1-1-2015

Genomic, Evolutionary and Functional Analyses of Diapause in Drosophila Melanogaster

Xiaqing Zhao
University of Pennsylvania, xiaqing@sas.upenn.edu

Genomic, Evolutionary and Functional Analyses of Diapause in Drosophila Melanogaster

Abstract
Understanding the genetic basis of adaptation has been and remains to be one major goal of ecological and evolutionary genetics. The variation in diapause propensity in the model organism Drosophila melanogaster represents different life-history strategies underlying adaptation to regular and widespread environmental heterogeneity, thus provides an ideal model to study the genetic control of ecologically important complex phenotype. This work employs global genomic and transcriptomic approaches to identify genetic polymorphisms co-segregating with diapause propensity, as well as genes that are differentially regulated at the transcriptional level as a function of the diapause phenotype. I show that genetic polymorphisms co-segregating with diapause propensity are found throughout all major chromosomes, demonstrating that diapause is a multi-genic trait. I show that diapause in D. melanogaster is an actively regulated phenotype at the transcriptional level, suggesting that diapause is not a simple physiological or reproductive quiescence. I also demonstrate that genetic polymorphisms co-segregating with diapause propensity, as well as genes differentially expressed as a function of diapause are enriched for clinally varying and seasonal oscillating SNPs, supporting the hypothesis that natural variation in diapause propensity underlies adaptation to spatially and temporally varying selective pressures. In addition to global genomic and transcriptomic screens, I also performed functional analysis of one candidate polymorphism on the gene Crystalllin, which represents an intersection of multiple global screens related to seasonal adaptation. I show that this polymorphism affects patterns of gene expression and a subset of fitness-related phenotypes including diapause, in an environment-specific manner. Taken together, this work provide a holistic view of the genetic basis of a complex trait underlying climatic adaptation in wild populations of D. melanogaster, linking genetic polymorphism, gene regulation, organismal phenotype, population dynamics and environmental parameters.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biology

First Advisor
Paul S. Schmidt

Keywords
diapause, Drosophila melanogaster, genome wide association study, seasonal adaptation, transcriptomic profile

Subject Categories
Biology | Evolution | Genetics

GENOMIC, EVOLUTIONARY AND FUNCTIONAL ANALYSES OF DIAPAUSE IN *DROSOPHILA MELANOGASTER*

Xiaqing Zhao

A DISSERTATION

in

Biology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2015

Supervisor of Dissertation:	Graduate Group Chairperson:
Paul S. Schmidt	Michael A. Lampson
Associate Professor of Biology	Associate Professor of Biology

Dissertation Committee:

Walter F. Eanes, Professor of Ecology and Evolution, Stony Brook University
Timothy A. Linksvayer, Assistant Professor of Biology, University of Pennsylvania
Joshua B. Plotkin, Professor of Biology, University of Pennsylvania
R. Scott Poethig, Professor of Biology, University of Pennsylvania
Paul S. Schmidt, Associate Professor of Biology, University of Pennsylvania
Paul D. Sniegowski, Professor of Biology, University of Pennsylvania

GENOMIC, EVOLUTIONARY AND FUNCTIONAL ANALYSES OF DIAPAUSE IN *DROSOPHILA MELANOGASTER*

ACKNOWLEDGEMENT

I would like to thank my thesis advisor, Paul Schmidt, for his intellectual contributions to my research that made this dissertation possible, for his enthusiasm in a broad range of scientific topics that inspired me to think and do more, and for his good-natured humor that made the past six year incredibly fun. I would also like to thank Alan Bergland and Dmitri Petrov for their generous intellectual input and technical support. I thank the members of my thesis committee, Joshua Plotkin, Paul Sniegowski, Scott Poethig, Tim Linksvayer and Walter Eanes, for their service and advices that improved the quality of my thesis work. Special thanks to Josh, who shared his views on both science and scientific careers at the time when they were the most needed. Peter Petraitis has offered enormous help in statistics throughout my graduate career. Mia Levine provided insightful perspective that improved the work. Yang Jiang and Yue Hou were extremely patient and helpful in getting me started with programming. Many colleagues in the Department of Biology who have generously shared their expertise and resources to make some of my work possible: Qi Zheng, Brian Gregory, Ozlem Onder, Dustin Brisson, Leeanne McGurk, Nancy Bonini, Changsong Yang and Tatyana Svitkina. I would like to extend my appreciation to members of the *Drosophil*a research community, whose work I have and have not directly cited, that made my work possible.

I would like to thank all the other members of the Schmidt lab, Emily Behrman, Katherine O'Brien, Vinayak Mathur, Subhash Rajpurohit, Annalise Paaby and Megan Phifer-Rixey, for their academic and social support. All the other colleagues in Leidy and Lynch that I had the fortune to overlap with will be remembered with affection as well.

This work is supported by the Dissertation Research Fellowship and the Dissertation Completion Fellowship provided by the Graduate Division of School of Arts and Sciences, University of Pennsylvania.

ABSTRACT

GENOMIC, EVOLUTIONARY AND FUNCTIONAL ANALYSES OF DIAPAUSE IN *DROSOPHILA MELANOGASTER*

Xiaqing Zhao

Paul S. Schmidt

Understanding the genetic basis of adaptation has been and remains to be one major goal of ecological and evolutionary genetics. The variation in diapause propensity in the model organism *Drosophila melanogaster* represents different life-history strategies underlying adaptation to regular and widespread environmental heterogeneity, thus provides an ideal model to study the genetic control of ecologically important complex phenotype. This work employs global genomic and transcriptomic approaches to identify genetic polymorphisms co-segregating with diapause propensity, as well as genes that are differentially regulated at the transcriptional level as a function of the diapause phenotype. I show that genetic polymorphisms co-segregating with diapause propensity are found throughout all major chromosomes, demonstrating that diapause is a multi-genic trait. I show that diapause in *D. melanogaster* is an actively regulated phenotype at the transcriptional level, suggesting that diapause is not a simple physiological or reproductive quiescence. I also demonstrate that genetic polymorphisms co-segregating with diapause propensity, as well as genes differentially expressed as a function of diapause are enriched for clinally varying and seasonal oscillating SNPs,

supporting the hypothesis that natural variation in diapause propensity underlies adaptation to spatially and temporally varying selective pressures. In addition to global genomic and transcriptomic screens, I also performed functional analysis of one candidate polymorphism on the gene *Crystalllin*, which represents an intersection of multiple global screens related to seasonal adaptation. I show that this polymorphism affects patterns of gene expression and a subset of fitness-related phenotypes including diapause, in an environment-specific manner. Taken together, this work provide a holistic view of the genetic basis of a complex trait underlying climatic adaptation in wild populations of *D. melanogaster*, linking genetic polymorphism, gene regulation, organismal phenotype, population dynamics and environmental parameters.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

General Introduction

Genetic basis of adaptation

Understanding the genetic basis of adaptation has been and remains to be one major goal of ecological and evolutionary genetics. Various types of adaptation have been observed. Novel environments, including derived habitats, altered biotic factors and human changes to natural environments, pose positive selection on variants that are rare in the original populations but yield higher fitness in the new environments, so that they reach high frequencies in the new habitats. Regular and widespread environmental heterogeneity, such as latitudinal and altitudinal environmental gradients, and daily, lunar or annual cycles of environmental parameters, may result in the allele that is favored in one environment to become disfavored in another, thus maintaining phenotypic and genetic heterogeneity. Irregular and unpredictable environmental heterogeneity such as aseasonal exigencies are largely overcome by adaptations whose primary function is survival in the adverse environments.

Empirically, genetic analyses of fitness-related phenotypes between differentiated populations have revealed major loci under selection. For example, the rock pocket mice (*Chaetodipus intermedius*) living in the deserts of the American southwest have an ancestral light coat color, keeping the mice hidden from their visual predators. Later, a series of volcanic eruptions have formed black lava flows through the pocket mice territory, and the derived melanic mice has risen to high frequency on the dark-colored lava (Nachman *et al.* 2003). Association studies with candidate genes showed that four

non-synonymous substitutions on the gene *Mc1R* were responsible for the dark fur in one lava-dwelling population (Nachman *et al.* 2003). Strong correlation between *Mc1R* allele frequency and habitat color, as well as the absence of correlation between alleles of neutral genes and habitat suggested that natural selection works to match the coat color to habitat color, despite high level of migration between light and dark populations (Hoekstra *et al.* 2004). Similarly, in most humans, the ability to digest lactose declines rapidly after weaning, while some individuals, particularly those descended from populations that have traditionally practiced cattle domestication, maintain lactase activity into adulthood (Swallow 2003; Tishkoff *et al.* 2007). A linkage disequilibrium and haplotype analysis of European families showed that variations in *cis*-acting elements of *LCT*, the gene encoding lactase, are responsible for the variation in lactose persistence, through regulating the transcription level of the gene (Enattah *et al.* 2002). Genotype-phenotype association study of the African populations also identified variants in the regulatory elements of *LCT* to be associated with the ability to digest lactose, yet they arose independently from the European alleles, providing an example of convergent evolution due to selective pressures associated with cattle domestication and adult milk consumption (Tishkoff *et al.* 2007).

The above-mentioned cases represent examples of adaptive changes on genes with large effects. In nature, adaptive traits with Mendelian inheritance may be exceptions rather than the norm (Orr 2005). Many fitness-related phenotypes, especially those underlie adaptation to environmental heterogeneity are quantitative and have highly complex genetic architecture (Rockman 2012; Hendry 2013). For example, flowering

time in *Arabidopsis thaliana*, which is quantitative and varies among genotypes, is a major determinant of fitness in seasonal habitat, as the correct timing of transition from vegetative to reproductive development is critical to synchronize the plant development with environmental conditions suitable for growth or reproduction (Engelmann & Purugganan 2006). Quantitative trait locus (QTL) mapping and association studies have identified dozens of loci that harbor natural polymorphisms that alter flowering time (Engelmann & Purugganan 2006; Ehrenreich *et al.* 2009), with a few loci of moderate to large effect and a large number of loci with smaller effects. The effects of these loci may be non-additive and environment-specific: epistatic effects among candidate genes were detected (Caicedo *et al.* 2004), and genotype-by-environment interactions of flowering time were also observed (Weinig *et al.* 2002).

For complex traits like flowering time, a comprehensive profile that link genetic polymorphism, phenotypic variation, environmental selective pressures and organismal fitness remains a challenge to evolutionary biologists. Population genomic studies screen the genomes of populations from distinct environments to identify outlier loci segregating with environments, which are potentially linked to loci under selection (Stinchcombe & Hoekstra 2008). QTL mapping, genome wide association studies (GWAS) as well as transcriptional and proteomic profiles of fitness-related phenotypes can identify loci and genes underlying ecologically relevant phenotypic variation (Stapley *et al.* 2010). Having identified genes and polymorphisms that show evidence of being involved in adaptation, it would be desirable to functionally characterize the phenotype and fitness consequences of these loci, to bring ecological genetic studies a full-circle.

Wild populations of Drosophila melanogaster and adaptation to environmental heterogeneity

The model organism *Drosophila melanogaster* originated in tropical areas of Africa, and has spread into temperate regions on multiple continents over a historical time period (David & Capy 1988). The worldwide expansion from tropics has required adaptation to the highly heterogeneous environments over broad geographical ranges, as well as adaptation to pronounced summer/winter seasonality in temperate habitats. As such, *D. melanogaster* provides an excellent system to study the evolutionary responses to regular and widespread environmental heterogeneity.

A number of latitudinal clines have been described in *D. melanogaster*, both at the phenotypic level (Capy *et al.* 1993; James *et al.* 1997; Karan *et al.* 1998; Azevedo *et al.* 1998; Robinson *et al.* 2000; Schmidt *et al.* 2005a) and genetic level (Berry & Kreitman 1993; Verrelli & Eanes 2001; Frydenberg *et al.* 2003; Sezgin *et al.* 2004; McKechnie *et al.* 2010). The direct link between clinal genetic markers and fitness-related phenotypes (Paaby *et al.* 2010; Lee *et al.* 2013; Paaby *et al.* 2014), as well as repeated clinal variation at multiple continents (Turner *et al.* 2008; Reinhardt *et al.* 2014) suggest that despite demographic processes (Bergland *et al.* 2015; Kao *et al.* 2015), local adaptation to spatially varying selective pressures has contributed to the observed phenotypic and genetic clines in *D. melanogaster*.

Repeated and predictable changes in phenotype (Schmidt & Conde 2006) and allele frequency (Cogni *et al.* 2013; 2014) across seasons have also been observed in *D.*

melanogaster. Genomic screen has identified hundreds of SNPs that oscillate predictably in frequency over multiple years (Bergland *et al.* 2014), indicating the widespread operation of balancing selection over seasonal and annual timescales.

Despite pervasive evidence of adaptation to spatial and temporal environmental heterogeneity in temperate populations of *D. melanogster*, it remains unclear what traits are directly under selection to generate the observed latitudinal clines and seasonal oscillations in phenotypes and allele frequencies.

