significant GxE between temperature and three measures of kinetic performance, providing further evidence that the natural cline in *Idh*-1 allele frequencies in *A. socius* is the result of adaptation to environmental conditions. Differences in kinetic performance across temperatures also mirror many of the geographic patterns observed in allele frequencies. This study further supports the hypothesis that the natural distribution of *Idh*-1 alleles in *A. socius* is driven by natural selection on differential enzymatic performance. This example is one of several which clearly document a functional basis for observed clines in allele frequencies, and provides further evidence for the non-neutrality of some allozyme alleles.

4.2 Introduction

Individuals within populations are under selection pressure to adapt to their environment; these adaptations can be morphological, physiological, or behavioral in nature. However, these diverse adaptations all have a molecular basis; as such, the study of molecular adaptation to environmental conditions is an active area of research within evolutionary biology. One biochemical adaptation that lends itself to empirical study is the kinetic performance of different enzyme alleles under a range of environmental conditions, such as temperature.

Different enzyme alleles at a protein-coding locus with the same metabolic function are commonly referred to as allozymes. A point mutation in the protein-coding sequence of an enzyme may lead to an amino acid substitution, altering the charge, weight, and folding of the protein, changing the relative electrophoretic mobility of the enzyme and thus permitting the detection of allozyme alleles (Hubby and Lewontin 1966). Amino acid substitutions may also affect the function of the protein, altering optimal ranges for temperature, pH, or substrate concentration. Most amino acid changes will likely be deleterious and quickly eliminated by purifying selection (Harris 1966; Lewontin and Hubby 1966). However, some substitutions may not significantly affect the function of the protein, and are therefore selectively neutral, while others may improve enzyme function and be favored by selection.

Originally, evolutionary biologists and geneticists thought genetic diversity in populations would be quite low due to purifying selection (Lewontin and Hubby 1966). However, early studies of protein polymorphism revealed unexpected polymorphisms at most loci studied, in a range of organisms including mice, humans, and fruit flies (Harris 1966; Hubby and Lewontin 1966; Lewontin and Hubby 1966). As a result of these high levels of diversity revealed, allozymes were thought to be neutral (Kreitman and Akashi 1995; Eanes 1999). However, others believed that allozymes would be subject to selection, and subsequent studies have supported selection in many cases (e.g., Watt 1994; Kreitman and Akashi 1995). Thus, a debate over whether allozymes were neutral or were subject to selection began soon after their

discovery, and it has been stated that “few subjects in biology have been more strongly debated than the evolutionary significance of protein polymorphisms” (Powers et al. 1991).

As enzyme function depends on temperature, pH, substrate concentration, and other environmental factors, some amino acid substitutions will result in an enzyme that functions best under certain conditions, and these may be favored locally by natural selection, depending on environmental factors. In populations inhabiting heterogeneous environments, multiple alleles at one enzyme locus may be maintained by balancing selection (Hotz and Semlitsch 2000; Schmidt et al. 2000; Wheat et al. 2006). Evidence for selection acting upon allozyme loci includes the presence of clines in allozyme allele frequencies (Powers et al. 1991), correlations between environmental variables and allozyme allele frequencies (Riddoch 1993), differences in chemical properties between allozyme alleles (Fields 2001), and differential performance and/or fitness differences between individuals with different allozyme genotypes (e.g., Watt 1992, 1994; Eanes 1999; Hotz and Semlitsch 2000; Huestis and Marshall 2006b). These types of evidence are often combined within a system, and several lines of evidence together provide support for the hypothesis that selection is acting on certain allozyme loci.

Another important evolutionary question which is often overlooked is ‘why are common alleles common?’. Such common alleles may be prevalent across a wide environmental landscape for many reasons, including ancestral inertia, recent range expansion, genetic drift, purifying selection, or some combination of these or other processes. While it is not always possible to determine the underlying processes that drive or maintain the existence of common allozyme alleles, experiments testing for differential enzyme performance of alleles across a wide-range of environmental conditions can shed light on the possibility of commonness being maintained by natural selection.

Numerous studies providing strong evidence for both neutrality and selection of allozyme loci are found in the literature, and only a few will be detailed here. Cases of neutrality include widespread surveys of allelic variation in white spruce (Jaramillo-Correa et al. 2001),

Peromyscus mice (Storz and Nachman 2003), and several others (Hoffmann et al. 1995; Kreitman and Akashi 1995). In contrast, strong evidence for selection has been found for two well-studied loci, phosphoglucose isomerase (*Pgi*; Watt 1992; Riddoch 1993) and alcohol dehydrogenase (*Adh*; McDonald and Kreitman 1991; Kreitman and Akashi 1995) in a wide range of organisms, and for other loci on a smaller scale (e.g., Powers et al. 1991; Hotz and Semlitsch 2000). Another allozyme locus, isocitrate dehydrogenase (*Idh*), has been studied less than the well-known examples above, but evidence of natural selection acting on this locus has been found across a range of taxa, including bacteria (Dean and Golding 1997; Zhu et al. 2005), plants (Hawkins et al. 1991; Bergmann and Gregorius 1993), invertebrates (da Cunha and de Oliveira 1996; Sokolova and Portner 2001), and vertebrates (Pemberton et al. 1991; Vrijenhoek et al. 1992). Isocitrate dehydrogenase is a metabolic enzyme in the Krebs cycle, and its activity is therefore one of several key steps in the generation of ATP from glucose (Eanes 1999).

In the striped ground cricket, *Allonemobius socius*, there is a naturally-occurring cline in *Idh*-1 allele frequencies, hypothesized to have resulted from natural selection (Huestis and Marshall 2006b). In *A. socius*, there is significant geographic variation in the allele frequency distributions of two *Idh*-1 alleles (1.8 and 2.2), while a third allele (2.0) is found at a frequency of approximately 50% in most locations (Figure 4.1). This geographic variation in the frequencies of the 1.8 and 2.2 alleles is coupled with significant correlations between these allele frequencies and two important environmental variables, temperature and rainfall (Huestis and Marshall 2006b). An empirical study of homozygous individuals revealed a significant genotype-by-environment interaction between *Idh*-1 genotype and temperature on fitness in the direction hypothesized based on the environmental correlations (Huestis and Marshall 2006b), providing experimental evidence supporting temperature-driven selection on the *Idh*-1 locus in this species. I wanted to address the hypotheses that 1) the 2.0 allele is common across all thermal environments because it performs well under a wide range of conditions and 2) the clinal

distributions of the 1.8 and 2.2 alleles were due to differences in performance across temperatures.

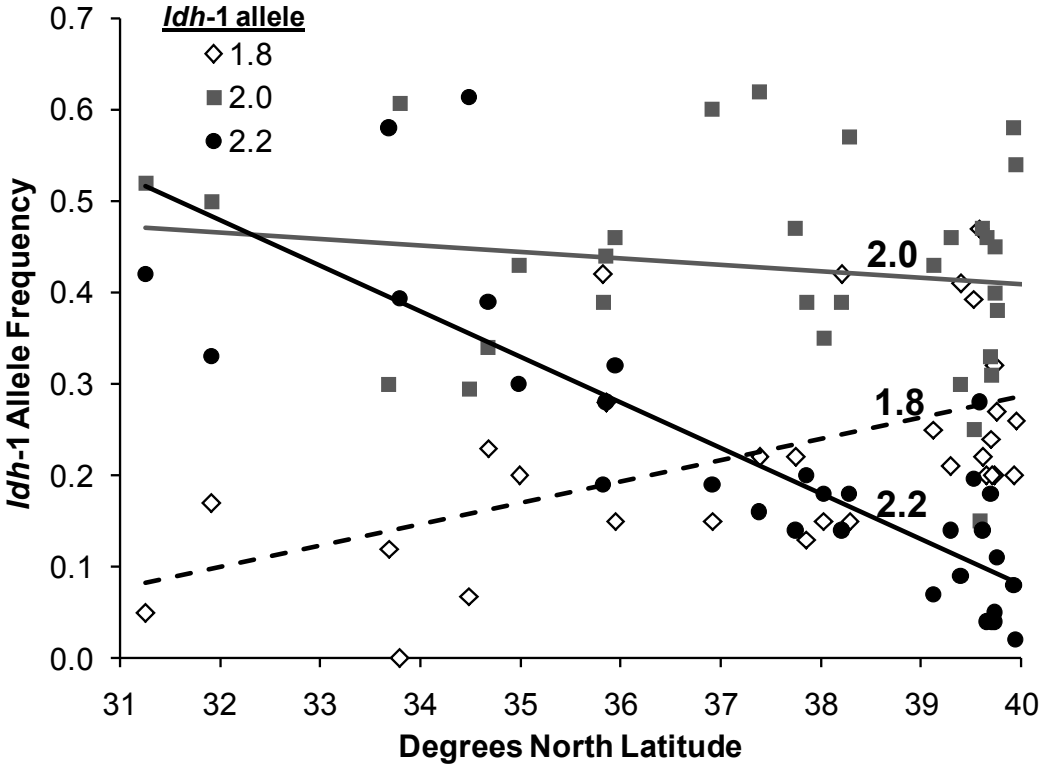

Figure 4.1. Geographic variation in allele frequencies at the *Idh*-1 locus in the cricket *Allonemobius socius*. Points are allele frequency data from field-collected populations (data from [14]), and lines are least-squares regressions. The 1.8 allele is symbolized by open diamonds and a dashed line, the 2.0 allele by grey squares and a grey line, and the 2.2 allele by black circles and a black line.

To further explore the hypothesis that selection has shaped allelic distributions of the *Idh*-1 locus in *A. socius*, I performed enzyme kinetics assays at a range of ecologically-relevant temperatures to explore the molecular basis of the GxE interaction between *Idh*-1 genotype and temperature. For all 3 kinetic parameters examined (K_m, V_{max}, and enzyme efficiency), there was a significant GxE between *Idh*-1 genotype and temperature which affected enzyme performance. Additionally, there were significant differences in performance parameters between alleles at both high and low temperatures. Together, these results provide additional

evidence that natural selection underlies the naturally-occurring geographic distribution of *Idh*-1 alleles in *A. socius*.