The variation of diapause propensity in Drosophila melanogaster

The adult ovarian diapause in *D. melanogaster* may be one phenotype underlying adaptation to the latitudinal and seasonal gradients in climatic parameters and associated selection regimes. Diapause is induced by reduced temperature and shortened photoperiod, and results in arrested ovary development, life span extension, negligible senescence, and elevated stress tolerance (Saunders *et al.* 1989; Saunders & Gilbert 1990; Tatar & Yin 2001; Schmidt *et al.* 2005a; b). The propensity of diapause expression has genetic-basis and is highly variable within the species: some genotypes become reproductively quiescent when exposed to diapausing-inducing conditions while others proceed with vitellogenesis and reproductive development despite of the stressful environment (Williams & Sokolowski 1993). In wild populations, the propensity to enter reproductive diapause exhibits a strong latitudinal cline, with greater than 90% incidence in temperate areas and ~30% in neotropical habitats (Schmidt *et al.* 2005a). Diapause incidence also varies with season: with high propensity to enter diapause (~90%) in

spring and lower incidence (40%-50%) in fall (Schmidt & Conde 2006). Furthermore, diapause incidence is shown to be a powerful predictor of a suite of other traits associated with organismal fitness: even without exposure to diapause-inducing conditions, genotypes that show a high diapause rate are constitutively longer-lived, less fecund, and more stress resistant than the lines that show a low diapause incidence (Schmidt *et al.* 2005b; 2008). These correlated traits have also been shown to be clinal (e.g. (Hoffmann *et al.* 2002; Schmidt *et al.* 2005b) as well as seasonal (Behrman *et al.* 2015). The physiological consequences of diapause induction in *D. melanogaster*, the clinal and seasonal variation in diapause incidence, and the genetic correlations between diapause propensity and other fitness-related traits strongly support that variation in the tendency of expressing diapause is associated with climatic adaptation in temperate habitats: individuals with high diapause propensity have higher survival rate in high latitude and in the harsh winter, while individuals with low diapause propensity have higher fitness in low latitude and summer due to the genetic correlation with high fecundity.

While diapause has been extensively characterized at the physiological level, the genetic and molecular mechanisms that perceive the environmental change to induce diapause expression and stimulate the associated phenotypic cascade are largely unknown (Denlinger 2002). Given the available genetic tools, diapause in *D. melanogaster* is an ideal system for outlining molecular mechanisms of ecologically relevant phenotype and adaptive evolution.

Three candidate diapause genes, *couch potato* (*cpo*) (Schmidt *et al*. 2008), *timeless* (*tim*) (Tauber *et al.* 2007) and *Dp110* (Williams *et al.* 2006) were previously

shown to harbor variants associated with differential diapause propensity. These genes were identified either by a combination of QTL mapping and complementation test, or based on *a priori* knowledge of candidate pathways. They do not explain all of the phenotypic variance, and their effects may be specific to certain genetic background. A relatively unbiased and complete picture of genes involved in the regulation of diapause is lacking. Also, it is not clear whether candidate diapause genes show patterns of climatic adaptation.

Overview of the thesis

This thesis represents a comprehensive effort to investigate the genetic basis of an adaptive trait. With both "bottom-up" and "top-down" approaches, genetic polymorphisms that co-segregate with the natural variation of diapause propensity, as well as genes and transcripts that are differentially regulated as a function of the diapause phenotype were identified; and the functional significance of a polymorphism on candidate diapause gene was evaluated.

Chapter 1 is a global transcriptional profiling of naturally occurring variation in diapause expression in two tissue types associated with neuroendocrine signaling, heads and ovaries. I show that diapause in *D. melanogaster* is an actively regulated phenotype at the transcriptional level, suggesting that diapause is not a simple physiological or reproductive quiescence. Differentially expressed genes and pathways are highly distinct in heads and ovaries, suggesting that the diapause response may be comprised of functional modules associated with specific tissues. A subset of differentially expressed

genes are significantly enriched for clinally varying SNPs and seasonally oscillating SNPs, consistent with the hypothesis that diapause is a driving phenotype of climatic adaptation. Taken together, these results demonstrate that diapause is a complex phenotype actively regulated in multiple tissues, and support the hypothesis that natural variation in diapause propensity underlies adaptation to spatially and temporally varying selective pressures.

Chapter 2 is a functional characterization of a natural polymorphism on a candidate diapause gene *Crystallin* (*Cry*). *Cry* represents an intersection of a transcriptomic profile of diapause, a proteomic profile of diapause and a genome-wide screen of SNPs with allele frequencies oscillating with season. By phenotypically profiling populations that are fixed for either allele of the seasonal SNP on *Cry*, I show that this polymorphism affects patterns of gene expression, and that it affects a subset of fitness-related phenotypes, including diapause. The effects of the polymorphism on gene expression and organismal phenotypes are dependent on temperature and photoperiod, which are the hallmarks of seasonality. This work provides a rare incidence connecting population genomics screens, global expression profiles of fitness-related traits, functional validation of candidate genes, and ecologically relevant environmental parameters.

Chapter 3 is a genome-wide association study and global transcriptional profile of diapause in the same populations. I show that genetic polymorphisms co-segregating with diapause propensity are found throughout all major chromosomes, demonstrating that diapause is a complex trait regulated by multiple genes. The genetic polymorphisms co-

segregating with diapause propensity show signatures of clinal and seasonal adaptation. By combing the association study and transcriptional profile, I also show evidence that a subset of diapause SNPs may affect the phenotype by altering expression levels of candidate genes.

Overall, the three chapters provide a holistic view of the genetic basis of a complex trait underlying climatic adaptation in wild populations of *D. melanogaster*, linking genetic polymorphism, gene regulation, organismal phenotype, population dynamics and environmental parameters.

Chapter One

Global transcriptional profiling of diapause and climatic adaptation in *Drosophila melanogaster*

Introduction

Understanding the mechanisms by which populations adapt to distinct environments has remained a major goal of evolutionary and ecological genetics. A number of environmental parameters vary predictably with latitude, and the examination of natural populations across latitudinal gradients has been widely used to elucidate fundamental aspects of spatially variable selection (Blanckenhorn & Fairbairn 1995; Clapham *et al.* 1998; Huey *et al.* 2000; Loeschcke *et al.* 2000; García Gil *et al.* 2003; Stinchcombe *et al.* 2004; Zhang *et al.* 2008). Similarly, seasonality creates predictable temporal variation in a suite of environmental parameters, and changes in natural populations for various fitness-related traits among seasons demonstrate evolutionary responses to seasonal shifts in selective pressures (Hairston & Dillon 1990; Grant & Grant 2002; Korves *et al.* 2007; Brown *et al.* 2013; Tarwater & Beissinger 2013).

The model organism *Drosophila melanogaster* originated in tropical areas of Africa and have subsequently spread into temperate regions on multiple continents (David & Capy 1988). Natural populations now experience highly heterogeneous environments over broad geographical ranges, thus providing an excellent system to study the evolutionary responses to environmental variability. A number of latitudinal clines have been described in *D.melanogaster*, both at the phenotypic (Capy *et al.* 1993;

James *et al.* 1997; Karan *et al.* 1998; Azevedo *et al.* 1998; Robinson *et al.* 2000; Schmidt *et al.* 2005a) and genetic levels (Berry & Kreitman 1993; Verrelli & Eanes 2001; Frydenberg *et al.* 2003; Sezgin *et al.* 2004; McKechnie *et al.* 2010); clinal patterns have been documented on most continents, including Australia (Mitrovski & Hoffmann 2001; Hoffmann *et al.* 2002; Frydenberg *et al.* 2003; Umina *et al.* 2005; Paaby *et al.* 2010), the Indian subcontinent (Rajpurohit *et al.* 2008a; b; Rajpurohit & Nedved 2013), North America (Schmidt *et al.* 2005a; b; Paaby *et al.* 2010), South America (Folguera *et al.* 2008; Goenaga *et al.* 2013) and Africa (Fabian *et al.* 2015). The direct link between clinal genetic markers and fitness-related phenotypes (Paaby *et al.* 2010; Lee *et al.* 2013; Paaby *et al.* 2014), as well as repeated clinal variation at multiple continents (Turner *et al.* 2008; Reinhardt *et al.* 2014) suggest that despite demographic processes (Bergland *et al.* 2015; Kao *et al.* 2015), local adaptation to spatially varying selective pressures has contributed to the observed phenotypic and genetic clines in *D. melanogaster*. Repeated and predictable changes in phenotype (Schmidt & Conde 2006; Behrman *et al.* 2015) and allele frequency (Cogni *et al.* 2013; 2014) over seasons have also been found in *D. melanogaster*. Genomic screen have identified hundreds of SNPs that oscillate predictably in frequency over multiple years (Bergland *et al.* 2014), indicating the widespread operation of balancing selection over seasonal and annual timescales as well as across spatial gradients.

The selective regimes associated with high and low latitudes are in many ways similar to those associated with winter and summer seasonality: both high latitude and winter seasons are associated with lower temperature and scarce resources, while both

low latitude and summer seasons are associated with elevated temperature and abundant resources. A natural question is whether the observed latitudinal clines and seasonal oscillations are generated by the same selection pressures, and if so, what are the common underlying mechanisms of adaptation.

The adult ovarian diapause in *D. melanogaster* may be one phenotype underlying adaptation to the cyclical seasonal and geographic gradients in climatic parameters and associated selection regimes. Diapause is induced by shortened photoperiod and moderately low temperature, and results in reproductive quiescence, extreme life span extension (more than 5-fold), negligible senescence, increased lipid storage and increased stress tolerance (Saunders *et al.* 1989; Saunders & Gilbert 1990; Tatar *et al.* 2001a; Schmidt *et al.* 2005a; b). Notably, diapause expression is highly variable within and among *D. melanogaster* populations, and the variation has a significant genetic component: when exposed to the diapause-inducing conditions, some genotypes tend to become reproductively quiescent whereas others tend to proceed with vitellogenesis and reproductive development (Williams & Sokolowski 1993). The propensity to enter reproductive diapause exhibits a strong latitudinal cline in eastern North American populations: greater than 90% in temperate areas and ~30% in neotropical habitats (Schmidt *et al.* 2005a). Diapause incidence also varies with season: with high propensity to enter diapause (~90%) in spring and lower incidence (40%-50%) in fall (Schmidt & Conde 2006). Similarly, genetic polymorphism in candidate diapause genes also varies clinally (Fabian *et al.* 2012; Bergland *et al.* 2014) and seasonally (Cogni *et al.* 2013). Furthermore, diapause incidence is shown to be a powerful predictor of a suite of other

traits associated with organismal fitness: even without exposure to diapause-inducing conditions, genotypes that show a high diapause rate are constitutively longer-lived, less fecund, and more stress resistant than the lines that show a low diapause incidence (Schmidt *et al.* 2005b; 2008). These correlated traits have also been shown to be clinal (e.g. (Hoffmann *et al.* 2002; Schmidt *et al.* 2005b) as well as seasonal (Behrman *et al.* 2015).

The clinal and seasonal patterns of diapause incidence and its association with other life-history traits may represent a fundamental trade-off between somatic maintenance and reproduction: under stressful and unfavorable conditions, hardiness is favored although associated with lower fecundity; while under benign conditions, high reproductive rate is favored at the cost of aspects of survival. It remains unclear what traits are directly under selection to generate the observed latitudinal clines and seasonal oscillations in phenotype and allele frequencies. Given the physiological consequences of diapause induction in *D. melanogaster*, the clinal and seasonal variation in diapause incidence, and the genetic correlations between diapause propensity and other fitness-related traits, we hypothesize that diapause is one major determinant of adaptation to spatial and temporal environmental heterogeneity in this taxon. Therefore genes differentially regulated as a function of the diapause phenotype are likely under spatially and/or temporally varying selectively pressure, thus genetic polymorphisms on these genes are likely to show clinal and seasonal patterns.

Unlike many other arthropod systems in which diapause is constitutive and invariant within populations under stressful environments (Tauber *et al.* 1986; Denlinger

2002; Ragland *et al.* 2010; Poelchau *et al.* 2013a), the intra-specific variation of diapause propensity in *D. melanogaster* provides an excellent opportunity to identify the genes and pathways associated with the observed genetic variance for diapause expression, and to identify the downstream targets that may ultimately result in the observed phenotypes, via comparisons between diapausing and non-diapausing genotypes exposed to identical environmental parameters. Three candidate genes for diapause in *D. melanogaster* have been previously identified: *Dp110* (Williams *et al.* 2006), *timeless* (*tim*) (Tauber *et al.* 2007), *couch potato* (*cpo*) (Schmidt *et al.*, 2008). Similarly, two major pathways have been implicated in diapause expression: the circadian clock (Tauber *et al.* 2007; Pegoraro *et al.* 2014) and insulin-insulin like growth factor signaling that plays a central role in the neuroendocrine regulation (Tatar *et al.* 2001b; Williams *et al.* 2006; Schmidt 2011). However, a comprehensive genomic and transcriptomic analysis of diapause in natural populations of *Drosophila* is lacking (but see (Baker & Russell 2009), which utilizes diapause as a way to synchronize developmental stages to elucidate fundamental aspects of egg development).