4.3 Methods

4.3.1 Study System

Crickets of the genus *Allonemobius* range from southern Canada to the southern United States, primarily in the East, and are abundant in appropriate habitat throughout their range. Due to their large distribution, abundance in the field, and ease of laboratory maintenance, members of the *A. socius* complex have been used as a model system for several aspects of evolutionary biology, including studies of speciation and hybrid zone dynamics (Howard and Furth 1986; Howard and Waring 1991; Britch et al. 2001), *Wolbachia* (Marshall 2004, 2007), life-history evolution (Mousseau and Roff 1989a; Mousseau 1991; Olvido et al. 1998; Huestis and Marshall 2006a), and morphological variation (Mousseau and Roff 1989a,b, 1995; Fedorka and Mousseau 2002b; Olvido et al. 2003). Additionally, given their naturally widespread distribution, the *A. socius* complex is an ideal model system for studying geographic variation in life-history traits and genetic diversity. Within the *A. socius* complex, clines have been found in ovipositor length (Mousseau and Roff 1995), diapause occurrence (Mousseau and Roff 1989a; Mousseau 1991; Bradford and Roff 1995), allozyme alleles (Howard and Furth 1986; Britch et al. 2001; Huestis and Marshall 2006b), and nuclear and mitochondrial markers (Marshall 2004).

Members of the *A. socius* complex are morphologically cryptic, and species were originally discovered and described using allozymes (Howard 1982, 1983; Howard and Furth 1986); therefore, much data about the geographic distribution of allozyme alleles are readily available. One locus in particular, isocitrate dehydrogenase (*Idh*-1), is strongly clinal within *A. socius*, with the 1.8 allele being common in the north and east and the 2.2 allele being common to the south and west (Figure 4.1; Huestis and Marshall 2006b).

The geographic distributions of these alleles are also correlated with two important environmental variables, temperature and rainfall, with 1.8 at highest frequency in cooler, drier

locations while 2.2 is associated with hotter, wetter locations (Huestis and Marshall 2006b). A third allele, 2.0, is found at intermediate frequencies in most locations (~50%; Figure 4.1) and not significantly distributed in relation to geography or climate. In a previous study, I found a significant interaction between *Idh*-1 genotype and temperature on fitness, such that individuals homozygous for the 1.8 allele laid more eggs at a cool temperature relative to the two faster alleles, and individuals homozygous for the two faster alleles laid more eggs at a warm temperature than 1.8 individuals (Huestis and Marshall 2006b). However, the molecular basis for this GxE was not examined. Here I use enzyme kinetics to assay differences in performance between these 3 alleles across a range of ecologically-relevant temperatures.

4.3.2 Kinetic Parameters and Calculations

The initial velocity of an enzyme-catalyzed reaction depends on the initial substrate concentration; this relationship is typically hyperbolic, with a linear increase at lower concentrations until the reaction approaches saturation, at which point further increases in substrate will not increase reaction velocity (see Figure 4.2A). Two important parameters are typically calculated using kinetic assay data: V_{max}, the initial reaction velocity at saturated substrate concentration, and the Michaelis constant, K_m, which is a measure of the affinity of the enzyme for the substrate and the rate at which the substrate is converted to product (Cornish-Bowden 2004). These parameters are obtained through a double-reciprocal plot of velocity against substrate concentration for the linear portion of the original curve (see Figure 4.2B). First, least-squares regression is performed on the double-reciprocal data and the slope and y-intercept calculated. V_{max} is then calculated from the regression equation, using the formula: V_{max}=1/y-intercept (Cornish-Bowden 2004). Next, K_m is obtained from the equation: $K_m=V_{max}$*slope (Cornish-Bowden 2004).

Because the y-intercept changes with enzyme concentration (Figure 2B), K_m and V_{max} are also influenced by the initial enzyme concentration, such that the values of both parameters increase when enzyme concentration increases (Table 4.1). However, the slope of the least-

squares line is not dependent on the initial enzyme concentration (Figure 4.2B, Table 4.1), and is equal to K_m/V_{max}, a measure of enzyme efficiency. As V_{max} is a measure of reaction velocity, a relatively higher value of V_{max} means the reaction can progress faster, and therefore higher values of V_{max} indicate better enzymatic performance. Conversely, K_m contains a measure of enzyme-substrate affinity, and is the amount of substrate needed to achieve one-half V_{max}. Therefore, an enzyme with a lower K_m needs less substrate to achieve a given rate than one with a higher K_m. Lastly, because enzyme efficiency (as defined above) is calculated from reciprocal plots, a smaller value of this parameter (either due to a lower K_m and/or a higher V_{max}) means the enzyme is more efficient than a higher value. Here I report all 3 kinetic parameters (efficiency, K_m, and V_{max}), but note that K_m and V_{max} may be influenced by variation in amount of enzyme present in each individual. Specifically, differences in these two parameters when comparing two individuals could be due to differences in the relative amounts of enzyme in individuals (see Table 4.1) or real differences in performance between individuals for these two parameters.

Table 4.1. Estimation of 3 enzyme kinetic parameters from data presented in Figure 4.2.

parameter	1 cricket	2 crickets	ratio (2/1)
slope	0.0081	0.0080	0.9877
V_{max}	5.6180	20.0803	3.5743
K_m	0.0455	0.1606	3.5302

This table demonstrates that slopes are nearly equal with approximately double the amount of enzyme, while both K_m and V_{max} increase by a factor of approximately 3.5.

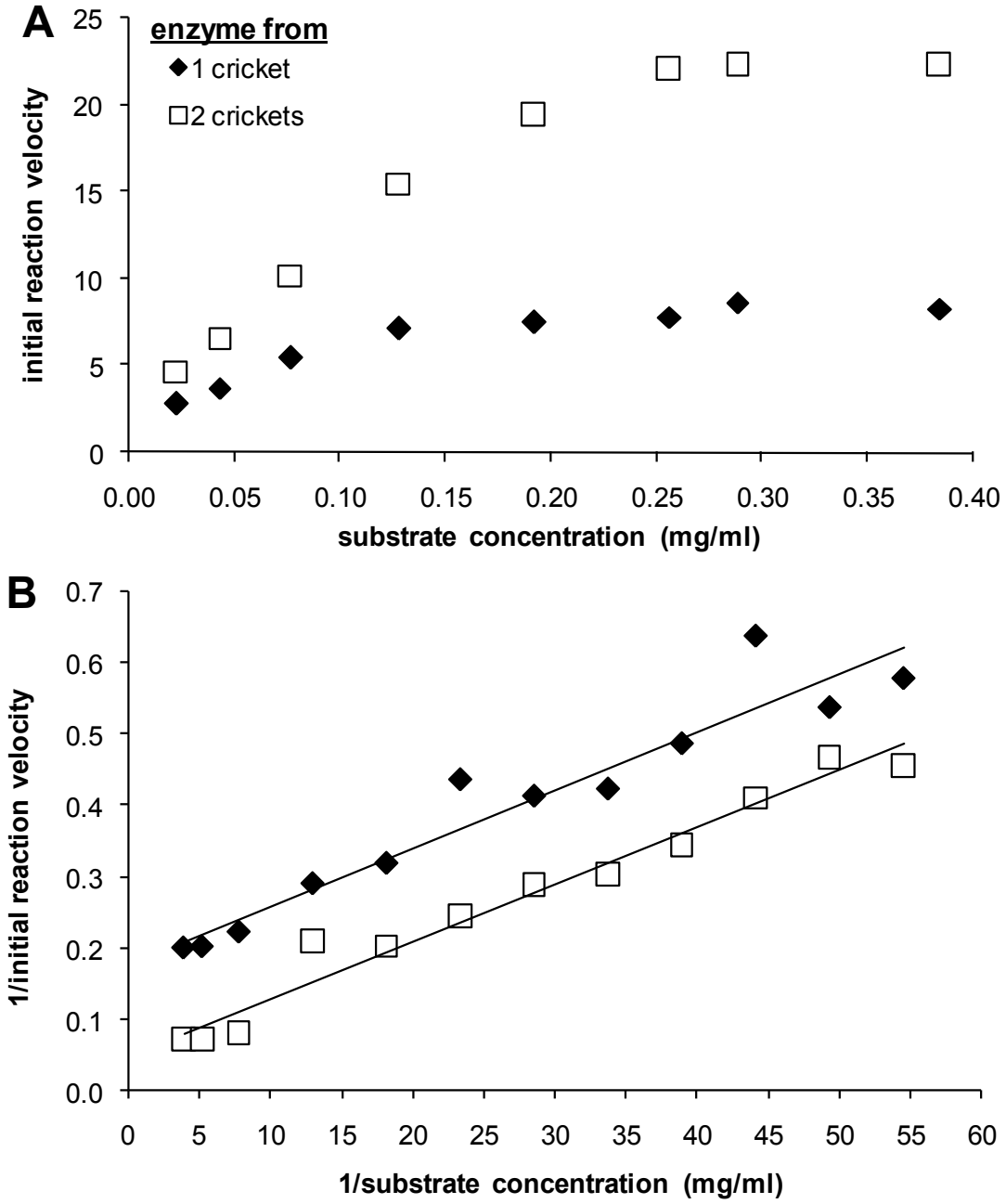

Figure 4.2. Example enzyme kinetics data. Experimental *Idh*-1 kinetic data using homogenates from 1 cricket (2 legs; solid diamonds) and 2 crickets (4 legs; open squares). (A) Initial reaction velocity as a function of substrate concentration, showing the expected hyperbolic relationship. Reaction velocity increases linearly until approaching the saturation point. (B) Lineweaver-Burk plot, the double-reciprocal method for calculating K_m, the Michaelis constant, and V_{max}, the maximum reaction velocity, from kinetic assay data.

4.3.3 Experimental Methods

4.3.3.1 Experimental Animals

To further detail the relationship between temperature and enzyme activity in *A. socius*, I conducted an enzyme kinetics experiment using tissue homogenates derived from laboratory-raised *Idh*-1 homozygotes. Juvenile crickets were collected in July 2006 from a field population in western North Carolina (35.199°N, 81.371°W), an area known to have all 3 alleles at roughly equal frequencies (see Figure 1 in Huestis and Marshall 2006b). Field-collected juveniles were raised to adulthood at 27°C in sex-specific cages to prevent mating. Adults were individually genotyped for malate dehydrogenase (*Mdh*-1; diagnostic of the *A. socius* complex relative to other species in the genus) and isocitrate dehydrogenase to screen for individuals homozygous at the *Idh*-1 locus. Allozyme electrophoresis and staining was performed using standard methods for *Allonemobius* (Howard and Furth 1986).