Maternal photoperiod has been shown to influence whether offspring express diapause in other taxa (Henrich & Denlinger 1982; Saunders 1987; Rockey *et al.* 1989; Oku *et al.* 2003), and epigenetic processes are known to contribute to the control of diapause (Reynolds *et al.* 2013; Yocum *et al.* 2015). Accordingly, we hypothesized that higher-level regulatory mechanisms such as local sharing of regulatory elements and chromatin structure may also be involved in the regulation of diapause in *D.*

melanogaster, and seek to test if this is reflected in the differential expression profile of diapausing and non-diapausing individuals.

Here, we performed high-coverage, genome-wide transcriptomic profiling of heads and ovaries, two tissue types associated with neuroendocrine signaling, in diapausing and non-diapausing individuals. To our knowledge, this is the first comprehensive expression profile of diapause in an organism that has naturally different perception of the same environmental cues. We use these data to address several fundamental questions: (1) What are the genes and transcripts that are differentially expressed as a function of diapause phenotype? (2) Are the genes previously identified as being associated with variance in diapause propensity differentially regulated in diapausing and non-diapausing genotypes? (3) Is there evidence of chromatin-based gene regulation associated with the diapause phenotype? (4) Do the genes that are differentially expressed as a function of diapause also segregate for SNPs that vary predictably in frequency with latitude and season? (5) Are the observed patterns parallel between heads and ovaries or are they tissue-specific? Our data demonstrate that diapause is an actively regulated phenotype at the transcriptional level, suggesting that it is not a simple dormancy characterized by physiological quiescence. All genes previously shown to affect diapause have differentially expressed isoform(s) in at least one of the two tissues, confirming their association with the phenotype. Patterns of differential expression are spatially clustered across the major chromosomes, suggesting that chromatin-level regulation may contribute to the regulation and response of diapause. Genes down-regulated in heads of diapausing flies are significantly enriched for clinal

and seasonal SNPs, consistent with the hypothesis that diapause is a driving phenotype of climatic adaptation. Taken together, these results support the idea that diapause is a fundamental and complex phenotype that involves multiple levels of regulation, and underlies adaptation to spatially and temporally varying selective pressures.

Materials and Methods

Fly populations

D. melanogaster were collected as isofemale lines by direct aspiration on wind fallen fruit from four natural populations of northern regions of the east coast of the US: Rocky Ridge Orchard in Bowdoin, ME (44.03°N, 69.95°W), Champlain Orchards in Shoreham, VT (43.86°N, 73.35°W), Westward Orchards in Harvard, MA (42.49°N, 71.56°W) and Lyman Orchards in Middlefield, CT (41.50°N, 72.71°W). All populations were sampled in October of 2009. Fifty isofemale lines from each orchard were used to construct an outbred population. Ten females and ten males from a single age cohort for each isofemale line were released into laboratory population cages and were allowed to interbreed for six generations. The outbred cage was constructed less than six months after the flies were collected from the wild. Four cages were created from the F1 of the initial cohort, and the population was maintained as four experimental replicates.

Diapause assay, cDNA library construction and sequencing

Vials containing standard cornmeal-molasses medium were placed in the population cage and flies were allowed to lay eggs on the surface for approximately 4 hours. These eggs were reared at low density at 25°C, 12 hour Light: 12 hour Dark. Females were collected with 2 hours post-eclosion and placed at 11°C at a photoperiod of 9h light: 15h dark. After being exposed to the diapause-inducing conditions of low temperature and short days for four weeks, all experimental females were dissected and the developmental status of the ovaries was assessed (King 1970). A female was scored as diapausing (D) if the most advanced oocyte was arrested before vitellogenesis (before stage 8); a female was scored as nondiapausing (ND) if vitellogenin was observed in either ovary (stage 8 or later).

The heads and ovaries of either diapausing or nondiapausing females were pooled and separately flash-frozen in liquid nitrogen, yielding 4 samples: D_head, ND_head, D_ovary and ND_ovary. A total of 92 flies went into the pooled diapause samples (D_head and D_ovary), and a total of 106 flies went into the pooled nondiapause samples (ND_head and ND_ovary). cDNA libraries were prepared as previously described (Elliott *et al.* 2013), and sequenced on the Illumina Hiseq 2000 platform using standard protocols for 100bp single-end read sequencing.

Read processing, alignment, counts estimation and differential expression

As some of the fragments in the libraries were shorter than 100bp, the 3' adapters were sequenced in part or full in some of the reads. Therefore raw reads were first

trimmed using the cutadapt program (Martin 2011). Trimmed reads were mapped to the *D. melanogaster* reference transcriptome and expression levels of genes and isoforms were estimated following the RNA-Seq by Expectation-Maximization (RSEM) pipeline version 1.2.8 (Li & Dewey 2011). The reference transcriptome was constructed with the *D. melanogaster* reference genome version 5.54 and the Ensembl annotation version 5.74. The mean and standard deviation of read length in each library was calculated using fastqstats of the ea-utils toolset (Aronesty 2013), and provided to the RSEM pipeline.

The empirical Bayesian approach-based tool EBSeq (Version 1.5.3) (Leng *et al.* 2013) was implemented to call for genes and isoforms that have differential expression levels between D_head and ND_head, D_ovary and ND_ovary. The reads per kilobase per million reads (RPKM) values given by RSEM were used as input. Library sizes were normalized using the median normalization approach provided by EBSeq. Genes and isoforms that have adjusted *p*-values (FDR) smaller than 0.05 were considered differentially expressed (DE) as a function of the diapause phenotype. Due to the similarities of the isoforms, it is difficult to assign short reads unambiguously to individual isoforms (Roberts & Pachter 2011). Therefore we confine our examination of the isoform-level expression profiles to a general analysis, and do not make further set-based inferences.

Since the most obvious phenotypic difference between D and ND is the development and maturation of the ovary, and genes involved in vitellogenesis and oogenesis are at much higher expression levels in developed ovaries compared to development-arrested ovaries (Baker & Russell 2009), it is possible that the drastic

difference of these genes in D and ND ovaries alone can affect the detection of other differentially expressed genes. To address this issue, we excluded the genes under the Gene Ontology term "Vitellogenesis" (GO:0007296) and "Oogenesis" (GO: 0048477) and re-ran the differential expression analysis on the ovary data. The estimation of the expression levels for the other genes is not affected by the removal of vitellogenesis and oogenesis genes, and the list of DE genes is not affected. Therefore, these genes were kept in the analysis.

To estimate the inversion frequencies in the samples, we called SNPs from the RNAseq data using the SAMtools mpileup version 1.1 (Li *et al.* 2009), and assessed the frequencies of major cosmopolitan inversions using the diagnostic SNPs provided by Kapun et al. (Kapun *et al.* 2014).

Functional category and pathway enrichment

Gene ontology (GO) and Kyoto encyclopedia of gene and genomes (KEGG) pathway enrichment analysis was performed using the GOseq package (Young *et al.* 2010) that accounts for transcript length bias associated with RNA-seq data. Only genes that had RPKM values greater than one in at least one of the D or ND samples in the respective tissues were used as the background gene list for that tissue. Genes that were of significantly high and low expression in diapausing head/ovary samples were separately tested to identify which categories of genes are enhanced and suppressed as a function of diapause. The *p* values were corrected using the Benjamini and Hochberg

procedure (Benjamini & Hochberg 1995), and the false discovery rate threshold was set to be 0.05.

Location-dependent expression

To investigate whether genes located closely on the chromosome are similarly regulated between D and ND, we examined the log2 fold changes of expression levels in D vs. ND along each major chromosome arm, and calculated the levels of spatial autocorrelation of the log2 fold changes using Moran's I (Moran 1950). Null distributions of spatial autocorrelation for each chromosome arm were generated by randomly scrambling the locations of genes within the chromosome. Tandemly-arrayed duplicated genes were removed from the analysis due to the difficulty of unambiguously mapping reads to these genes, as well as the potential for local co-expression. The list of tandemly-arrayed duplicated genes was obtained from Quijano et al. 2008 (Quijano *et al.* 2008). Moran's I calculation was implemented with Moran.I function of the R package "ape" (Paradis *et al.* 2004) version 3.1.

Seasonal and clinal overlap with differentially expressed genes

Seasonal and clinal genes were identified based on the results of Bergland et al. 2014 (Bergland *et al.* 2014), where generalized linear models with binomial error structure were used to call SNPs that oscillate with season or change with latitude. SNPs with a seasonal q value of 0.3 or less were considered to be seasonal, and those with a

clinal *q* value of 0.01 or less were considered to be clinal (as per Bergland et al. 2014). Genes with at least one seasonal/clinal SNP on them, or within 5Kb upstream or downstream of an identified SNP were considered seasonal/clinal genes respectively.

We then tested if differentially expressed genes are enriched in the seasonal/clinal set. Because DE genes and seasonal/clinal SNPs are spatially clustered, to eliminate the possibility that the signal of enrichment is driven by a few clusters of DE genes that happen to harbor clusters of seasonal/clinal SNPs, we implemented a block-bootstrap procedure: we generated 500 subsets of DE genes where at most 1 DE gene was sampled from each 100 Kb consecutive interval of the genome. For each DE gene, a gene that was not differentially expressed, but of similar length and normalized mean RPKM value, was established as a control. 500 sets of control genes were generated for each DE list to ensure the reproducibility of the results. The numbers of genes that are seasonal/clinal were numerated in both the DE and control gene sets, and odds ratios of DE genes being seasonal/clinal relative to the control genes are calculated. Estimates of the mean $\log_2$ odds ratios and the 95% confidence intervals were taken across the 500 block bootstraps. Up- and down-regulated genes in either tissue type were then tested for enrichment separately.

Results

Differentially expressed genes

Our data demonstrate that diapause is actively regulated at the transcriptional level in both heads and ovaries in adult *Drosophila melanogaster* (Figure 1). The

numbers of genes having relatively higher and lower expression levels in diapausing individuals are approximately the same in the both head and ovary, indicating that diapause is not a simple dampening of biological processes, but represents an alternative physiological state associated with the concomitant up-regulation and down-regulation of many genes.

A large number of genes are differentially expressed (DE) between diapausing (D) and non-diapausing (ND) individuals and this is true both in head and ovary (Table 1). Because *D. melanogaster* has only been in temperate environments for a relatively short period of time (David & Capy 1988), one hypothesis is that as a potentially recent adaptation to the novel temperate environment, diapause in this species might only involve a small number of genes. However, our results suggest that as in other insects (Ragland *et al.* 2010; Poelchau *et al.* 2013a; b), the diapause phenotype in *D. melanogaster* is associated with differential expression of hundreds of genes.

The expression profiles of D and ND are highly distinct between head and ovary, and differentially expressed genes in the two tissues are not the same (Supplementary Table 1). The RPKM fold change of D over ND in the head is not correlated to that in the ovary (*p* value of Pearson's product-moment correlation test=0.9817, *cor*=-0.000222). Only 165 genes are differentially expressed in both head and ovary, of which 41 show opposite regulation in the two tissues; by chance alone, we would expect to see ~120 genes overlapping in the DE lists in the two tissues. The inconsistency of transcriptional regulation in head and ovary as a function of the diapause phenotype advocates analysis of different tissue and body parts in the dissection of mechanisms of diapause.

Because diapause incidence is clinal (Schmidt *et al.* 2005a), and distributions of some of the cosmopolitan inversions are also clinal (Knibb 1982; Kapun *et al.* 2014), one hypothesis is that the expression of diapause is associated with specific inversion status. We assessed the frequencies of the seven major cosmopolitan inversions (*In(2L)t*, *In(2R)Ns*, *In(3L)P*, *In(3R)C*, *In(3R)K*, *In(3R)Mo*, *In(3R)Payne*), and found that *In(2R)Ns*, *In(3R)C*, *In(3R)K*, *In(3R)Mo* are at low frequencies in our experimental population. The frequencies of *In(2L)t*, *In(2R)Ns*, *In(3R)C*, *In(3R)K*, *In(3R)Mo* are not different between diapausing and non-diapausing samples (data not shown). The regions containing the inversion-specifc SNPs of *In(3L)P* and *In(3R)Payne* are not transcribed so we are not able to assess their frequencies from our data. However, according to previous results (Kapun *et al.* 2014), they are at low frequencies in the northern populations of North America. In addition, the differentially expressed genes are found across all major chromosomes and are not clustered around inversion break points. Based on these results, we conclude that inversion status is not driving the variation of diapause propensity, and does not have pronounced effects on diapause-associated patterns of gene expression in northern populations of North America.

Differentially expressed isoforms

To fully characterize the differences in transcriptional profiles of diapausing and non-diapausing *Drosophila*, we examined patterns of transcription at the gene level incorporating all isoforms. As a complementary analysis, we also examined differential expression of each individual isoform. This is especially interesting in the expression

profiles of diapause, as candidate diapause genes may affect the phenotype in transcript-specific manners (Tauber *et al.* 2007).

Our results show that there are a large number of isoforms differentially expressed between D and ND in both the head and ovary. Compared to the gene level DE detection, the isoform level tests identified many more genes with at least 1 differentially expressed isoform (Table 2). As expected, the majority of genes (more than 60%) identified from the gene-level tests have at least 1 differentially expressed isoform (Figure 2). However, there are also some genes that do not have any significantly DE isoforms, but when all isoforms are considered together, the genes are differentially expressed. There are also hundreds of genes carrying isoforms that have opposite patterns of regulation in D and ND (Table 2), and these antagonistic effects may cancel out and make the gene not DE in the gene-level tests. Our results clearly demonstrate that the examination of gene-level DE associated with diapause fails to capture fundamental aspects of the transcriptional complexity.

Functional categories of differentially expressed genes

To place the comparisons in gene expression into a biologically meaningful context, we identified enriched functional categories using gene ontology (GO) enrichment analysis as well as the Kyoto encyclopedia of genes and genomes (KEGG) pathways analysis. To distinguish between functional categories of genes that are up- or down-regulated as a function of the diapause phenotype, we performed the enrichment

tests separately for each class of genes in both heads and ovaries. We see non-overlapping GO terms (Supplementary Table 2) and KEGG pathways (Table 3) in different tissue types for up- and down-regulated genes.

Several metabolic pathways are down-regulated in diapause heads according to KEGG enrichment, including carbohydrate (starch, sucrose, galactose) as well as amino acid (glycine, serine threonine, phenylalanine) metabolism. This indicates that flies in diapause may generate less energy. There is also a down-regulation of "fatty acid degradation" and "other glycan degradation" in diapause heads, suggesting a shift towards storage in diapausing flies. This is consistent with physiological characterizations (Tauber *et al.* 1986) as well as transcriptional profiling of diapause in other taxa (Ragland *et al.* 2010; Poelchau *et al.* 2013a).

Candidate genes and pathways

Although none of the previously identified candidate genes (*cpo*, *tim* and *Dp110*) show differential expression at the gene level (Table 4), all of them have at least one differentially expressed isoforms in at least one tissue (Table 5). Two isoforms of *cpo* show opposite regulation in the ovaries: the *cpo*-RO transcript is up-regulated in diapausing ovaries, whereas the *cpo*-RS transcript is down-regulated in this same tissue. The two isoforms of *tim*, *tim*-RL and *tim*-RM, are differentially regulated in opposite directions in both heads and ovaries, and the regulation of each isoform shows antagonistic directions in different parts of the body. This suggests that the effects of *tim*

on diapause are transcript and tissue specific. Taken together, genome-wide transcriptional analysis of naturally occurring variation in diapause induction demonstrates that the three previously known candidate genes underlying diapause propensity are differentially expressed; furthermore, the data is of sufficient resolution to detect that it is specific isoforms that are associated with the observed phenotypic variation.

The insulin insulin-like growth factor signaling pathway (IIS) regulates dauer formation in *C. elegans* (Kimura *et al.* 1997; Apfeld & Kenyon 1998) and dauer formation is in many ways similar with diapause in insects (Tatar & Yin 2001; Sim & Denlinger 2013). Manipulations of different components of the IIS pathway in *D.melanogaster* phenocopy many aspects of diapause such as arrested cell cycle (LaFever & Drummond-Barbosa 2005; Hsu *et al.* 2008), extended lifespan (Clancy *et al.* 2001; Tatar *et al.* 2001b; Hwangbo *et al.* 2004; Giannakou *et al.* 2004; Lee *et al.* 2008; Grönke *et al.* 2010) and increased fat storage (Böhni *et al.* 1999). Therefore, the expression of genes in the insulin-signaling pathway in diapausing and non-diapausing flies is of particular interest.

We examined all genes belonging to the GO category "insulin receptor signaling pathway" (GO: 0008286) to determine if any differential regulation is present in either heads or ovaries as a function of the diapause phenotype. None of these genes are differentially expressed in the heads. However, multiple components of the signaling pathway exhibit differential expression in the ovaries: the insulin-like peptide *Ilp-4* is up-

regulated, and *Ilp7* and *Ilp8* are down-regulated, in diapausing flies (Supplmentary Table 4). Since insulin-like peptides are known to control the germ cell cycle (LaFever & Drummond-Barbosa 2005), this result suggests that these peptides are related to the ovarian developmental arrest. As with the evaluation of the candidate genes for diapause, an examination of specific transcripts demonstrates more complex patterns of expression for genes of this pathway in both heads and ovaries. The genes *dock*, *foxo*, *InR*, *Pdk1*, *Pten* and *RpL8* have differentially expressed transcripts in the heads; the genes *chico*, *dock*, *melt*, *Pdk1*, *Pi3K21B* and *Pi3K92E* (*Dp110*) have differentially expressed transcripts in the ovaries (Supplementary Table 4). Of these genes, *chico* affects body size and lipid storage (Böhni *et al.* 1999); the expression of *chico* and *foxo* are related to lifespan (Clancy *et al.* 2001; Hwangbo *et al.* 2004; Giannakou *et al.* 2004); natural variants of *InR* are associated with various aspects of life-history evolution (Paaby *et al.* 2014); and *Pi3K92E* is a candidate diapause gene. Our data suggest that the IIS genes identified as differentially expressed in our transcriptional profile are excellent candidates for the regulation of diapause, and that their effects may be transcript-specific.

Chromosome location-dependent expression

To test if there is any higher-order regulation of expression in diapause, we examined the spatial autocorrelation of log2 fold changes of expression levels in D vs. ND along all major chromosomes. The values of Moran's I, which is a measure of the direction and level of spatial autocorrelation, are located on the far right side of the null

distributions generated by randomizing the locations of genes within the chromosome; the sole exception is head data on chromosome X (Figure 3). The pattern of spatial autocorrelation is not driven by local co-expression of tandemly arrayed duplicated genes, as they were excluded from the analysis. This result suggests that chromosome location-based co-regulation of gene expression is present in the transcriptional regulation of diapause.

Clinal and seasonal overlap

To test the hypothesis that genes differentially regulated in diapausing and non-diapausing *Drosophila* are likely to be under spatially and/or temporally varying selective pressure, we identified the overlaps of differentially expressed genes with clinal and seasonal genes according to Bergland (Bergland *et al.* 2014). We found hundreds of differentially expressed genes harbor clinally varying polymorphisms, and dozens of differentially expressed genes harbor seasonally varying polymorphisms (summarized in Table 6, listed in Supplementary Table 5).

There is a significant enrichment of clinal genes in genes that are down-regulated in diapausing heads (p=0.022) (Figure 4). The 357 clinal genes that are down-regulated in diapausing heads are primarily located throughout chromosomes 2 and 3, and the locations of these genes are not associated with the cosmopolitan inversion breakpoints. These genes are enriched for GO terms “proteolysis”, “oxidation-reduction process”, and

KEGG “metabolic pathways”. The other classes of differentially expressed genes are not enriched for clinal polymorphisms.

Similarly, we also observe a significant enrichment of seasonal genes in genes that are down-regulated in diapausing heads (p=0.01), but not in other classes of differentially expressed genes (Figure 4). As with the clinal overlap gene list, the 98 seasonal genes that are down-regulated in diapausing heads are primarily located on chromosomes 2 and 3, and the locations of these genes are not associated with the cosmopolitan inversions. Unlike the clinal overlap gene list, however, no gene ontology or pathway enrichment was identified for this set of genes. Many of these genes' functions are completely unknown either from experimental evidence or from prediction based on sequence similarities. Interestingly, 29 out of the 98 down-regulated genes have seasonal SNP(s) in vicinity of the known *cis*-eQTL peaks of these genes (Supplementary Table 5) (*cis*-eQTL regions contributing to regulatory variation in *D.melanogaster* female head were obtained from (King *et al.* 2014). This suggests that these seasonally fluctuating SNPs, or sequence variants linked to them, may be under seasonally oscillating selection because they are associated with different gene expression levels that are associated with the diapause and non-diapause phenotypes. These results prompt future research on function of these candidate genes, as well as the interplay between segregating variation, expression level, phenotype and dynamics of selection.

Discussion

Diapause is the most intensively studied adaptation to seasonality in arthropods, and is a complex physiological syndrome that in adults involves sequestered reproduction, extended lifespan, altered metabolism and increased stress tolerance. Global transcriptional profiles of diapause have been performed on whole bodies of the flesh fly *Sarcophaga crassipalpis* (Ragland *et al.* 2010) and the Asian tiger mosquito *Aedes albopictus* (Poelchau *et al.* 2013a; b). These studies identified metabolic pathway transitions, cell cycle arrest and altered endocrine regulation as fundamental features of the diapause program. Both species have constitutive diapause, therefore the difference between the diapause and non-diapause phenotypic states were inferred from comparisons between insects exposed to different environments as well as insects at different diapause stages. In *D.melanogaster*, transcriptomic profiling of female abdomens has identified lipid metabolism, insulin signaling and structural constituents of chitin-based cuticle being altered by the transition from diapause-inducing conditions to diapause-terminating conditions (Baker & Russell 2009).

Here we provide the first global transcriptomic profiling of naturally occurring variation in diapause expression that is elicited under the exact same environmental parameters. Like the aforementioned studies, we find that diapause is an actively regulated transcriptional state where almost as many genes are up-regulated as down-regulated. We have also identified that aspects of metabolism are altered as a function of diapause. Components of the IIS pathway were identified as differentially expressed in our study, consistent with previous investigations. Unlike the aforementioned studies, our

analysis focused on how differential effective perception of low temperature and a short photoperiod affected patterns of transcription in the adult head and ovary, tissues/structures associated with juvenile hormone and ecdysteroid signaling rather than systemic metabolism (e.g., fat body, flight muscle, digestive system). While both the adult head and ovary may be involved in both environmental perception and the downstream transcriptional regulation of this perception (Liu *et al.* 1988; Plautz *et al.* 1997), diapause expression in *Drosophila* is regulated by measuring the length of night (Saunders 2013; Meuti & Denlinger 2013) and may involve an inhibition of junvenile hormone release in the corpora allata (Saunders & Gilbert 1990; Yamamoto *et al.* 2013). Further investigations have also implicated ecdysteroids in the regulation of diapause in this taxon (Richard *et al.* 1998). Thus, transcriptional profiles of flies differentially responding to the same environmental parameters eliminate confounding factors such as diet, age, developmental state, temperature and photoperiod. Confining the transcriptional profile in heads and ovaries reduces the influence of systemic metabolic processes in other parts of the body. This type of global transcriptional screen may reveal genes and pathways differentially responding to specific environmental cues, as well as those under distinct neuroendocrine control.

Biological processes revealed by GO and KEGG enrichment tests

Among genes that are up-regulated in diapausing head compared to non-diapausing head, GO terms such as “ribosome biogenesis”, “ribonucleoprotein complex

biogenesis", "rRNA processing" and "rRNA metabolic process" are among the most significantly enriched biological processes. Ribosome biogenesis is a major metabolic activity and accounts for large proportion of energy consumption in the cell (Wullschleger *et al.* 2006). Therefore, at first glance it is paradoxical that genes related to ribosome biogenesis are up-regulated in diapausing heads, where energy consumption is presumably reduced. However, in the flesh fly *S. crassipalpis*, the ribosomal protein P0 was found to be up-regulated in diapausing pupae brains (Flannagan *et al.* 1998), and the up-regulation of P0 is tightly linked to the down-regulation of O_2 consumption (Craig & Denlinger 2000). Ribosome biogenesis proteins were thought to contribute to the cellular maintenance of the insect during the diapause period (Denlinger 2002). Although the homologous gene in *D. melanogaster*, *RpLP0*, is not significantly up-regulated in diapause heads, the up-regulation of these functional categories suggests that low metabolic rate and similar response to hypoxia occurs in diapausing heads in *D. melanogaster*.

Vitellogenesis and egg-formation related terms are enriched in genes that are down-regulated in diapausing head. Although vitellogenesis and egg formation obviously does not occur in the head, these genes are likely pleiotropic and may have important functions outside the ovary. In fact, the expression levels of the yolk proteins *Yp1*, *Yp2* and *Yp3* are dozens of times higher in heads than in ovaries, and experimental evidence has shown that these genes are involved in neurogenesis (Neumüller *et al.* 2011). The consistency of the shut down of some vitellogensis and egg-formation related genes in

diapausing heads and ovaries indicates some shared mechanisms of gene regulation in different parts of the body.