Individuals homozygous for each of the 3 alleles (1.8, 2.0, and 2.2) were placed in genotype-specific mating cages and allowed to mate for 2 weeks, producing homozygous offspring. Offspring were raised to adulthood at 27°C and ~10 individuals per genotype were randomly chosen and frozen at -80°C. For these individuals, *Idh*-1 genotype was confirmed with allozyme electrophoresis on head tissue as above. Enzyme homogenates from 3 individuals of each genotype were generated by homogenizing both rear legs in 25 µl 0.2M tris-citrate buffer, pH 8.0 (Howard and Furth 1986), centrifuging for 2 minutes at 10,000x*g*, and removing the supernatant for use in enzyme kinetics assays.

4.3.3.2 Kinetics Assays

Enzyme kinetics assays were performed in 96-well microplates using a temperature-controlled microplate reader (Bio-Tek Instruments, Winooski, VT), similar to the procedure described by Oppert et al. (1997). A pilot study was conducted using a wide range of substrate concentrations and standard *Idh* staining media optimized for *Allonemobius* (Howard and Furth 1986; 0.2M tris-citrate buffer, pH 8.0 with 1mg/ml $MgCl_2$, 0.2mg/ml NADP, 0.2mg/ml NBT, and

0.04mg/ml PMS; all reagents from Sigma-Aldrich) to determine the range of concentrations which produced a linear relationship between initial reaction velocity and substrate concentration (see Figure 4.2A; Cornish-Bowden 2004). This staining solution produces a purple color as the reaction progresses, and absorbance was read at 595nm using the microplate reader. Although this assay measures combined activity for *Idh*-1 and *Idh*-2, the *Idh*-2 locus is monomorphic in *A. socius* (and all species of *Allonemobius*; Howard 1983; Howard and Waring 1991; Britch et al. 2001; Huestis and Marshall 2006b); therefore, all differences observed between *Idh*-1 genotypes should result from variation in performance of *Idh*-1 alleles. From these preliminary data, four concentrations of isocitric acid were used for further assays (0.077, 0.055, 0.043, and 0.035 mg/ml).

Next, assays were performed on 3 individuals per genotype (i.e., 1.8, 2.0, and 2.2 homozygotes) at 7 ecologically-relevant temperatures (18, 21, 24, 27, 30, 33, and 36°C). For each individual, at each temperature and substrate concentration, a 1 µl aliquot of enzyme homogenate was added to 25 µl of staining media and used in an assay following the protocol outlined above. All samples for each temperature were conducted simultaneously with absorbance readings taken for every 15 sec for 20 minutes (yielding 80 absorbance readings per aliquot).

Using standard Michaelis-Menten kinetics (e.g., Levitzki and Koshland 1969; Cornish-Bowden 2004) implemented by the program KC3 (Bio-Tek Instruments), initial reaction velocity of the enzymatic conversion of isocitrate to α-ketoglutarate was calculated using the linear portion of the curve (approximately 5-15 minutes). For each individual at each assay temperature, the increase in velocity with increasing isocitric acid concentration was plotted and slopes calculated (see *Kinetic Parameters and Calculations* above; Figure 4.2B), yielding a measure of enzyme efficiency – i.e., the inverse of the change in the rate of the reaction with increasing substrate concentration. K_m, and V_{max} were also calculated from the assay data using standard methods as described above.

4.3.4 Statistical Analyses

Kinetic data were analyzed using repeated-measures analysis of variance (ANOVA) with *Idh*-1 genotype as a between-subjects factor and assay temperature and the interaction of temperature and *Idh*-1 genotype as within-subjects factors. Enzymatic efficiency, K_m, and V_{max} were analyzed using separate ANOVA's. Post-hoc ANOVA's were then used to test for significant differences between alleles at each temperature and groupings assigned using Ryan-Einot-Gabriel-Welsch (REGWQ) multiple-range tests (SAS Institute 2006). All statistical analyses were performed with SAS Learning Edition 4.1 (SAS Institute 2006). Statistics were considered significant at $P<0.05$.

4.4 Results

There was a significant interaction between *Idh*-1 genotype and assay temperature that affected enzyme efficiency, measured as the increase in velocity with increased substrate concentration (Table 4.2). Because of the inverse relationships plotted by this method, a lower value indicates greater efficiency. Enzyme efficiency was significantly different between alleles at the two lowest temperatures tested (18 and 21°C), while there was no difference at the higher temperatures tested (24-36°C; Figure 4.3A). At 18°C, 2.0 individuals outperformed 1.8 and 2.2 individuals, which were not significantly different from each other, while at 21°C, 2.0 and 2.2 individuals were not significantly different and more efficient than 1.8 individuals (Figure 4.3A).

Table 4.2. Repeated-measures ANOVA on kinetic efficiency across a temperature gradient.

Between Subjects				
Source	df	MS	*F*	*P*
Genotype	2	0.00303	5.07	0.0627
Error	5	0.00060		
Within Subjects				
Source	df	MS	*F*	*P*
Temperature	6	0.00996	39.89	<0.0001
Genotype x Temperature	12	0.00108	4.31	0.0009
Error (Temperature)	30	0.00025		

Within-subject *P*-values are with Huynh-Feldt corrections.

There was also a significant GxE between *Idh*-1 genotype and assay temperature for K_m, the Michaelis constant (Table 4.3). K_m is a measure of the binding affinity between the enzyme and the substrate, and a lower value of K_m indicates a higher affinity. There were significant differences in K_m between alleles at 3 temperatures (the two lowest temperatures, 18 and 21°C, and the highest temperature, 36°C; Figure 4.3B). Similar to the results for efficiency (see above), at 18°C, 2.0 individuals had lower K_m than 1.8 and 2.2 individuals, which were not different from each other, while at 21°C, 2.0 and 2.2 individuals were not significantly different from each other and had lower K_m than 1.8 individuals (Figure 4.3B). At 36°C, individuals of all three alleles were significantly different from each other, with 2.0 having the lowest K_m, followed 1.8 and 2.2 (Figure 4.3B).

Table 4.3. Repeated-measures ANOVA on K_m across a temperature gradient.

Between Subjects

Source	df	MS	*F*	*P*
Genotype	2	0.07596	2.94	0.1431
Error	5	0.02582		

Within Subjects

Source	df	MS	*F*	*P*
Temperature	6	0.17803	7.79	0.0010
Genotype x Temperature	12	0.06115	2.67	0.0426
Error (temperature)	30	0.02286		

Within-subject *P*-values are with Huynh-Feldt corrections.

Lastly, the maximum reaction velocity, V_{max}, was affected by a significant GxE between *Idh*-1 genotype and assay temperature (Table 4.4). At 36°C, 2.2 had a significantly higher V_{max} than 1.8 and 2.0, which were not significantly different (Figure 4.3C). These data indicate that the 2.2 allele encodes an enzyme with the highest velocity but also the highest K_m at 36°C, an expected tradeoff between velocity and affinity.

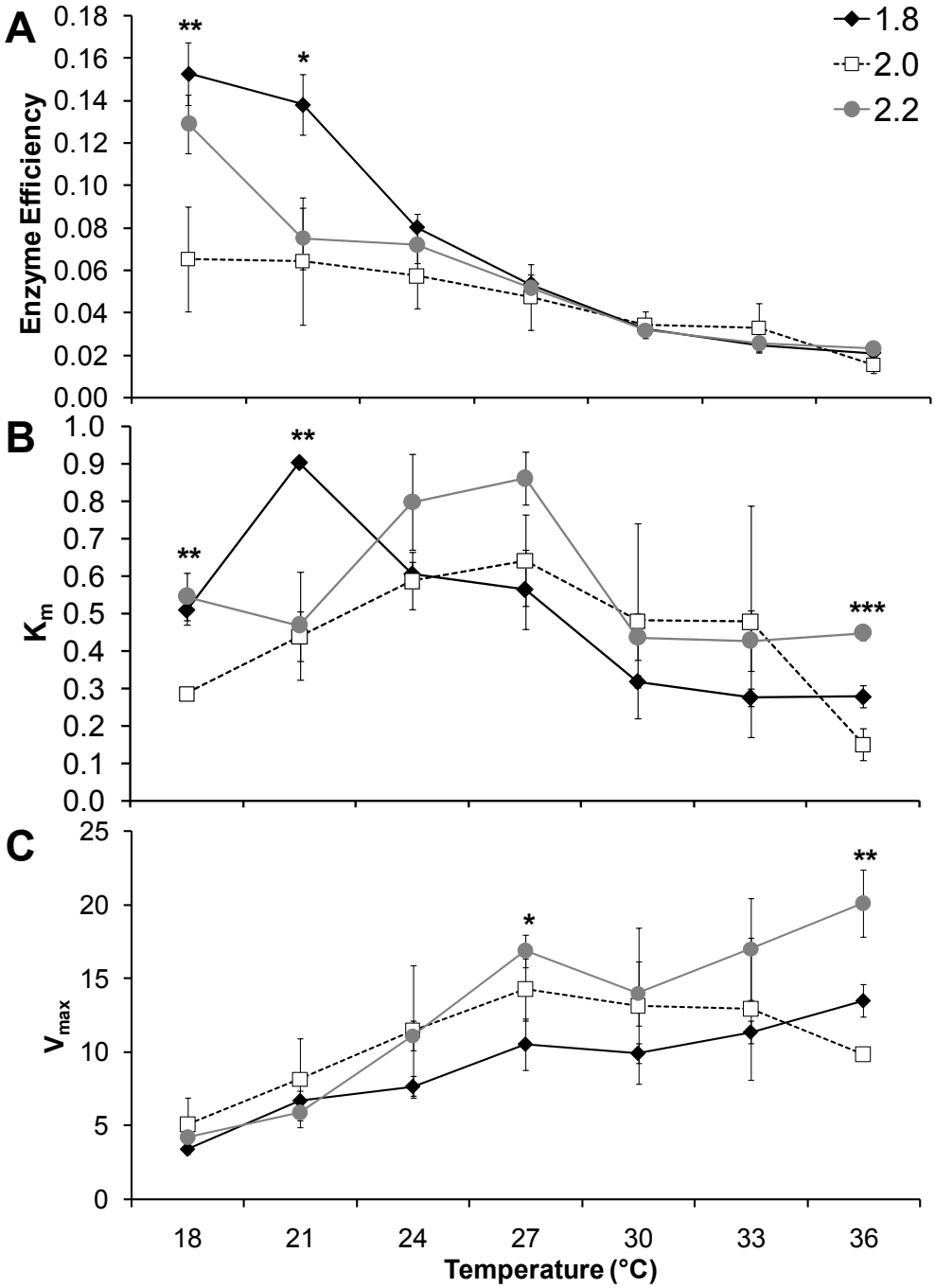

Figure 4.3. Kinetic performance across a temperature gradient of 3 *Idh*-1 alleles in the cricket *Allonemobius socius*. Enzyme efficiency (A) was measured as the increase in reaction velocity with increasing substrate concentration (see Methods). K_m (B) and V_{max} (C) were calculated using standard Michaelis-Menten methods. Means ± SE are given (*n*=3 individuals per genotype). Individual ANOVA's were used to test for differences between genotypes at each temperature and significant differences indicated (**P*<0.1, ***P*<0.05, ****P*<0.01).

Table 4.4. Repeated-measures ANOVA on V_{max} across a temperature gradient.

Between Subjects				
Source	df	MS	*F*	*P*
Genotype	2	73.7718	1.67	0.2787
Error	5	44.2408		
Within Subjects				
Source	df	MS	*F*	*P*
Temperature	6	119.7032	25.46	<0.0001
Genotype x Temperature	12	13.5814	2.89	0.0133
Error (temperature)	30	4.7018		

Within-subject *P*-values are with Huynh-Feldt corrections.

4.5 Discussion

Previous studies have shown that the effect of temperature on enzyme performance can affect many aspects of organismal fitness, including growth rate and size at maturity (van der Have and de Jong 1996). Thus, based on previous geographic and empirical data for *A. socius*, I hypothesized that the observed distributions of *Idh-1* alleles, including the high frequency of the 2.0 allele and the cline in the frequency of the 2.2 allele, had resulted from natural selection on differential enzymatic performance across thermal environments. By conducting kinetics assays, I found a significant GxE between *Idh*-1 genotype and temperature on 3 measures of enzyme kinetic performance in *A. socius*. Additionally, I found significant differences in performance among alleles at the two lowest and one highest temperatures.

Specifically, at 18°C, 2.0 was more efficient because of higher substrate affinity (lower K_m) of the Idh enzyme than the 1.8 and 2.2 alleles. At 21°C, the 2.0 and 2.2 alleles both had higher substrate affinities that that of 1.8 for those two parameters; there was no significant difference among alleles in V_{max} at low temperatures. At 36°C, there was no significant difference in overall efficiency between alleles. However, K_m and V_{max} were significantly higher for the 2.2 allele. K_m and V_{max} for 2.0 were lower than the other alleles at this temperature, while the 1.8 allele had intermediate values for K_m and V_{max} when compared to the other two. Overall, I found that K_m and V_{max} varied at critical temperatures for all 3 alleles, in contrast to the results

of Johns and Somero (2004), who found that a cold-adapted allele of lactate dehydrogenase (*Ldh*-4) had a higher K_m than warm-adapted alleles across all assay temperatures in Pacific damselfishes.

These results, in general, are consistent with hypothesized results based on the naturally-occurring distribution of *Idh-1* allele frequencies and previous fitness data (Huestis and Marshall 2006b). Specifically, the 2.0 allele was the most efficient allele at lower temperatures ($\leq$ 27°C) and equivalent in efficiency to the other two alleles at higher temperatures (> 27°C; Figure 4.3A); these data indicate that the 2.0 allele performs well across the widest range of temperatures, and it is therefore not unexpected that it occurs at approximately 50% frequency in all populations in the eastern United States (Figure 4.1; Huestis and Marshall 2006b). Conversely, the 2.2 allele had the highest maximum reaction velocity (V_{max}) at higher temperatures ($\geq$ 27°C); this allele occurs at approximately equal frequency with the 2.0 allele at lower latitudes, but is not as common to the north (Figure 4.1). However, it is somewhat surprising that the 1.8 allele did not perform better at lower temperatures (18 and 21°C), given the earlier fitness data and its geographic distribution (Huestis and Marshall 2006b). At the lowest assay temperature (18°C), the 1.8 allele was equivalent to the 2.2 allele in both efficiency and substrate affinity, while the V_{max} of all 3 alleles were not significantly different. Thus, we hypothesize that the increased frequency of the 1.8 allele in northern latitudes may be more reflective of processes such as genetic drift, rather than a strict by-product of natural selection. In all, our findings support the hypotheses that 1) the 2.0 allele is common across all thermal environments due to superior performance, for at least some kinetic parameters, relative to the other two alleles at many different temperatures, and 2) the clinal distribution of the 2.2 allele is driven by selection on thermal performance.

In general, enzymatic performance depends on temperature and typically decreases at high temperatures due to degradation and inactivation of enzyme molecules (Daniel et al. 2001; Ma et al. 2001; Peterson et al. 2004, 2007). Interestingly, we did not observe a decrease in

performance, even in our highest temperature assays (Figure 4.3). Therefore, the temperature range over which I conducted my assays appears not to have exceeded the thermostability threshold of the isocitrate dehydrogenase enzyme over the time period of the assay, which is not surprising given that temperatures were chosen to reflect the natural conditions experienced by field populations.

At extreme reaction temperatures (cold or hot), researchers typically observe a tradeoff between K_m and V_{max}. Although an enzyme may tightly bind a substrate, the rate at which the substrate is converted to product may be lower (Watt 1977, 1983, 1992; Xu et al. 2003a,b). These data appear to show this tradeoff at our highest assay temperature (36°C), as the 2.2 allele has both the highest K_m and V_{max}, while 2.0 has the lowest (Figure 4.3). These data may be influenced by overall differences in amounts of Idh enzyme produced by individuals of different genotypes. However, if differences in performance of alleles at 36°C were due to differences in Idh concentrations among the samples, we would expect alleles to show similar patterns of performance for these two parameters across all temperatures, not just 36°C. Overall, I found that at low temperatures all three alleles have a similar V_{max} but significant differences in efficiency and K_m, while at high temperatures all three alleles have the same efficiency but significant differences in K_m and V_{max} (see Figure 4.3). These findings point to a clear performance tradeoff, in that a single allele cannot perform optimally for all measures of kinetic performance across all temperatures.

When one genotype produces different phenotypes over a range of environmental conditions, the relationship between the phenotype produced and the environment is known as a reaction norm (Schlichting and Pigliucci 1998). Similarly, a genotype-by-environment interaction (GxE) occurs when different genotypes produce different reaction norms across the same environmental conditions. Reaction norms and GxE interactions are thought to be adaptive for species living in temporally-variable environments or with wide geographic ranges, as populations may experience different environmental conditions across the species' range or

during the year. Such environmental variation can lead to balancing selection on alternative alleles and/or the evolution of phenotypic plasticity. Given GxE and spatially- or temporally-variable environments across a species' range, balancing selection is predicted to maintain genetic diversity in natural populations (Karl and Avise 1992; Schmidt et al. 2000; Wheat et al. 2006). It appears that all 3 alleles are being maintained in *A. socius*, possibly due to differences in GxE across temperatures. To further test the hypothesis that natural selection (and/or genetic drift) is acting on the *Idh*-1 locus in *A. socius*, I am currently sequencing the protein-coding region for all alleles from multiple individuals spanning the geographic range of this species.

The link between allozyme allele frequencies, differential thermal performance of alleles, and natural environmental gradients has been shown in several other recent studies. For example, Piccino et al. (2004) found significant differences in phosophoglucomutase (*Pgm*-1) allele frequencies between populations of the polychaete *Alvinella pompejana* which inhabited either newly-created or older hydrothermal vents. The enzyme allele at highest frequency in populations living near recently-established, warmer vents was both more thermostable and had higher activity at warmer temperatures than the allele found in populations dwelling in older, cooler vents (Piccino et al. 2004). Similar results have been found for *Ldh* in species of Pacific damselfish of the genera *Chromis* (Johns and Somero 2004) and *Sphyraena* (Holland et al. 1997) inhabiting different thermal regimes. Together, these and other studies indicate that some allozyme loci are not neutral markers of diversity, but rather that the geographic distributions of alleles can be a consequence of environmental conditions and the differential performance of alleles in those environments. Thus, common alleles may be common due to their enhanced performance relative to other, less-common alleles, while clinal distributions can be attributed to either selection or drift across thermal gradients. The distribution of *Idh*-1 alleles in *A. socius* appears to be another example of this phenomenon, as we are now able to link a GxE between genotype and temperature on enzyme kinetic performance to the natural distribution of alleles in this species.

REFERENCES

Andrés, J. A., L. S. Maroja, S. M. Bogdanowicz, W. J. Swanson, and R. G. Harrison. 2006. Molecular evolution of seminal proteins in field crickets. Molecular Biology and Evolution 23:1574–1584.

Arakane, Y., S. Muthukrishnan, K. J. Kramer, C. A. Specht, Y. Tomoyasu, M. D. Lorenzen, M. Kanost, and R. W. Beeman. 2005. The *Tribolium* chitin synthase genes *TcCHS1* and *TcCHS2* are specialized for synthesis of epidermal cuticle and midgut peritrophic matrix. Insect Molecular Biology 14:453–463.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. 2000. Gene Ontology: tool for the unification of biology. Nature Genetics 25:25–29.

Audic, S., and J. M. Claverie. 1997. The significance of digital gene expression profiles. Genome Research 7:986–995.

Bachem, C. W. B., R. S. van der Hoeven, S. M. de Bruijn, D. Vreugdenhil, M. Zabeau, and R. G. F. Visser. 1996. Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. Plant Journal 9:745–753.

Bachem, C. W. B., R. J. F. J. Oomen, and R. G. F. Visser. 1998. Transcript imaging with cDNA-AFLP: a step-by-step protocol. Plant Molecular Biology Reporter 16:157–173.

Bensch, S., and M. Akesson. 2005. Ten years of AFLP in ecology and evolution: why so few animals? Molecular Ecology 14:2899–2914.

Bergmann, F., and H. R. Gregorius. 1993. Ecogeographical distribution and thermostability of isocitrate dehydrogenase (Idh) alloenzymes in European Silver Fir (*Abies alba*). Biochemical Systematics and Ecology 21:597–605.

Berrigan, D., and E. L. Charnov. 1994. Reaction norms for age and size at maturity in response to temperature: a puzzle for life historians. Oikos 70:474–748.

Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Molecular Biology and Evolution 23:773–783.

Bradford, M. J., and D. A. Roff. 1993. Bet hedging and the diapause strategies of the cricket *Allonemobius fasciatus*. Ecology 74:1129–1135.

Bradford, M. J., and D. A. Roff. 1995. Genetic and phenotypic sources of life history variation along a cline in voltinism in the cricket *Allonemobius socius*. Oecologia 103:319–326.