The phototransduction pathway is up-regulated in diapausing ovaries. Phototransduction genes are expected to play a role in the regulation of diapause as they are involved in light sensing, and the perceived change in photoperiod is an anticipatory environmental cue used in both the initiation and termination of diapause (Denlinger 2002; Pegoraro *et al.* 2014). The conundrum is why the up regulation of phototransduction occurs in the ovary instead of head. It was previously shown that circadian genes are expressed in multiple tissues (Liu *et al.* 1988) and circadian oscillators are present throughout the body of *D. melanogaster* (Plautz *et al.* 1997). It may be that the ovaries sense light directly, and that increased light sensing in diapausing ovaries affects the expression of diapause.

Cell cycle arrest in the ovaries

GO terms such as "chromosome organization", "nucleosome assembly", "protein-DNA complex assembly" and "chromatin assembly" are highly over represented in genes that are down-regulated in diapausing ovaries. A closer examination reveals that these GO enrichments are driven by the shut down of a group of ~90 histones located on the right end of chromosome 2L from cytogenetic location 39D3 to 39E1. Canonical histones are greatly up-regulated during S phase of the cell cycle for the assembly of newly replicated chromatin (Osley 1991), but the two genes H3.3A and H3.3B that encode the

variant H3.3 are expressed at a relatively constant level throughout the cell cycle (Ahmad & Henikoff 2002). In the ovary data, there is a bulk loss of canonical histone transcripts in diapausing individuals but H3.3 is not differentially expressed between diapausing and non-diapausing individuals: this suggests arrested cell division (somatic and germline) during diapause. In fact, GO terms such as "cell cycle", "cell division", "meiotic cell cycle" and "mitotic cell cycle" are also enriched in genes that are down-regulated in diapausing ovaries.

However, many of these histone encoding genes are characterized by a high degree of sequence similarity, and thus it is possible that the mapping software is not able to accurately differentiate between them. If true, assigned expression values may have been generated for similar histones by dividing reads equally among them; thus, the observed differential expression of these functional classes may be driven by a subset of the associated genes. Despite that, the observed data for head and ovary are still quite distinct: the canonical histones are shut down in diapausing ovaries but not in diapausing heads, suggesting that the cell cycle arrest is specific to the ovaries.

Chromosome-location dependent expression regulation

Neighboring genes in the *D. melanogaster* genome can exhibit similar expression patterns across different experimental conditions, and such similarity is not explained by gene function or homology (Spellman & Rubin 2002; Boutanaev *et al.* 2002; Kalmykova *et al.* 2005; Levine *et al.* 2010). Although the mechanism of this neighborhood effect is

not clear, we speculated that it can be explained by regulation at the level of chromatin structure, where chromosome regions containing genes open for translational regulation increase the accessibility of the *cis*-regulatory elements of their neighboring genes. Alternatively, local sharing of *cis*- and *trans*- regulatory elements could activate the target genes as well as its adjacent genes (Oliver *et al.* 2002; Michalak 2008).

We have also observed a high level of spatial autocorrelation of transcriptional regulation as a function of the diapause phenotype. Whether the co-regulation of genes involved is constitutive or specific to the diapause phenotype is beyond the scope of our data and analyses. However, we can make two speculations: 1) higher order regulation of gene expression, such as regulatory element sharing and chromatin remodeling, is present in diapause; 2) since the transcriptional regulation is not always perfectly precise to individual genes, it is possible that some genes are differentially expressed as a result of "transcriptional hitch hiking", where the neighboring genes are the targets of transcriptional regulation (Oliver *et al.* 2002). Regardless, the spatial autocorrelation of transcriptional profiles does not conflict with our previous conclusion that diapause involves the differential expression of many genes, as DE genes are found across all chromosomes.

Seasonal and clinal overlap

Evidence suggests that the ancestral source of *D. melanogaster* is the subtropical Miombo and Mopane woodland of Zambia and Zimbabwe, where the genetic diversity is

the highest (Pool *et al.* 2012). Since there is wet-dry seasonality in this region, it is possible that *D. melanogaster* has an ancestral seasonal syndrome or response that is associated with the wet/dry seasonality. Migration out of subtropical Africa to the high latitude temperate regions in North America, Europe, Australia and the Indian subcontinent with pronounced winters has resulted in the exposure of *D. melanogaster* to novel environmental conditions, as well as selective pressures, associated with temperate climates. The physiological and genetic mechanism of adaptation to temperate climate is not clear but increased diapause propensity is likely involved, and this may in part be achieved by selection on gene expression levels.

Our global transcriptional profiling of diapause has suggested that, as an adaptation to seasonality, diapause in *D. melanogaster* is a globally regulated syndrome. The genes that are differentially expressed as a function of the naturally occurring variation in diapause phenotype represent a non-random set, as genes that are down-regulated in diapausing heads are more likely than expected by chance to show both seasonal and clinal patterns of adaptation. Contrary to the traditional hypothesis that genes that are shut down during diapause may simply represent a suspension of development or reduced metabolic activity (Denlinger 2002), our data demonstrate that it is the transcriptionally suppressed rather than transcriptionally activated genes that exhibit a robust signal of climatic adaptation. We suggest that subsequent functional analysis of this set of genes could provide further insight into the mechanisms of climatic adaptation in *Drosophila* and other taxa.

Tables

Table 1. Summary of the gene-level tests of differential expression

	Head	Ovary
Candidate genes tested for DE*	12204	10856
Differentially expressed genes between D and ND (FDR<0.05)	1094	1173
Up-regulated genes in D (FDR<0.05)	510	426
Down-regulated genes in D (FDR<0.05)	584	747

*To eliminate noise and improve model fitting, only genes with RPKM>10 in either D or ND sample are tested for differential expression.

Table 2. Summary of the isoform-level tests of differential expression

	Head	Ovary
Candidate isoforms tested for DE[1]	21087	18956
DE isoforms between D and ND (FDR<0.05)	3762 (2608)[1]	3603 (2491)
Up-regulated isoforms in D (FDR<0.05)	1598 (1384)	1784 (1407)
Down-regulated isoforms in D (FDR<0.05)	2164 (1705)	1819 (1522)
Genes not DE at gene level, but have DE isoforms	1939	1641
Genes with both up- and down-regulated isoforms	481	438

1. To eliminate noise and improve model fitting, only isoforms with RPKM>10 in either D or ND sample are tested for differential expression.

2. Number in parenthesis is the number of genes corresponding to the isoforms listed before.

Table 3. Enriched KEGG pathways

Tissue	Pathway	FDR	No. DE in category	Total in category
Head	Up-regulated in D			
	ribosome biogenesis in eukaryotes	5.37E-27	32	71
	RNA polymerase	2.27E-07	9	24
	ribosome	4.65E-05	12	81
	pyrimidine metabolism	5.01E-05	12	73
	Down-regulated in D			
	metabolic pathways	8.98E-09	74	876
	glycine, serine and threonine metabolism	2.47E-07	10	27
	lysosome	4.69E-05	13	78
	other glycan degradation	7.32E-05	6	19
	fatty acid degradation	3.82E-04	7	28
	one carbon pool by folate	1.21E-03	4	11
	starch and sucrose metabolism	1.39E-03	9	54
	phenylalanine metabolism	1.51E-03	4	10
	galactose metabolism	2.03E-03	6	26
	glycerolipid metabolism	3.61E-03	7	41
Ovary	Up-regulated in D			
	Phototransduction	3.34E-07	7	25
	Down-regulated in D			
	No enrichment			

Table 4. Gene-level expression of candidate diapause genes (expression levels are normalized and in units of RPKM)

	Head				Ovary			
Gene	D	ND	adj.*p*	DE	D	ND	adj.*p*	DE
cpo	46708	41759	0.996	FALSE	2674	1567	0.377	FALSE
tim	87867	8842	0.997	FALSE	5731	3611	0.714	FALSE
Dp110	1987	2119	0.991	FALSE	4820	3855	0.974	FALSE

Table 5. Transcript-level expression of candidate diapause genes (expression levels are normalized and in units of RPKM.)

		Head				Ovary			
Gene	Transcript	D	ND	adj.*p*	DE	D	ND	adj.*p*	DE
cpo	*cpo*-RN	167	238	0.498	FALSE	367	202	0.785	FALSE
	cpo-RO	5559	3835	0.594	FALSE	370	0	0.000	TRUE
	cpo-RP	6125	5429	0.992	FALSE	596	332	0.815	FALSE
	cpo-RQ	378	295	0.925	FALSE	18	5	0.566	FALSE
	cpo-RR	Low	Low	NA	NA	Low	Low	NA	NA
	cpo-RS	2091	3038	0.081	FALSE	0	198	0.000	TRUE
	cpo-RT	12130	10747	0.995	FALSE	112	67	0.865	FALSE
	cpo-RU	Low	Low	NA	NA	Low	Low	NA	NA
	cpo-RV	21426	17070	0.987	FALSE	1382	669	0.579	FALSE
	cpo-RW	96	0	0.000	TRUE	Low	Low	NA	NA
	cpo-RX	Low	Low	NA	NA	Low	Low	NA	NA
	cpo-RY	Low	Low	NA	NA	Low	Low	NA	NA
tim	*tim*-RB	Low	Low	NA	NA	Low	Low	NA	NA
	tim-RL	22	0	0.000	TRUE	0	35	0.000	TRUE
	tim-RM	351	548	0.027	TRUE	149	38	0.002	TRUE
	tim-RN	0	32	0.000	TRUE	31	7	0.162	FALSE
	tim-RO	625	483	0.912	FALSE	1224	827	0.951	FALSE
	tim-RP	Low	Low	NA	NA	Low	Low	NA	NA
	tim-RR	8026	7547	0.995	FALSE	4693	2487	0.841	FALSE
	tim-RS	Low	Low	NA	NA	Low	Low	NA	NA
Dp110	*Pi3K92E*-RA	1193	1539	0.693	FALSE	3835	3623	0.981	FALSE
	Pi3K92E-RB	848	525	0.058	FALSE	1293	0	0.000	TRUE
	Pi3K92E-RC	Low	Low	NA	NA	Low	Low	NA	NA

Table 6. Summary of the overlap of differentially expressed genes and clinal/seasonal genes

	Head	Ovary
Clinal genes tested for differential expression*	7468	6565
Clinal genes up-regulated in diapause	289	279
Clinal genes down-regulated in diapause	357	395
Seasonal genes tested for differential expression*	1689	1469
Seasonal genes up-regulated in diapause	54	87
Seasonal genes down-regulated in diapause	98	75

* To eliminate noise and improve model fitting, only genes with RPKM>10 in either D or ND sample are tested for differential expression.

Figures

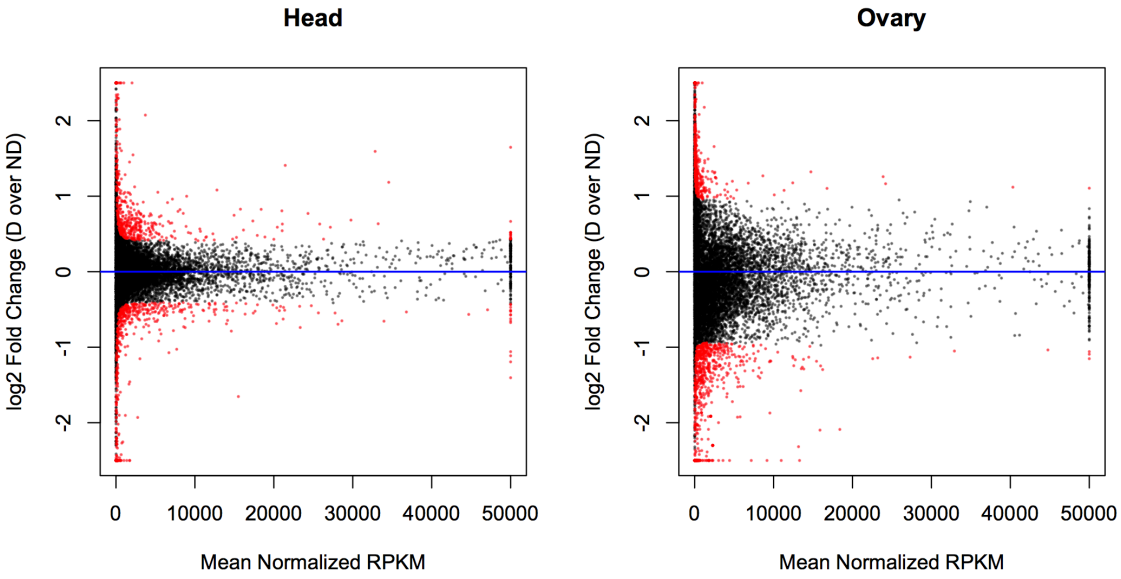

Figure 1. Log2 fold change of expression levels versus mean normalized RPKM values in head and ovary.

Genes that are significantly differentially expressed (FDR<0.05) in D and ND are shown in red. The expression levels are normalized by library sizes. Genes with relatively higher expression levels in diapausing samples as compared to nondiapausing samples have positive log2 fold change values, and genes with lower expression levels in D have negative log2 fold change values. In the head, 6237 genes fall above the line of y=0 and 5967 genes fall below the line. In the ovary, 5272 genes fall above the line and 5584 genes fall below.