Bradford, M. J., and D. A. Roff. 1997. An empirical model of diapause strategies of the cricket *Allonemobius socius*. Ecology 78:442–451.

Braswell, W. E., J. A. Andres, L. S. Maroja, R. G. Harrison, D. J. Howard, and W. J. Swanson. 2006. Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome 49:1069–1080.

Britch, S. C., M. L. Cain, and D. J. Howard. 2001. Spatio-temporal dynamics of the *Allonemobius fasciatus-A. socius* mosaic hybrid zone: a 14-year perspective. Molecular Ecology 10:627–638.

Britch, S. C., E. J. Swartout, D. D. Hampton, M. L. Draney, J. Chu, J. L. Marshall, and D. J. Howard. 2007. Genetic architecture of conspecific sperm precedence in *Allonemobius fasciatus* and *A. socius*. Genetics 176:1209–1222.

Campbell, C. L., and W. C. Wilson. 2002. Differentially expressed midgut transcripts in *Culicoides sonorensis* (Diptera: Ceratopogonidae) following Orbivirus (Reoviridae) oral feeding. Insect Molecular Biology 11:595–604.

Carroll, S. B., J. K. Grenier, and S. D. Weatherbee. 2005. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design, 2nd edition. Blackwell Publishing, Oxford.

Cayre, M., C. Strambi, A. Tirard, M. Renucci, P. Charpin, R. Augier, and A. Strambi. 1995. Effects of juvenile hormone on polyamines of the fat body and neural tissue of the cricket *Acheta domesticus*. Comparative Biochemistry and Physiology Part A 111:241–250.

Chen, B., T. Kayukawa, H. Jiang, A. Montiero, S. Hoshizaki, and Y. Ishikawa. 2005. *DaTrypsin*, a novel clip-domain serine proteinase gene up-regulated during winter and summer diapauses of the onion maggot, *Delia antiqua*. Gene 347:115–123.

Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.

Cornish-Bowden, A. 2004. Fundamentals of Enzyme Kinetics, 3rd edition. Portland Press, London.

daCunha, G. L., and A. K. DeOliveira. 1996. Citric acid cycle: a mainstream metabolic pathway influencing life span in *Drosophila melanogaster*? Experimental Gerontology 31:705–715.

Daibo, S., M. T. Kimura, and S. G. Goto. 2001. Upregulation of genes belonging to the drosomycin family in diapausing adults of *Drosophila triauraria*. Gene 278:177–184.

Daniel, R. M., M. J. Danson, and R. Eisenthal. 2001. The temperature optima of enzymes: a new perspective on an old phenomenon. Trends in Biochemical Sciences 26:223–225.

Danks, H. V. 1987. Insect dormancy: an ecological perspective. Biological Survey of Canada Monograph Series No. 1.

Dean, A. M., and G. B. Golding. 1997. Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. Proceedings of the National Academy of Sciences of the United States of America 94:3104–3109.

Denlinger, D. L. 2002. Regulation of diapause. Annual Review of Entomology 47:93–122.

Denlinger, D. L., K. H. Joplin, R. D. Flannagan, S. P. Tammariello, M.-L. Zhang, G. D. Yocum, and K.-Y. Lee. 1995. Diapause-specific gene expression. *In* Molecular Mechanisms of Insect Metamorphosis and Diapause (A. Suzuki, H. Kataoka, and S. Masumoto, eds.). Industrial Publishing and Consulting, Inc., Tokyo.

Denlinger, D. L., J. P. Rinehart, and G. D. Yocum. 2001. Stress proteins: a role in insect diapause? *In* Insect Timing: Circadian Rhythmicity to Seasonality (D. L. Denlinger, J. M. Giebultowicz, and D. S. Saunders, eds.). Elsevier, Amsterdam.

Diatchenko, L., Y. F. C. Lau, A. P. Campbell, A. Chenchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, N. Gurskaya, E. D. Sverdlov, and P. D. Siebert. 1996. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proceedings of the National Academy of Sciences of the United States of America 93:6025–6030.

Diatchenko, L., S. Lukyanov, Y. F. C. Lau, and P. D. Siebert. 1999. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. Methods in Enzymology 303:349–380.

Donson, J., Y. W. Fang, G. Espiritu-Santo, W. M. Xing, A. Salazar, S. Miyamoto, V. Armendarez, and W. Volkmuth. 2002. Comprehensive gene expression analysis by transcript profiling. Plant Molecular Biology 48:75–97.

Dzitoyeva, S., N. Dimitrijevic, and H. Manev. 2003. Identification of a novel *Drosophila* gene, *beltless*, using injectable embryonic and adult RNA interference (RNAi). BMC Genomics 4:33.

Eanes, W. F. 1999. Analysis of selection on enzyme polymorphisms. Annual Review of Ecology and Systematics 30:301–326.

Emrich, S. J., W. B. Barbazuk, L. Li, and P. S. Schnable. 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Research 17:69–73.

Fedorka, K. M., and T. A. Mousseau. 2002a. Material and genetic benefits of female multiple mating and polyandry. Animal Behaviour 64:361–367.

Fedorka, K. M., and T. A. Mousseau. 2002b. Nuptial gifts and the evolution of male body size. Evolution 56:590–596.

Fedorka, K. M., and T. A. Mousseau. 2004. Female mating bias results in conflicting sex-specific offspring fitness. Nature 429:65–67.

Fedorka, K. M., W. E. Winterhalter, and T. A. Mousseau. 2007. The evolutionary genetics of sexual size dimorphism in the cricket *Allonemobius socius*. Heredity 99:218–223.

Fields, P. A. 2001. Review: Protein function at thermal extremes: balancing stability and flexibility. Comparative Biochemistry and Physiology A 129:417–431.

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

Flannagan, R. D., S. P. Tammariello, K. H. Joplin, R. A. Cikra-Ireland, G. D. Yocum, and D. L. Denlinger. 1998. Diapause-specific gene expression in pupae of the flesh fly *Sarcophaga crassipalpis*. Proceedings of the National Academy of Sciences of the United States of America 95:5616–5620.

Freeman, W. M., S. J. Walker, and K. E. Vrana. 1999. Quantitative RT-PCR: pitfalls and potential. BioTechniques 26:112–125.

Fulton, B. B. 1931. A study of the genus *Nemobius* (Orthoptera: Gryllidae). Annals of the Entomological Society of America 24:205–237.

Funaguma, S., S. I. Hashimoto, Y. Suzuki, N. Omura, S. Sugano, K. Mita, S. Katsuma, and T. Shimada. 2007. SAGE analysis of early oogenesis in the silkworm, *Bombyx mori*. Insect Biochemistry and Molecular Biology 37:147–154.

Gerisch, B., and A. Antebi. 2004. Hormonal signals produced by DAF-9/cytochrome P450 regulate *C. elegans* dauer diapause in response to environmental cues. Development 131:1765–1776.

Gheysen, G. and B. Vanholme. 2007. RNAi from plants to nematodes. Trends in Biotechnology 25:89–92.

Gibson, U. E. M., C. A. Heid, and P. M. Williams. 1996. A novel method for real time quantitative RT-PCR. Genome Research 6:995–1001.

Gindling, H. L., G. A. Rosenthal, and D. L. Dahlman. 1995. Purification of L-arginine kinase from the tobacco budworm, *Heliothis virescens* (Noctuidae) and its function in L-canavanine detoxification. Insect Biochemistry and Molecular Biology 25:933–938.

Götz, S., J. M. García-Gómez, J. Terol, T. D. Williams, M. José Nueda, M. Robles, M. Talón, J. Dopazo, and A. Conesa. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research 36: 3420–3435.

Gracey, A. Y., and A. R. Cossins. 2003. Application of microarray technology in environmental and comparative physiology. Annual Review of Physiology 65:231–259.

Graff, J., S. Jemielity, J. D. Parker, K. M. Parker, and L. Keller. 2007. Differential gene expression between adult queens and workers in the ant *Lasius niger*. Molecular Ecology 16:675–683.

Gregory, P. G., and D. J. Howard. 1994. A post-insemination barrier to fertilization isolates two closely-related ground crickets. Evolution 48:705–710.

Guerrero, F. D., K. G. Bendele, A. C. Chen, A. Y. Li, R. J. Miller, E. Pleasance, R. Varhol, M. E. Rousseau, and V. M. Nene. 2007. Serial analysis of gene expression in the southern cattle tick following acaricide treatment of larvae from organophosphate resistant and susceptible strains. Insect Molecular Biology 16:49–60.

Habu, Y., S. Fukada-Tanaka, Y. Hisatomi, and S. Iida. 1997. Amplified restriction fragment length polymorphism-based mRNA fingerprinting using a single restriction enzyme that recognizes a 4-bp sequence. Biochemical and Biophysical Research Communications 234:516–521.

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

Harris, H. 1966. Enzyme polymorphisms in man. Proceedings of the Royal Society of London B 164:298–310.

Harris, M. A., J. Clark, A. Ireland, J. Lomax, M. Ashburner, R. Foulger, K. Eilbeck, S. Lewis, B. Marshall, C. Mungall, J. Richter, G. M. Rubin, J. A. Blake, C. Bult, M. Dolan, H. Drabkin, J. T. Eppig, D. P. Hill, L. Ni, M. Ringwald, R. Balakrishnan, J. M. Cherry, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. Engel, D. G. Fisk, J. E. Hirschman, E. L. Hong, R. S. Nash, A. Sethuraman, C. L. Theesfeld, D. Botstein, K. Dolinski, B. Feierbach, T. Berardini, S. Mundodi, S. Y. Rhee, R. Apweiler, D. Barrell, E. Camon, E. Dimmer, V. Lee, R. Chisholm, P. Gaudet, W. Kibbe, R. Kishore, E. M. Schwarz, P. Sternberg, M. Gwinn, L. Hannick, J. Wortman, M. Berriman, V. Wood, N. de la Cruz, P. Tonellato, P. Jaiswal, T. Seigfried, and R. White. 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 32:D258–D261.

Hawkins, B. J., G. B. Sweet, D. H. Greer, and D. O. Bergin. 1991. Genetic variation in the frost hardiness of *Podocarpus totara*. New Zealand Journal of Botany 29:455–458.

Heid, C. A., J. Stevens, K. J. Livak, and P. M. Williams. 1996. Real time quantitative PCR. Genome Research 6:986–994.

Hockham, L. R., J. A. Graves, and M. G. Ritchie. 2001. Variable maternal control of facultative egg diapause in the bushcricket *Ephippiger ephippiger*. Ecological Entomology 36:143–147.

Hoekstra, H. E., and J. A. Coyne. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution 61:995–1016.

Hoffmann, A. A., C. M. Sgro, and S. H. Lawler. 1995. Ecological population genetics: the interface between genes and the environment. Annual Review of Genetics 29:349–370.

Hoffmann, A. A., J. Shirriffs, and M. Scott. 2005. Relative importance of plastic vs. genetic factors in adaptive differentiation: geographical variation for stress resistance in *Drosophila melanogaster* from Eastern Australia. Functional Ecology 19:222–227.

Holland, L. Z., M. McFallNgai, and G. N. Somero. 1997. Evolution of lactate dehydrogenase-A homologs of barracuda fishes (genus *Sphyraena*) from different thermal environments: differences in kinetic properties and thermal stability are due to amino acid substitutions outside the active site. Biochemistry 36:3207–3215.

Horike, N., and H. Sonobe. 1999. Ecdysone 20-monooxygenase in eggs of the silkworm, *Bombyx mori*: enzymatic properties and developmental changes. Archives of Insect Biochemistry and Physiology 41:9–17.

Hotz, H., and R. D. Semlitsch. 2000. Differential performance among LDH-B genotypes in *Rana lessonae* tadpoles. Evolution 54:1750–1759.

Howard, D. J. 1982. Speciation and coexistence in a group of closely-related ground crickets. Ph.D. Dissertation, Yale University, New Haven, Connecticut.

Howard, D. J. 1983. Electrophoretic survey of eastern North American *Allonemobius* (Orthoptera: Gryllidae): evolutionary relationships and the discovery of three new species. Annals of the Entomological Society of America 76:1014–1021.

Howard, D. J. 1986. A zone of overlap and hybridization between two ground cricket species. Evolution 40:34–43.

Howard, D. J., and D. G. Furth. 1986. Review of the *Allonemobius fasciatus* (Orthoptera: Gryllidae) complex with the description of two new species separated by electrophoresis, songs, and morphometrics. Annals of the Entomological Society of America 79:472–481.

Howard, D. J., and G. L. Waring. 1991. Topographic diversity, zone width, and the strength of reproductive isolation in a zone of overlap and hybridization. Evolution 45:1120–1135.

Huang, J. H., X. X. Miao, W. R. Jin, P. Couble, K. Mita, Y. Zhang, W. B. Liu, L. J. Zhuang, Y. Shen, C. Keime, O. Gandrillon, P. Brouilly, J. Briolay, G. P. Zhao, and Y. P. Huang. 2005. Serial analysis of gene expression in the silkworm, *Bombyx mori*. Genomics 86:233–241.

Hubby, J. L., and R. C. Lewontin. 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. Genetics 54:577–594.

Hudson, M. E. 2007. Sequencing breakthroughs for genomic ecology and evolutionary biology. Molecular Ecology Notes 10.1111/j.1471-8286.2007.02019.x.

Huestis, D. L., and J. L. Marshall. 2006a. Interaction between maternal effects and temperature affects diapause occurrence in the cricket *Allonemobius socius*. Oecologia 146:513–520.

Huestis, D. L., and J. L. Marshall. 2006b. Is natural selection a plausible explanation for the distribution of *Idh*-1 alleles in the cricket *Allonemobius socius*? Ecological Entomology 31:91-98.

Huestis, D. L., and J. L. Marshall. From gene expression to phenotype in insects: approaches for transcriptome analysis. BioScience, *in press*.

Huse, S. M., J. A. Huber, H. G. Morrison, M. L. Sogin, and D. M. Welch. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology 8:R143.

Jaramillo-Correa, J. P., J. Beaulieu, and J. Bousquet. 2001. Contrasting evolutionary forces driving population structure at expressed sequence tag polymorphisms, allozymes and quantitative traits in white spruce. Molecular Ecology 10:2729–2740.

Jasper, H., V. Benes, C. Schwager, S. Sauer, S. Clauder-Munster, W. Ansorge, and D. Bohmann. 2001. The genomic response of the *Drosophila* embryo to JNK signaling. Developmental Cell 1:579–586.

Johns, G. C., and G. N. Somero. 2004. Evolutionary convergence in adaptation of proteins to temperature: A_4-lactate dehydrogenases of pacific damselfishes (*Chromis* spp.). Molecular Biology and Evolution 21:314–320.

Kang, L., X. Y. Chen, Y. Zhou, B. W. Liu, W. Zheng, R. Q. Li, J. Wang, and J. Yu. 2004. The analysis of large-scale gene expression correlated to the phase changes of the migratory locust. Proceedings of the National Academy of Sciences of the United States of America 101:17611–17615.

Karl, S. A., and J. C. Avise. 1991. Balancing selection at allozyme loci in oysters: implications from nuclear RFLP's. Science 256:100–102.

Kim, H. J., H. J. Je, S. Y. Park, I. H. Lee, B. R. Jin, H. K. Yun, C. Y. Yun, Y. S. Han, Y. J. Kang, and S. J. Seo. 2004. Immune activation of apolipophorin-III and its distribution in hemocyte from *Hyphantria cunea*. Insect Biochemistry and Molecular Biology 34:1011–1023.

Kreitman, M., and H. Akashi. 1995. Molecular evidence for natural selection. Annual Review of Ecology and Systematics 26:403–422.

Kropf, M., J. W. Kadereit, and H. P. Comes. 2002. Late Quaternary distributional stasis in the submediterranean mountain plant *Anthyllis montana* L. (Fabaceae) inferred from ITS sequences and amplified fragment length polymorphism markers. Molecular Ecology 11: 447–463.

Kuhn, E. 2001. From library screening to microarray technology: strategies to determine gene expression profiles and to identify differentially regulated genes in plants. Annals of Botany 87:139–155.

Lash, A. E., C. M. Tolstoshev, L. Wagner, G. D. Schuler, R. L. Strausberg, G. J. Riggins, and S. F. Altschul. 2000. SAGEmap: a public gene expression resource. Genome Research 10:1051–1060.

Levitzki, A., and D. E. Koshland Jr. 1969. Negative cooperativity in regulatory enzymes. Proceedings of the National Academy of Sciences of the United States of America 62:1121–1128.

Lewontin, R. C., and J. L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595–609.

Liang, P., and A. B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971.

Liang, P., D. Bauer, L. Averboukh, P. Warthoe, M. Rohrwild, H. Muller, M. Strauss, and A. B. Pardee. 1995. Analysis of altered gene expression by differential display. Methods in Enzymology 254:304–321.

Liao, V. H. C., and J. H. Freedman. 2002. Differential display analysis of gene expression in invertebrates. Cellular and Molecular Life Sciences 59:1256–1263.

Liefting, M., and J. Ellers. 2008. Habitat-specific differences in thermal plasticity in natural populations of a soil arthropod. Biological Journal of the Linnean Society 94:265–271.

Lisitsyn, N., N. Lisitsyn, and M. Wigler. 1993. Cloning the differences between two complex genomes. Science 259:946–951.

Lycett, G. J., L. A. McLaughlin, H. Ranson, J. Hemingway, F. C. Kafatos, T. G. Loukeris, and M. J. I. Paine. 2006. *Anopheles gambiae* P450 reductase is highly expressed in oenocytes and *in vivo* knockdown increases permethrin susceptibility. Insect Molecular Biology 15:321–327.

Ma, Y. F., D. E. Evans, S. J. Logue, and P. Langridge. 2001. Mutations of barley β-amylase that improve substrate-binding affinity and thermostability. Molecular Genetics and Genomics 266:345–352.

Maeda, S., A. Nakashima, R. Yamada, N. Hara, Y. Fujimoto, Y. Ito, and H. Sonobe. 2008. Molecular cloning of ecdysone 20-hydroxylase and expression pattern of the enzyme during embryonic development of silkworm *Bombyx mori*. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 149:507–516.

Maestro, J. L., and X. Bellés. 2006. Silencing allatostatin expression using double-stranded RNA targeted to preproallatostatin mRNA in the german cockroach. Archives of Insect Biochemistry and Physiology 62:73–79.

Magurran, A. E. 1988. Ecological Diversity and its Measurement. Princeton University Press, Princeton, NJ.

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

Marshall, J. L. 2004. The *Allonemobius-Wolbachia* host-endosymbiont system: evidence for rapid speciation and against reproductive isolation driven by cytoplasmic incompatibility. Evolution 58:2409–2425.

Marshall, J. L. 2007. Rapid evolution of spermathecal duct length in the *Allonemobius socius* complex of crickets: species, population and *Wolbachia* effects. PLoS ONE 2:e720.

Marshall, J. L., D. L. Huestis, Y. Hiromasa, C. Oppert, S. A. Marshall, and B. Oppert. Adult RNAi in the cricket *Allonemobius*: knockdown of an ejaculate protein. BMC Biotechnology, *in review*.

Martin, K. J., and A. B. Pardee. 2000. Identifying expressed genes. Proceedings of the National Academy of Sciences of the United States of America 97:3789–3791.

Masaki, S. 1996. Geographical variation of life cycle in crickets (Ensifera: Grylloidea). European Journal of Entomology 93:281–302.

Masinde, G., X. M. Li, D. J. Baylink, B. Nguyen, and S. Mohan. 2005. Isolation of wound healing/regeneration genes using restrictive fragment differential display-PCR in MRL/MPJ and C57BL/6 mice. Biochemical and Biophysical Research Communications 330:117–122.

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

Metta, M., R. Gudavalli, J. M. Gibert, and C. Schlotterer. 2006. No accelerated rate of protein evolution in male-biased *Drosophila pseudoobscura* genes. Genetics 174:411–420.

Meyering-Vos, M., S. Merz, M. Sertkol, and K. M. Hoffmann. 2006. Functional analysis of the allatostatin A gene in the cricket *Gryllus bimaculatus* and the armyworm *Spodoptera frugiperda*. Insect Biochemistry and Molecular Biology 36:492–504.

Michaud, M. R., and D. L. Denlinger. 2006. Oleic acid is elevated in cell membranes during rapid cold-hardening and diapause in the flesh fly, *Sarcophaga crassipalpis*. Journal of Insect Physiology 52:1073–1082.

Mita, K., M. Morimoyo, K. Okano, Y. Koike, J. Nohata, H. Kawasaki, K. Kadono-Okuda, K. Yamamoto, M. G. Suzuki, T. Shimada, M. R. Goldsmith, and S. Maeda. 2003. The construction of an EST database for *Bombyx mori* and its application. Proceedings of the National Academy of Sciences of the United States of America 100:14121–14126.