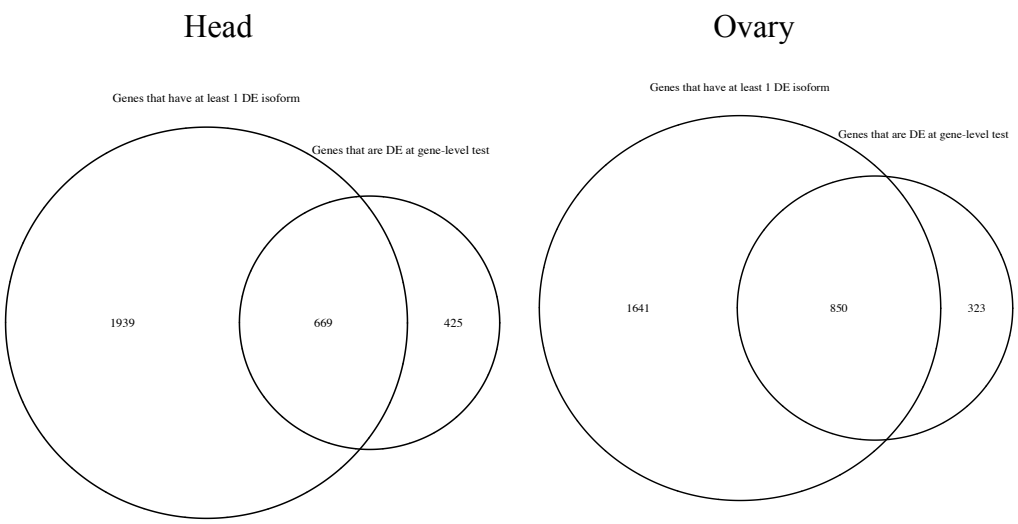

Figure 2. Venn diagrams of numbers of genes identified from the gene-level and isoform-level differential expression tests.

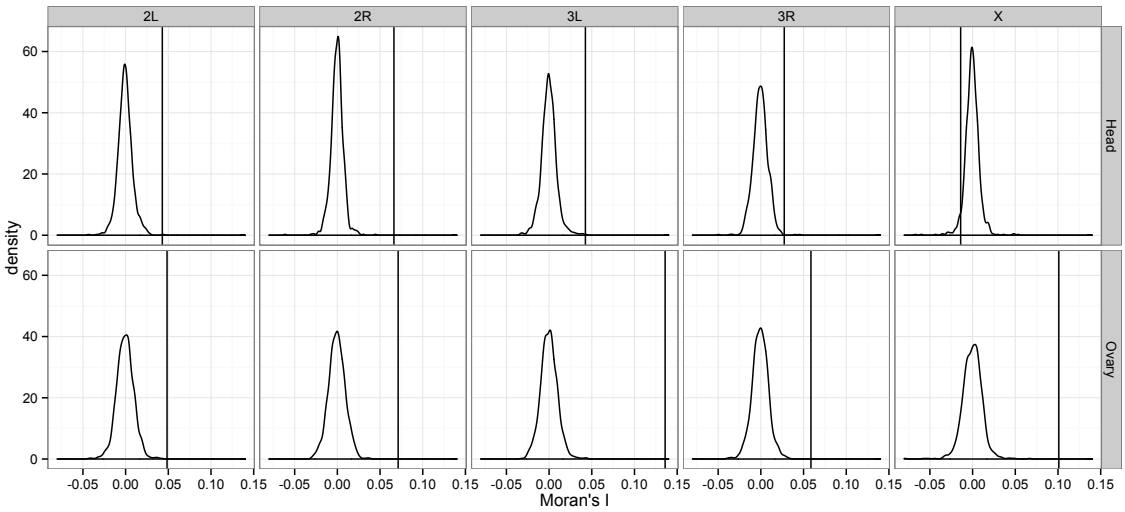

Figure 3. Spatial clustering of gene expression regulation.

The null distributions and real (shown as vertical lines) Moran's I of log2 fold change of D over ND on the major chromosome arms in head and ovary. A positive Moran's I indicates positive spatial autocorrelation. The null distribution was generated by scrambling the starting sites of genes within each chromosome 1000 times.

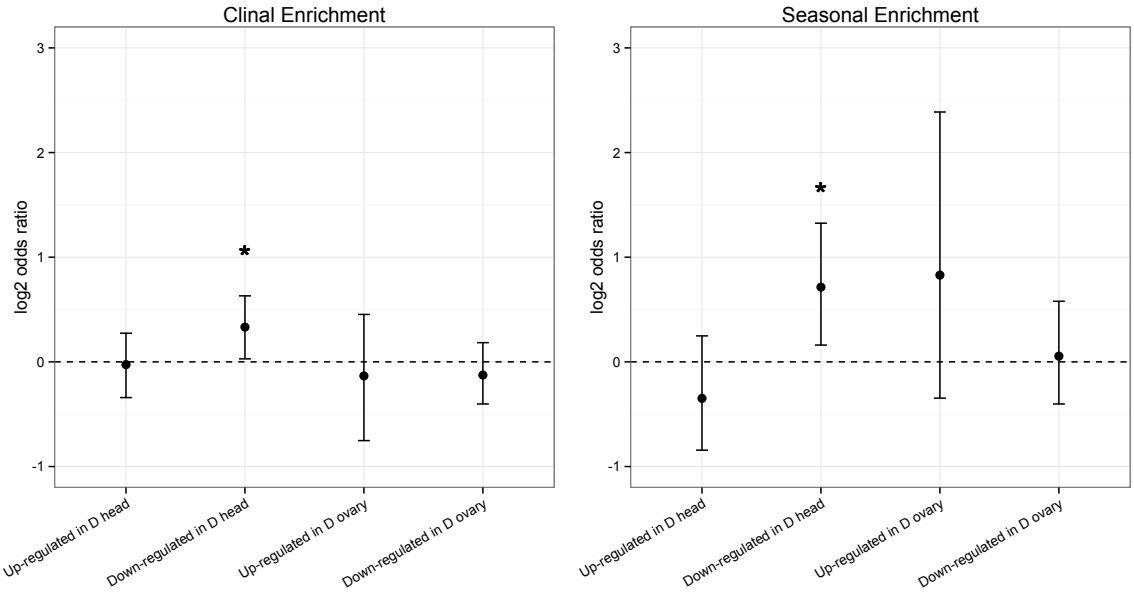

Figure 4. Clinal and seasonal enrichment of differentially expressed genes.

Enrichment ($\log_2$ odds ratio) of differentially expressed genes among genes that harbor, or are near clinally-varying (left) and seasonally-oscillating (right) polymorphisms. Error bars represent 95% confidence intervals based on 500 block bootstrap resampling.

Chapter Two

Natural variants of the gene *Crystallin* are associated with differential seasonal adaptation in *Drosophila melanogaster*

Introduction

Understanding the genetic basis of adaptation remains one of the most fundamental goals in modern evolutionary biology. To achieve this goal, several approaches have been developed. Population genomics screens the genomes of populations from distinct environments to identify outlier loci segregating with environments, which are potentially linked to loci under selection (Stinchcombe & Hoekstra 2008). Quantitative trait locus (QTL) mapping, genome wide association studies (GWAS) as well as transcriptomic and proteomic profiles of fitness-related phenotypes identify genes underlying ecologically relevant phenotypic variation, depicting the genotype-phenotype map of adaptive traits (Stapley *et al.* 2010). To comprehensively understand how individual polymorphisms segregating in populations are put together to affect phenotypes and fitness that eventually shape the allele frequencies in natural populations, one step forward is to functionally assay genes and polymorphisms identified from both population genomics screens and QTL/transcriptomic/proteomic analysis of fitness-related traits (Stinchcombe & Hoekstra 2008; Rockman 2012).

Adaptation to environmental heterogeneity has long been used as a context to elucidate aspects of the genetics of adaptation. A number of environmental parameters

vary predictably with latitude, and natural populations arrayed along broad latitudinal gradients often show clines in phenotype and genotype (Capy *et al.* 1993; Berry & Kreitman 1993; Blanckenhorn & Fairbairn 1995; James *et al.* 1997; Clapham *et al.* 1998; Karan *et al.* 1998; Azevedo *et al.* 1998; Huey *et al.* 2000; Robinson *et al.* 2000; Loeschcke *et al.* 2000; Verrelli & Eanes 2001; García Gil *et al.* 2003; Frydenberg *et al.* 2003; Stinchcombe *et al.* 2004; Sezgin *et al.* 2004; Schmidt *et al.* 2005a; Zhang *et al.* 2008; McKechnie *et al.* 2010; Fabian *et al.* 2012; Bergland *et al.* 2015). The parallel latitudinal clines observed across multiple continents (Turner *et al.* 2008; Reinhardt *et al.* 2014), as well as direct links between clinally varying phenotypes and genetic polymorphisms (Paaby *et al.* 2010; 2014) demonstrate that local adaptation to spatially varying selective forces plays an important role in shaping the population differentiation. Despite all the efforts in elucidating the spatially varying selective pressures, the mechanistic basis of the observed latitudinal clines are still poorly understood. One potential complication is that, since the latitudinal clines are formed over longer period of time, demographic processes such as population stratification and secondary contact can also shape patterns of clinality, which are difficult to disentangle from local adaptation (Bergland *et al.* 2015; Kao *et al.* 2015).

Temporal environmental heterogeneity can also pose variable selective pressures on organisms. Seasonality as a wide spread temporal environmental variant confers predictable cycling of selective forces over annual time scale (Tauber *et al.* 1986). The traits and alleles associated with high fitness in one season may be subsequently disfavored in another, resulting in the maintenance of genetic and phenotypic

polymorphism (Haldane & Jayakar 1963; Gillespie 1973; Gillespie & Langley 1974). Substantial empirical evidences support the phenotypic response to seasonality (Hairston & Dillon 1990; Grant & Grant 2002; Schmidt & Conde 2006; Korves *et al.* 2007; Brown *et al.* 2013; Tarwater & Beissinger 2013). Although the selective regimes associated with winter and summer seasonality are in many ways similar to those associated with high and low latitudes, unlike latitudinal environmental gradients, seasonal cycling of selective pressures occur over shorter time scales, and in many cases on the same residing populations, thus provides a more tractable context to link the changes in genotype and phenotype, to unravel the molecular variance underlying adaptation in natural environmental heterogeneity. Due to the short time frame of seasonal cycling of selective pressures, we predict that *cis*-regulatory variants, which confer higher level of plasticity as compared to structural variants, are likely to be involved in seasonal adaptation.

Two particular environmental parameters, temperature and photoperiod, vary predictably with season. Not only do they provide reliable cues for organisms to synchronize aspects of their life history with seasonal changes of selective pressures such as humidity and food availability (Tauber *et al.* 1986; Bradshaw & Holzapfel 2007), the variation in temperature and photoperiod also provides alternative selective regimes among successive generations and within same individuals at different points in a calendar year. Temperature and photoperiod are closely associated with the regulation of diapause and dormancy, which are the primary adaptation to seasonality that marks the trade-off between somatic maintenance and reproduction (Tauber *et al.* 1986; Denlinger 2002; Bradshaw & Holzapfel 2007). Therefore, temperature and photoperiod play critical

role in adaptively modifying physiological features, and we expect that the fitness consequences of some of the genetic polymorphisms/phenotypic variations underlying seasonal adaptation may be specific to certain environmental conditions, including temperature and photoperiod.

The model organism *Drosophila melanogaster* show pronounced response to seasonality. A number of genetic polymorphisms cycle with season (Cogni *et al.* 2013; 2014). Genomic screen of populations collected from different times of the years identified hundreds of SNPs across the genome that cycle in frequency consistently and repeatedly between spring and fall. These SNPs are enriched for functional genetic elements, showing signatures of adaptive response to seasonality (Bergland *et al.* 2014). Diapause propensity, which is potentially a driving phenotype of climatic adaptation in *D. melanogaster*, cycle with season (Schmidt & Conde 2006). Genetic polymorphism in candidate diapause gene fluctuate seasonally as well (Cogni *et al.* 2013). Some phenotypes that are known to vary seasonally such as lifespan, fecundity, development time and stress resistance (Behrman *et al.* 2015) are genetically correlated with diapause propensity (Schmidt *et al.* 2005b). These evidences strongly support that the variation in the tendency of expressing diapause in *D. melanogaster* is associated with adaptation to seasonality in temperate habitats: populations with high diapause propensity have higher survival rate in the harsh winter, while populations with low diapause propensity have higher fitness in summer due to the genetic correlation with high fecundity.