Moribe, Y., T. Niimi, O. Yamashita, and T. Yaginuma. 2001. *Samui*, a novel cold-inducible gene, encoding a protein with a BAG domain similar to silencer of death domains (SODD/BAG-4), isolated from *Bombyx* diapause eggs. European Journal of Biochemistry 268:3432–3442.

Mousseau, T. A. 1991. Geographic variation in maternal-age effects on diapause in a cricket. Evolution 45:1053–1059.

Mousseau, T. A., and H. Dingle. 1991. Maternal effects in insect life histories. Annual Review of Entomology 36:511–534.

Mousseau, T. A., and C. W. Fox. 1998. The adaptive significance of maternal effects. Trends in Ecology and Evolution 13:403–407.

Mousseau, T. A., and D. A. Roff. 1989a. Adaptation to seasonality in a cricket: patterns of phenotypic and genotypic variation in body size and diapause expression along a cline in season length. Evolution 43:1483–1496.

Mousseau, T. A., and D. A. Roff. 1989b. Geographic variability in the incidence and heritability of wing dimorphism in the striped ground cricket, *Allonemobius fasciatus*. Heredity 62:315–318.

Mousseau, T. A., and D. A. Roff. 1995. Genetic and environmental contributions to geographic variation in the ovipositor length of a cricket. Ecology 76:1473–1482.

Mutti, N. S., Y. Park, J. C. Reese, and G. R. Reeck. 2006. RNAi knockdown of a salivary transcript leading to lethality in the pea aphid, *Acyrthosiphon pisum*. Journal of Insect Science 6:38.

Nelson, C. M., K. E. Ihle, M. K. Fondrk, R. E. Page Jr., and G. V. Amdam. 2007. The gene vitellogenin has multiple coordinating effects on social organization. PLOS Biology 5:e62.

Niwa, R., T. Matsuda, T. Yoshiyama, T. Namiki, K. Mita, Y. Fujimoto, and H. Kataoka. 2004. CYP306A1, a cytochrome P450 enzyme, is essential for ecdysteroid biosynthesis in the prothoracic glands of *Bombyx* and *Drosophila*. Journal of Biological Chemistry 279:35942–35949.

Oh, J. H., Y. J. Jeon, S. Y. Jeong, S. M. Hong, J. S. Lee, S. K. Nho, S. W. Kang, and N. S. Kim. 2006. Gene expression profiling between embryonic and larval stages of the silkworm, *Bombyx mori*. Biochemical and Biophysical Research Communications 343:864–872.

Olvido, A. E., and T. A. Mousseau. 1998. Seasonal effects on oviposition behavior in *Allonemobius socius* (Orthoptera: Gryllidae): test of the sense of malaise hypothesis. Annals of the Entomological Society of America 91:488–492.

Olvido, A. E., S. Busby, and T. A. Mousseau. 1998. Oviposition and incubation environmental effects on embryonic diapause in a ground cricket. Animal Behaviour 55:331–336.

Olvido, A. E., E. S. Elvington, and T. A. Mousseau. 2003. Relative effects of climate and crowding on wing polymorphism in the southern ground cricket, *Allonemobius socius* (Orthoptera: Gryllidae). Florida Entomologist 86:158–164.

Oppert, B., K. J. Kramer, and W. H. McGaughey. 1997. Rapid microplate assay for substrates and inhibitors of proteinase mixtures. BioTechniques 23:70–72.

Pasyukova, E. G., C. Vieira, and T. F. C. MacKay. 2000. Deficiency mapping of quantitative trait loci affecting longevity in *Drosophila melanogaster*. Genetics 156:1129–1146.

Pemberton, J. M., S. D. Albon, F. E. Guinness, and T. H. Cluttonbrock. 1991. Countervailing selection in different fitness components in female red deer. Evolution 45:93–103.

Peterson, M. E., R. Eisenthal, M. J. Danson, A. Spence, and R. M. Daniel. 2004. A new intrinsic thermal parameter for enzymes reveals true temperature optima. Journal of Biological Chemistry 279:20717–20722.

Peterson, M. E., R. M. Daniel, M. J. Danson, and R. Eisenthal. 2007. The dependence of enzyme activity on temperature: determination and validation of parameters. Biochemical Journal 402:331–337.

Philippi, T., and J. Seger. 1989. Hedging one's evolutionary bets, revisited. Trends in Ecology and Evolution 4:41–44.

Piccino, P., F. Viard, P. M. Sarradin, N. Le Bris, D. Le Guen, and D. Jollivet. 2004. Thermal selection of PGM allozymes in newly founded populations of the thermotolerant vent polychaete *Alvinella pompejana*. Proceedings of the Royal Society of London Series B – Biological Sciences 271:2351–2359.

Pleasance, E. D., M. A. Marra, and S. J. M. Jones. 2003. Assessment of SAGE in transcript identification. Genome Research13: 1203–1215.

Powers, D. A., T. Lauerman, D. Crawford, and L. DiMichele 1991. Genetic mechanisms for adapting to a changing environment. Annual Review of Genetics 25:629–659.

Qiu, L. H., J. R. Gao, and J. M. Clark. 2005. Sequencing and characterization of a cDNA encoding a ferritin subunit of Colorado potato beetle, *Leptinotarsa decemlineata*. Archives of Insect Biochemistry and Physiology 60:140–150.

Qiu, Z., S. C. M. Tsoi, and T. H. MacRae. 2007. Gene expression in diapause-destined embryos of the crustacean, *Artemia franciscana*. Mechanisms of Development 124:856–867.

Reineke, A., and S. Löbmann. 2005. Gene expression changes in *Ephestia kuehniella* caterpillars after parasitization by the endoparasitic wasp *Venturia canescens* analyzed through cDNA-AFLPs. Journal of Insect Physiology 51:923–932.

Reineke, A., O. Schmidt, and C. P. W. Zebitz. 2003. Differential gene expression in two strains of the endoparasitic wasp *Venturia canescens* identified by cDNA-amplified fragment length polymorphism analysis. Molecular Ecology 12:3485–3492.

Ribeiro, J. M. C. 2003. A catalogue of *Anopheles gambiae* transcripts significantly more or less expressed following a blood meal. Insect Biochemistry and Molecular Biology 33:865–882.

Riddoch, B. J. 1993. The adaptive significance of electrophoretic mobility in phosphoglucose isomerase (PGI). Biological Journal of the Linnean Society 50:1–17.

Rinehart, J. P., G. D. Yocum, and D. L. Deninger. 2000. Developmental upregulation of inducible hsp70 transcripts, but not the cognate form, during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*. Insect Biochemistry and Molecular Biology 30:515–521.

Rinehart, J. P., R. M. Robich, and D. L. Denlinger. 2006. Enhanced cold and desiccation tolerance in diapausing adults of *Culex pipiens*, and a role for *Hsp70* in response to cold shock but not as a component of the diapause program. Journal of Medical Entomology 43:713–722.

Rinehart, J. P., A. Li, G. D. Yocum, R. M. Robich, S. A. L. Hayward, and D. L. Denlinger. 2007. Up-regulation of heat shock proteins is essential for cold survival during insect diapause. Proceedings of the National Academy of Sciences of the United States of America 104:11130–11137.

Robich, R. M., J. P. Rinehart, L. J. Kitchen, and D. L. Denlinger. 2007. Diapause-specific gene expression in the northern house mosquito, *Culex pipiens* L., identified by suppressive subtractive hybridization. Journal of Insect Physiology 53:235–245.

Robison, B. D., P. A. Wheeler, K. Sundin, P. Sikka, and G. H. Thorgaard. 2001. Composite interval mapping reveals a major locus influencing embryonic development rate in rainbow trout (*Oncorhynchus mykiss*). Journal of Heredity 92:16–22.

Roelofs, D., J. Marien, and N. M. van Straalen. 2007. Differential gene expression profiles associated with heavy metal tolerance in the soil insect *Orchesella cincta*. Insect Biochemistry and Molecular Biology 37:287–295.

Roff, D. A. 2002. Life History Evolution. Sinauer Associates, Inc., Sunderland, MA.

Roff, D. A., and M. J. Bradford. 2000. A quantitative genetic analysis of phenotypic plasticity of diapause induction in the cricket *Allonemobius socius*. Heredity 84:193–200.

Rompf, R., and G. Kahl. 1997. mRNA differential display in agarose gels. BioTechniques 23:28–32.

Ronaghi, M. 2001. Pyrosequencing sheds light on DNA sequencing. Genome Research 11:3-11.

Rosinski-Chupin, I., J. Briolay, P. Brouilly, S. Perrot, S. M. Gomez, T. Chertemps, C. W. Roth, C. Keime, O. Gandrillon, P. Couble, and P. T. Brey. 2007. SAGE analysis of mosquito salivary gland transcriptomes during *Plasmodium* invasion. Cellular Microbiology 9:708–724.

Rossiter, M. C. 1996. Incidence and consequences of inherited environmental effects. Annual Review of Ecology and Systematics 27:451–476.

Rueppell, O., T. Pankiw, D. I. Nielsen, M. K. Fondrk, M. Beye, and R. E. Page. 2004. The genetic architecture of the behavioral ontogeny of foraging in honeybee workers. Genetics 167:1767–1779.

SAS Institute. 2006. SAS Learning Edition 4.1. SAS Institute, Inc., Cary, NC.

Scharf, M. E., D. Wu-Scharf, B. R. Pittendrigh, and G. W. Bennett. 2003. Caste- and development- associated gene expression in a lower termite. Genome Biology 4:R62.

Schiga, S., and H. Numata. 1997. Seasonal changes in the incidence of embryonic diapause in the band-legged ground cricket, *Dianemobius nigrofasciatus*. Zoological Science 14:1015–1018.

Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic Evolution – a Reaction Norm Perspective. Sinauer Associates Inc., Sunderland, MA.

Schmidt, P. S., M. D. Bertness, and D. M. Rand. 2000. Environmental heterogeneity and balancing selection in the acorn barnacle *Semibalanus balanoides*. Proceedings of the Royal Society of London Series B-Biological Sciences 267:379–384.

Scott, J. G., and Z. M. Wen. 2001. Cytochromes P450 of insects: the tip of the iceberg. Pest Management Science 57:958–967.

Sokolova, I. M., and H. O. Portner. 2001. Temperature effects on key metabolic enzymes in *Littorina saxatilis* and *L. obtusata* from different latitudes and shore levels. Marine Biology 139:113–126.