To understand the genetic basis of seasonal adaptation in temperate populations of *D. melanogaster*, efforts were made in both population genomics screen and mechanistic

dissection of traits related to seasonality. Individuals of *D. melanogaster* collected in the spring and fall over multiple years were sequenced, and hundreds of SNPs with allele frequencies repeatedly oscillating across seasons were identified as potentially subject to strong seasonal selection (Bergland *et al.* 2014). Global transcriptional profile of genes differentially expressed as a function of the diapause phenotype has identified hundreds of genes showing different expression levels between diapausing and non-diapausing individuals (Zhao *et al.* under review). Global proteomic profile of diapause has identified dozens of proteins showing different abundances as a function of the diapause phenotype (unpublished data). Of the genes and polymorphisms identified as associated with seasonal adaptation and diapause propensity, one polymorphism stood out as an intriguing intersection of multiple screens: the gene *Crystallin* (*Cry*), which encodes a 52kDa glycoprotein, is down-regulated at both transcriptional level and protein level in heads of diapausing individuals (Zhao *et al.*)(Figure 1A, Figure 1B); a SNP on the intronic region of *Cry* fluctuates seasonally in three consecutive years, and the allele frequency becomes more "spring-like" than the fall allele frequency after an acute frost event, further supporting that this SNP or linked loci is under seasonally-fluctuating selection (Figure 1C) (Bergland *et al.* 2014).

In *Drosophila melanogaster*, *Cry* was first identified as the structural constituent of the corneal lens (Komori *et al.* 1992; Janssens & Gehring 1999). The *Drosophila* compound eye consists of approximately 800 ommatidia, and the corneal lens is the outmost structure of each ommatidium. It is an acellular convex lamellated structure that is responsible for focusing light onto the photoreceptors. Three proteins were found in the

D. melanogaster corneal lens and *Cry* is the most abundant. The *Cry* protein is synthesized in cells underlying the lens and then secreted into the lens (Janssens & Gehring 1999). *Cry* protein is located unevenly in the corneal lens, with higher density in the bands of high refractive index (Nilsson 1989; Komori *et al.* 1992). *Cry* is also shown to be an integral constituent of the peritrophic matrix. The peritrophic matrix is a layer composed of chitin, proteins and glycoproteins, that lines the insect intestinal lumen and protects the midgut epithelium from abrasive food particles, pathogens and toxins (Lehane 1997). *Cry* protein is observed in the peritrophic matrix (Kuraishi *et al.* 2011). Flies with loss-of-function mutation of *Cry* have thinner and more permeable peritrophic matrix, which drastically increases the susceptibility of the flies to oral bacterial infection (Kuraishi *et al.* 2011). The functions of *Cry* suggest that it may be involved in organismal fitness in multiple ways: compound eye is one primary organ of light sensing, and the corneal lens may alter the light input into the photoreceptors underneath the lens and cone, thus affect perception of light, and eventually lead to altered photoperiodic responses. The gut epithelium plays vital roles in processes such as nutrient assimilation, immunity and osmoregulation, all of which potentially have fitness consequences (Huang *et al.* 2015).

The seasonal SNP on *Cry* identified in Bergland et al. 2014 is found in the first intron, located in between of the 5' UTR and the coding region (chromosome location is 2L:11944664, nucleotide position from start of gene is 536, site referred to as Cry^{536} thereafter). The relatively more abundant C allele at this site (~70% on average in Pennsylvania populations) increases in frequency over winter in three consecutive years,

showing evidence of adaptation to winter seasonality thus referred to as $Cry^{536:C(W)}$ thereafter. The relatively less abundant T allele, accordingly, is favored over summer, therefore referred to as $Cry^{536:T(S)}$. This polymorphism is old: it is present in the ancestral populations of *D. melanogaster* in Zambia at similar frequency (73% of C) as the Pennsylvania populations used to identify seasonally oscillating SNPs (Lack *et al.* 2015). Furthermore, evidence suggests that this SNP is trans-specific between *D. melanogaster* and *D. simulans*: the T and C alleles, as well as an A allele are found in lines originated from putatively ancestral populations of *D. simulans* (Begun *et al.* 2007), although more recent sequencing efforts did not identify the T allele in *D. simulans* (Rogers *et al.* 2014; Palmieri *et al.* 2015). The long-term preservation of the polymorphism, and the repeated seasonal oscillation of allele frequency may result from distinct selective pressures associated with seasonality. However, it is not clear how alleles of this polymorphism differentially affect phenotypes, and eventually confer distinct fitness consequences of summer and winter. Functional analysis of this SNP may bridge the gaps between genotype, phenotype and fitness.

In this study, we evaluate the functional significance of the seasonal SNP on *Cry* by phenotypically profiling populations that are fixed for either allele of the SNP. We show that this polymorphism is a *cis*-regulatory element that affects patterns of gene expression, and that it affects a subset of fitness-related phenotypes, including diapause. The effects of the polymorphism on gene expression and organismal phenotypes are dependent on temperature and photoperiod, which are the hallmarks of seasonality. This work provides a rare incidence connecting population genomics screens, global

expression profiles of fitness-related traits, functional validation of candidate genes, and ecologically relevant environmental parameters.

Materials and Methods

Population cages construction

To test if the two alleles of the seasonal SNP on *Cry* are functionally distinct in a natural outbred genetic background, we used the *Drosophila* Genome Reference Panel (DGRP) (Mackay *et al.* 2012; Huang *et al.* 2014) lines to construct outbred populations so that one of the alleles of the seasonal SNP is fixed in the population and the rest of the genome is randomized for allele state (including SNP and indel). Since the allelic state of each polymorphism of each line is known, we were able to pool multiple lines that carry the same allele of the target SNP, and allow the flies to freely recombine for multiple generations to get the desired populations. 18 DGRP lines were selected for each of the population cages (the line composition of cages can be found from Supplementary Table 1), and two population cages consisting of independent combinations of lines were created for each allele as biological replicates. For each outbred population, 10 virgin females and 10 males were collected from each of the 18 DGRP lines, and released into a 30×30×30 cm protective viewing chamber. The flies were allowed to interbreed freely and lay eggs on standard medium. The eggs were collected to establish the next generation of the population. All populations were allowed to recombine for eight generations prior to phenotyping. We computationally examined the genotype of all other

polymorphic nucleotide positions of the DGRP lines; assuming all lines contributed to the gene pool, no other locus is alternatively fixed for different alleles between populations. Thus, these experimental populations are fixed only for the SNP of interest. Besides, no other SNP or indel is at frequency higher than 90% in populations fixed for the summer allele, but at frequency lower than 10% in populations fixed for the winter allele, and vice versa.

A SNP in the 3' UTR of *Cry* that is not seasonal and not in disequilibrium with the seasonal SNP (D=0.02) was chosen as the control SNP (2L:11946412) Populations fixed for either allele of the control SNP were constructed using the same methodology as the seasonal SNP. For populations fixed for alleles of the control SNP, we kept the heterozygosity of the site of the seasonal SNP high, and vice versa. Computational examination of other polymorphic nucleotide positions revealed that four other SNPs (2L:11946356, 2L:11946398, 2L:11946433, 2L:11946487) were fixed together with the control SNP. These five SNPs are found within a small range of 132bp and are in significant linkage disequilibrium: they co-segregate in all DGRP lines except the lines where the genotype of one or more of the SNPs is ambiguous.

Because populations homozygous for different alleles of the desired SNP have presumably randomized genetic background and differ in just a small region, we hypothesize that most of the phenotypes would not differ between either the alleles of the seasonal SNP, or the alleles of the control SNP. The traits affected by the allele state of the seasonal polymorphism should differ specifically between alleles of the seasonal SNP, but not between alleles of the control SNP. If both the seasonal and control SNP

have effects on certain traits, we cannot make any inference of the association between the phenotype and the allele state of the seasonal SNP, but attribute it to line effects or cage effects.

Functional tests of the seasonal SNP

A) Phenotypic assays of population cages

For all phenotypic assays, flies were obtained by placing bottles containing standard cornmeal-molasses media in the cages and allow females to lay eggs on the surface for up to 24 hours. These eggs were reared at low density (<40 individuals per vial) at 25°C, 12-hour light: 12-hour dark until eclosion. All phenotypic assays were conducted under 25°C, 12h light: 12h dark unless otherwise specified. For each phenotypic assay, we used mixed-model nested ANOVA (unless otherwise specified) to test for significant effects of genotype as well as other relevant main effects such as sex and interactions. The two replicate population cages are nested within genotype, and since the line compositions of the cages represent just two out of many possible line combinations for each genotype, cage nested within genotype is considered a random effect. The seasonal SNP and control SNP are tested separately, and subsequently compared to identify significant effects specific to the seasonal SNP. All statistics were performed using JMP version 11 (SAS Institute, Cary, NC).

Lifespan and early-life fecundity. Four replications of mixed-sex batches (20 virgin females and 20 males) were collected from each cage and transferred into plastic

bottles. A total of 32 bottles from the four seasonal SNP cages and four control SNP cages were maintained by changing standard media egg-laying plates every day (days 1-14) or every other day (day 15 onward) until all flies were dead. Bottles were inverted so that dead flies were collected on the media plates. Dead flies were scored at every plate change. The numbers of eggs laid were counted from day 2 to day 14 at every plate change. Per capita fecundity for each day was calculated by dividing the total number of eggs found on the plate by the number of live females in the bottle, and total fecundity of the first two weeks was calculated by taking the sum of per capital fecundity from day 2 to day 14. Lifespan was analyzed using a proportional hazards model, with three fixed effects: genotype, sex and genotype×sex interaction. Early-life fecundity was analyzed first on total fecundity of the first two weeks, using a mixed model ANOVA with the fixed effect of genotype and the random effect of cage (nested within genotype); and then on per capita fecundity of each day, using a repeated measures ANOVA, with fixed effects of genotype, days since eclosion, genotype×days since eclosion and random effects of cage (nested within genotype) and bottle (nested within cage).

Chill coma recovery. Freshly eclosed flies were collected over 24 hours from each cage, sorted into eight batches of ~10 males and ~10 females and aged for 4-5 days in standard media vials. To induce chill coma, vials containing the flies were inverted and placed on 0°C ice water mixture for three hours. The knocked-out flies were then transferred back to room temperature into transparent plastic petri dishes. Time to recovery (signified as transition to the upright position) was recorded using a video

camera, and was analyzed using a mixed model ANOVA with fixed effects of genotype, sex, genotype×sex and random effect of cage (nested within genotype).

Starvation resistance. Freshly eclosed flies were collected over 24 hours from each cage, and sorted into four batches of 10 males and 10 females and aged for 4-5 days in standard media vials. To measure starvation resistance, flies were transferred into media-free vials containing a cotton ball saturated with 2mL of water. The number of dead flies in each vial was checked every three hours after 24 hours until all flies died. Data were analyzed using a mixed model ANOVA with fixed effects of genotype, sex, genotype×sex and random effect of cage (nested within genotype).

Desiccation resistance. Freshly eclosed flies were collected over 24 hours from each cage, and sorted into five batches of 10 males and 10 females and aged for 4-5 days in standard media vials. To measure desiccation resistance, flies were transferred into media-free vials topped with dried silica gel and sealed with parafilm. The number of dead flies in each vial was checked every two hours after 8 hours until all flies died. The hours survived was analyzed using a mixed model ANOVA with fixed effects of genotype, sex, genotype×sex and random effect of cage (nested within genotype).

Diapause propensity. Females from each cage were collected within 2 hours post-eclosion and placed at 11°C at either short day photoperiod (9h light: 15h dark) or long day photoperiod (15h light: 9h dark). After being exposed to the diapause-inducing conditions of low temperature for four weeks, all experimental females were dissected and the developmental status of the ovaries was assessed (King 1970). A female was scored as diapausing (D) if the most advanced oocyte was arrested before vitellogenesis

(before stage 8); a female was scored as nondiapausing (ND) if vitellogenin was observed in either ovary (stage 8 or later). Diapause assays were conducted in three batches using females collected from three consecutive generations. The diapause incidence was analyzed using a mixed model ANOVA with fixed effects of genotype, photoperiod, genotype×photoperiod and random effects of cage (nested within genotype) and batch (generation).

B) Quantitative complementation test

As an additional approach to test for the allelic effects of the seasonal SNP on *Cry* in diapause and desiccation, we performed quantitative complementation tests. Line 24348 from the Bloomington Drosophila Stock Center was crossed with both alleles of the seasonal SNP. Line 24348 has the genotype w^{1118};*Df*(*2L*)*BSC323*/*CyO*, where *Df*(*2L*)*BSC323* is an 185Kb deficiency on chromosome 2L that uncovers an array of genes including *Cry*, and *CyO* is the balancer chromosome over which the deficiency is maintained. The crosses yielded four genotypes: $Cry^{536:T(S)}$/deficiency, $Cry^{536:C(W)}$/deficiency, $Cry^{536:T(S)}$/balancer, $Cry^{536:C(W)}$/balancer, where within each genotype of the seasonal SNP, flies carrying deficiency and balancer chromosome can be identified based on the wing phenotype. The F1 progeny of the crosses were used for diapause and desiccation assays as described above.