Stearns, S. C. 1992. The Evolution of Life Histories. Oxford University Press, Oxford.

Stekel, D. J., Y. Git, and F. Falciani. 2000. The comparison of gene expression from multiple cDNA libraries. Genome Research 10:2055–2061.

Storz, J. F., and M. W. Nachman. 2003. Natural selection on protein polymorphism in the rodent genus *Peromyscus*: evidence from interlocus contrasts. Evolution 57:2628–2635.

Su, Z.-H., M. Ikeda, Y. Sato, H. Saito, K. Imai, M. Isobe, and O. Yamashita. 1994. Molecular characterization of ovary trehalase of the silkworm, *Bombyx mori* and its transcriptional activation by diapause hormone. Biochemica et Biophysica Acta 1218:366–374.

Sun, D. H., R. Ziegler, C. E. Milligan, S. Fahrbach, and L. M. Schwartz. 1995. Apolipophorin-III is dramatically up-regulated during the programmed death of insect skeletal muscle and neurons. Journal of Neurobiology 26:119–129.

Suzuki, M. G., T. Terada, M. Kobayashi, and T. Shimada. 1999. Diapause-associated transcription of *BmEts*, a gene encoding an ETS transcription factor homolog in *Bombyx mori*. Insect Biochemistry and Molecular Biology 29:339–347.

Tachibana, S. I., H. Numata, and S. G. Goto. 2005. Gene expression of heat-shock proteins (Hsp23, Hsp70 and Hsp90) during and after larval diapause in the blow fly *Lucilia sericata*. Journal of Insect Physiology 51:641–647.

Tanaka, K., S. Ichinari, K. Wartain, S. Yoshimatsu, and T. Suzuki. 2007. Arginine kinase from the beetle *Cissites cephalotes* (Olivier). Molecular cloning, phylogenetic analysis and enzymatic properties. Insect Biochemistry and Molecular Biology 37:338–345.

Tanaka, S. 1984. Seasonal variation in embryonic diapause of the striped ground cricket, *Allonemobius fasciatus*. Physiological Entomology 9:97–105.

Tanaka, S. 1986a. Developmental characteristics of two closely related species of *Allonemobius* and their hybrids. Oecologia 69:388–394.

Tanaka, S. 1986b. Uptake and loss of water in diapause and non-diapause eggs of crickets. Physiological Entomology 11:343–351.

Tanaka, S. 1987. Changing responses to temperature and moisture of diapausing and developing eggs of *Allonemobius fasciatus* (Orthoptera: Gryllidae). Journal of Insect Physiology 33:635–641.

Tanaka, S. 1992. Factors influencing embryonic diapause in the striped ground cricket, *Allonemobius fasciatus* (Orthoptera, Gryllidae). Japanese Journal of Entomology 60:149–155.

Teixeira, L. A. F., and S. Polavarapu. 2005. Expression of heat shock protein 70 after heat stress during pupal diapause in *Rhagoletis mendax* (Diptera: Tephritidae). Annals of the Entomological Society of America 98:966–972.

Tittiger, C. 2004. Functional genomics and insect chemical ecology. Journal of Chemical Ecology 30:2335–2358.

Toth, A. L., K. Varala, T. C. Newman, F. E. Miguez, S. K. Hutchison, D. A. Willoughby, J. F. Simons, M. Egholm, J. H. Hunt, M. E. Hudson, and G. E. Robinson. 2007. Wasp gene

expression supports an evolutionary link between maternal behavior and eusociality. Science 318:441–444.

Tsai, S. J., and M. C. Wiltbank. 1996. Quantification of mRNA using competitive RT-PCR with standard curve methodology. BioTechniques 21:862–866.

Uno, T., A. Nakasuji, S. Masami, and Y. Aizono. 2004. Expression of cytochrome *c* oxidase subunit 1 gene in the brain at an early stage in the termination of pupal diapause in the sweet potato hornworm, *Agrius convolvuli*. Journal of Insect Physiology 50:35–42.

van der Have, T. M., and G. de Jong. 1996. Adult size in ectotherms: temperature effects on growth and differentiation. Journal of Theoretical Biology 183:329–340.

Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. Science 270:484–487.

Via, S., and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. Evolution 39:505–522.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research 23:4407–4414.

Vrijenhoek, R. C., E. Pfeiler, and J. D. Wetherington. 1992. Balancing selection in a desert stream-dwelling fish, *Poeciliopsis monacha*. Evolution 46:1642–1657.

Watt, W. B. 1977. Adaptation at Specific Loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. Genetics 87:177–194.

Watt, W. B. 1983. Adaptation at specific loci. II. Demographic and biochemical elements in the maintenance of the *Colias* PGI polymorphism. Genetics 103:691–724.

Watt, W. B. 1992. Eggs, enzymes, and evolution: natural genetic variants change insect fecundity. Proceedings of the National Academy of Sciences of the United States of America 89:10608–10612.

Watt, W. B. 1994. Allozymes in evolutionary genetics: self-imposed burden or extraordinary tool. Genetics 136:11–16.

Weber, A. P. M., K. L. Weber, K. Carr, C. Wilkerson, and J. B. Ohlrogge. 2007. Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. Plant Physiology 144:32–42.

Weisner, R. J., J. C. Rüegg, and I. Morano. 1992. Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues. Biochemical and Biophysical Research Communications 183:553–559.

Welsh, J., K. Chada, S. S. Dalal, R. Cheng, D. Ralph, and M. McClelland. 1992. Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Research 20:4965–4970.

West-Eberhard, M. J. 2003. Developmental Plasticity and Evolution. Oxford University Press, Oxford.

Wheat, C. W., W. B. Watt, D. D. Pollock, and P. M. Schulte. 2006. From DNA to fitness differences: sequences and structures of adaptive variants of *Colias* phosphoglucose isomerase (PGI). Molecular Biology and Evolution 23:499–512.

Winterhalter, W. E., and T. A. Mousseau. 2007. Patterns of phenotypic and genetic variation for the plasticity of diapause incidence. Evolution 61:1520–1531.

Xu, Y., G. Feller, C. Gerday, and N. Glansdorff. 2003a. Metabolic enzymes from psychrophilic bacteria: challenge of adaptation to low temperatures in ornithine carbamoyltransferase from *Moritella abyssi*. Journal of Bacteriology 185:2161–2168.

Xu, Y., G. Feller, C. Gerday, and N. Glansdorff. 2003b. *Moritella* cold-active dihydrofolate reductase: are there natural limits to optimization of catalytic efficiency at low temperature? Journal of Bacteriology 185:5519–5526.

Yamashita, O. 1996. Diapause hormone of the silkworm, *Bombyx mori*: structure, gene expression, and function. Journal of Insect Physiology 42:669–679.

Yamashita, O., K. Shiomi, Y. Ishida, N. Katagiri, and T. Niimi. 2001. Insights for future studies on embryonic diapause promoted by molecular analyses of diapause hormone and it's action in *Bombyx mori*. *In* Insect Timing: Circadian Rhythmicity to Seasonality (D. L. Denlinger, J. M. Giebultowicz, and D. S. Saunders, eds.). Elsevier Science, Amsterdam.

Yang, P., H. Tanaka, E. Kuwano, and K. Suzuki. 2008. A novel cytochrome P450 gene (CYP4G25) of the silkmoth *Antheraea yamamai*: cloning and expression pattern in pharate first instar larvae in relation to diapause. Journal of Insect Physiology 54:636–643.

Yang, Z. F., F. T. Zhang, L. L. Zhu, and G. C. He. 2006. Identification of differentially expressed genes in brown planthopper *Nilaparvata lugens* (Hemiptera: Delphacidae) responding to host plant resistance. Bulletin of Entomological Research 96:53–59.

Yocum, G. D. 2003. Isolation and characterization of three diapause-associated transcripts from the Colorado potato beetle, *Leptinotarsa decemlineata*. Journal of Insect Physiology 49:161–169.

Yocum, G. D., K. H. Joplin, and D. L. Denlinger. 1998. Upregulation of a 23 kDa small heat shock protein transcript during pupal diapause in the flesh fly, *Sarcophaga crassipalpis*. Insect Biochemistry and Molecular Biology 28:677–682.

Zeppa, S., G. Guidi, A. Zambonelli, L. Potenza, L. Vallorani, R. Pierleoni, C. Sacconi, and V. Stocchi. 2002. Identification of putative genes involved in the development of *Tuber borchii* fruit body by mRNA differential display in agarose gel. Current Genetics 42:161–168.

Zhong, D. B., D. M. Menge, E. A. Temu, H. Chen, and G. Y. Yan. 2006. Amplified fragment length polymorphism mapping of quantitative trait loci for malaria parasite susceptibility in the yellow fever mosquito *Aedes aegypti*. Genetics 173:1337–1345.

Zhou, Q., C. G. Wu, B. Dong, F. Q. Liu, and J. H. Xiang. 2008. The encysted dormant embryo proteome of *Artemia sinica*. Marine Biotechnology 10:438–446.

Zhou, X., F. M. Oi, and M. E. Scharf. 2006. Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. Proceeding of National Academy of Sciences 103:4499–4504.

Zhu, G. P., G. B. Golding, and A. M. Dean. 2005. The selective cause of an ancient adaptation. Science 307:1279–1282.

Zhu, Y., T. J. Johnson, A. A. Myers, and M. R. Kanost. 2003. Identification by subtractive suppression hybridization of bacteria-induced genes expressed in *Manduca sexta* fat body. Insect Biochemistry and Molecular Biology 33:541–559.

BIOGRAPHICAL INFORMATION

Diana L. Huestis graduated from Eckerd College with High Honors in 2003 with a B.S. in Marine Science, a concentration in Biology, and a minor in Chemistry. As an undergraduate, Diana studied life-history evolution and population structure of freshwater turtles and salamanders of the genus *Eurycea*. Post-graduation, she immediately began pursuing her M.S. degree in Biology from the University of Texas – Arlington in 2003, and completed this degree in 2005. Her M.S. research examined 1) a GxE interaction between *Idh*-1 genotype and temperature on fitness in the cricket *Allonemobius socius* and 2) environmental and maternal effects on diapause in *A. socius*. This work allowed her to apply for and be awarded a US EPA STAR GRO doctoral fellowship to conduct the research presented in this dissertation. Upon successful completion of her Ph.D., Diana plans to pursue post-doctoral studies and will continue her research on the genetic basis of adaptive life-history traits and ecological genetics. Dr. Huestis plans to eventually obtain a faculty position where she will conduct research and teach at the college/university level.