Diapause incidence was analyzed using a mixed model ANOVA with fixed effects of genotype ($Cry^{536:T(S)}$ vs. $Cry^{536:C(W)}$), chromosome (deficiency vs. balancer),

photoperiod (short day vs. long day) and all the two-way interactions and three-way interaction, as well as the random effect of cage (nested within genotype). A significant genotype×chromosome term indicates failure to complement, or variation in diapause incidence in the sample of the wild alleles. A significant genotype×photoperiod×chromosome term suggests the failure to complement is dependent upon photoperiod. In addition to evidence of failure to complement, variance of the wild alleles in the balancer background should not be significantly greater than that in the deficiency background, since greater variance in the balancer background would suggest that the failure to complement is due to epistatic interaction between wild alleles and genes on the balancer chromosome, rather than allelic variation (Mackay 2001).

Desiccation resistance was analyzed following the same criteria, except that the fixed effects are genotype, chromosome, sex and all possible two-way and three-way interactions.

Quantitative PCR

To test if the transcriptional expression of *Cry* is different between the two alleles of the seasonal SNP under different environmental conditions, we performed quantitative PCR (qPCR) on adult females exposed to different combinations of temperature and photoperiod. Eggs were collected from each population cage and reared at low density (<40 eggs per vial) at 25°C, 12 hour light: 12 hour dark. Within 2 hours after eclosion, females were transferred to one of the four conditions: short day (9h light: 15 h dark) and

low temperature (11°C), long day (15h light: 9 h dark) and low temperature, short day and high temperature (25°C), long day and high temperature. Flies were aged for 7 days on standard media in their respective environment and snap frozen in liquid nitrogen and stored at -80°C. Twenty females were pooled into one sample, and four replicates of pooled samples were prepared for each cage/temperature/photoperiod combination. Total RNA was extracted using the Trizol reagent (Qiagen), treated with TURBO DNase (Ambion), and reverse transcribed using SuperScript III First Strand Synthesis Supermix (Invitrogen) following the manufacturer's protocol. Relative abundance of transcript was determined using SYBR Select Master Mix and an ABI ViiA 7 Real-time PCR system (Applied Biosystems) by the ΔC_T relative quantitation method. *GAPDH* was used as the endogenous control. Three technical replications were used for each sample. Primer sequences for the qPCR reaction are as follows: *Cry* forward, ACGTCAGCAGCCATGAAAC; *Cry* reverse, CACATTGCAGGTGAGGAGAC; *GAPDH* forward, AAAAAGCTCCGGGAAAAGG; *GAPDH* reverse, AATTCCGATCTTCGACATGG.

Corneal lens morphology

To test the hypothesis that natural allelic variation at *Cry* may affect some aspects of morphological phenotypes of the corneal lens, and to draw connections between genotype, expression, diapause and lens morphology, we performed histological analysis of the compound eyes of flies with both alleles of the seasonal SNP and a *Cry* mutant. To test if the effects of *Cry* on lens morphology are environment-sensitive, we exposed flies

of each genotype to different environments. Eggs were collected from each population and reared at low density at 25°C, 12 hour light: 12 hour dark. Within 2 hours after eclosion, flies were transferred to one of the four conditions: short day (9h light: 15 h dark) and low temperature (11°C), long day (15h light: 9 h dark) and low temperature, short day and high temperature (25°C), long day and high temperature. Like the diapause assay and quantitative PCR, only females were used in the analysis. Flies were aged for 7 days on standard media in their respective environment, and fixed with 37% formaldehyde (i.e. Formalin) (Fisher Scientific). Flies were than embedded in paraffin, longitudinally sectioned and hematoxylin and eosin (HE) stained following routine protocol at Penn Cancer Histology Core (https://somapps.med.upenn.edu/pbr/portal/hist/). Images of fly eyes were obtained using an Eclipse TE2000-U inverted microscope (Nikon) equipped with Plan Apo 20× /0.75 objective and MetaMorph Imaging software (Molecular Devices), and processed with ImageJ (Abramoff *et al.* 2004). Three flies of each genotype/photoperiod/temperature combination were analyzed. Within each individual fly, three sections were processed to account for the randomness of sectioning. Within each section, 10 ommatidia that are approximately equally spaced were measured for their thickness. The *Cry* mutant used in this study has a *Minos* transposon inserted in the first intron of *Cry* (genotype: w^{1118}; $Mi\{ET1\}Cry^{MB08319}$, Bloomington Drosophila Stock Center line number: 26106). This mutant was previously shown to produce less than 10% transcripts of that of wild-type flies (Metaxakis *et al.* 2005; Kuraishi *et al.* 2011).

Results

Allele effects on fitness-related phenotypes

Populations fixed for either allele of the seasonal SNP on *Cry* were tested on a panel of phenotypic traits, including lifespan, early-life fecundity, chill coma recovery rate, starvation resistance, desiccation resistance and diapause propensity. Of these traits, diapause propensity is of particular interest because multiple lines of evidence support that it is subject to strong seasonal selection (Schmidt & Conde 2006; Cogni *et al.* 2013). The other traits are chosen because they vary predictably with season and with latitudinal gradient (Schmidt *et al.* 2005b; Rajpurohit *et al.* 2008a; Behrman *et al.* 2015), are genetically correlated with diapause propensity (Schmidt *et al.* 2005b; 2008), and represent aspects of life-history strategies that may be under the distinct selective pressures associated with overwintering survivorship and summer population expansion.

The same tests were conducted on populations fixed for either allele of a control SNP on *Cry*, which has similar average allele frequency as Cry^{536} but does not cycle seasonally. The traits affected by the allele state of the seasonal polymorphism should differ specifically between alleles of the seasonal SNP, but not between alleles of the control SNP. If both the seasonal and control SNP have effects on certain traits, we cannot make any inference of the association between the phenotype and the allele state of the seasonal SNP, but attribute it to line effects or cage effects.

Diapause incidence Because diapause incidence varies among seasons, and the allele frequency of Cry^{536} oscillates with season as well, we hypothesized that the summer and winter alleles are associated with distinct diapause propensity, and thus

patterns of diapause incidence would be distinct between individuals homozygous for different alleles of Cry^{536}. Since *Cry* is a structural constituent of the compound eye, which is associated with light sensing, and photoperiod affects the diapause propensity of *D. melanogaster* (Saunders *et al.* 1989; Tauber *et al.* 2007) (but see (Emerson *et al.* 2009)), we further hypothesized that the seasonal alleles may exhibit non-parallel responses to changes in photoperiod.

Photoperiod significantly affected diapause incidence in the test populations: a larger proportion of flies expressed diapause under short day photoperiods as compared to under long day, consistent with previous observations that short day promotes the expression of diapause (Saunders & Gilbert 1990; Tauber *et al.* 2007). Individuals homozygous for $Cry^{536:T(S)}$ and $Cry^{536:C(W)}$ do not differ in their overall diapause incidence. However, there was a significant interaction between genotype and photoperiod: the diapause propensity of the summer allele was much more distinct between short day and long day, compared to that of the winter allele (Figure 2A, Table 1). For the control SNP, significant effect of photoperiod was also observed, suggesting that diapause is responsive to photoperiod regardless of genotype. The genotype×photoperiod interaction was not significant in the control SNP (Figure 2B, Table 1), thus the allelic effect on responsiveness to photoperiod is specific to the seasonal SNP.

The differential response to photoperiod of alleles of the seasonal SNP was also seen in the quantitative complementation test (Figure 2C, Table 2). The significant three-way interaction term of genotype×photoperiod×chromosome showed failure to complement, and comparison of variance between deficiency and balancer backgrounds

supported allelism, as variation among alleles and photoperiods in the deficiency background was much greater than variance in the balancer background. Although quantitative complementation tests have been used with a random array of multiple wild-type alleles (Geiger-Thornsberry & Mackay 2004; Paaby & Schmidt 2008), we were able to detect the significant interaction term and increased variance in the deficiency background with just two alleles.

Taken together, these data suggest that individuals homozygous for the summer allele respond more strongly to photoperiod in eliciting diapause compared to individuals homozygous for the winter allele. This is consistent with our hypothesis that seasonal alleles respond to changes in photoperiod in a non-parallel manner.

Stress Resistance Females were more resistant to desiccation stress than males regardless of genotype, and individuals homozygous for $Cry^{536:T(S)}$ and $Cry^{536:C(W)}$ did not differ in the overall desiccation resistance. Notably however, a significant genotype×sex interaction demonstrated a sex-specific allelic response to desiccation in the seasonal SNP: individuals homozygous for the winter allele are more resistant to desiccation stress than individuals homozygous for the summer allele, and this is only seen in females but in not in males (Figure 3A, Table 3). For the control SNP, the effect of sex was significant but the genotype×sex interaction was not (Figure 3B, Table 3), demonstrating that the sex-specific allelic effect on desiccation resistance is specific to the seasonal SNP. Similarly, in the quantitative complementation test females carrying the winter allele were more resistant to desiccation than those carrying the summer allele, and the difference between alleles was greater in the deficiency background than in the wild-type

balancer background (Figure 3C). However, these differences were not statistically significant (i.e., a significant Genotype×Chromosome×Sex term; Table 4).

Starvation resistance and chill coma recovery rate were not different between either the alleles of the seasonal SNP or the alleles of the control SNP (Table 5 and 6, Supplementary Figure 1 and 2, respectively), demonstrating that the seasonal SNP affect aspects of phenotype in a trait-specific manner.

Early-life fecundity and lifespan Total fecundity of the first two weeks post eclosion did not differ between alleles of the seasonal SNP (Table 7). Similarly, per capita fecundity from day 2 to day 14 was equivalent between $Cry^{536:T(S)}$ and $Cry^{536:C(W)}$ (Table 7, Supplementary Figure 3). Although fecundity varied as a function of age, the non-significant genotype×days term suggests that the variation among days is not specific to any genotype. Qualitatively similar patterns were seen for the control SNP.

Individuals homozygous for the summer allele showed longer lifespan than those homozygous for the winter allele, and the difference was more apparent in females than in males (Table 8, Supplementary Figure 4). However, a qualitatively similar pattern was also seen in the control SNP: one allele showed longer lifespan than the other, and this difference was driven by the female sex. Because significant genotype and genotype×sex terms were observed for both the candidate and control SNPs, we do not consider lifespan as affected specifically by the allelic state of the seasonal SNP. Lifespan could be one phenotype that tends to vary between different populations, and certain DGRP lines or cage effects might have driven the observed difference in lifespan between genotypes.

To summarize, most phenotypes we screened are not affected specifically by the allele state of the seasonal SNP. However, photoperiod influenced diapause incidence and individuals homozygous for $Cry^{536:T(S)}$ respond to photoperiod more strongly than those homozygous for $Cry^{536:C(W)}$ in eliciting diapause. Males in general are less resistant to desiccation stress than females, and females homozygous for the winter allele are more desiccation resistant than those homozygous for the summer allele. We show that the alleles of the seasonal SNP on *Cry* are functionally distinct in fitness-related phenotypes.

Allele effects on Cry transcriptional expression levels

Having shown the effects of the seasonal SNP on fitness-related phenotypes, we attempted to elucidate the molecular mechanism of these phenotypic effects. The seasonal SNP is found on the single intron of *Cry*, and introns are known to contain elements that mediate gene expression (Rose 2008). We hypothesized that individuals homozygous for different alleles of the seasonal SNP have different transcriptional expression levels of *Cry*, and the differential expression may be specific to environmental conditions. To test this hypothesis, we use a full-factorial design and quantitative PCR to examine the expression levels of the two alleles under various photoperiod and temperature combinations.

Regardless of genotype and photoperiod, the *Cry* transcript was approximately 20 times as abundant under low temperature (11°C) as compared to under high temperature (25°C) (Figure 4A, Table 9). This is in agreement with the modENCODE Treatments

RNAseq, where adults exposed to extended cold show up-regulated *Cry* expression (Li *et al.* 2014).

The summer allele shows significantly higher *Cry* expression level than the winter allele under short day photoperiod (p=0.0165), while under long day photoperiod, the pattern is qualitatively the same but not statistically significant (p=0.7637) (Figure 4B). Further decomposition of the environmental treatments showed that the distinct expression levels of the alleles under short day was mainly driven by flies exposed to high temperature and short day (p=0.0397). We therefore conclude that the two alleles of the seasonal SNP are functionally distinct in patterns of *Cry* expression, and the allelic effect on gene expression is environment-specific: individuals homozygous for the summer allele show higher *Cry* expression levels than those homozygous for the winter allele under short day photoperiod, and this pattern is more pronounced at high temperature compared to at low temperature.

Allelic effects on corneal lens morphology

Since *Cry* is the major constituent of the *Drosophila* corneal lens, we hypothesized that natural allelic variation at *Cry*, if of functional significance, may affect some aspects of the eye or lens morphology. Because the allelic effect on *Cry* expression depends upon environmental conditions, we further hypothesized that allelic effects on lens morphology may be similarly sensitive to the environment. To test these hypotheses, we exposed females homozygous for either allele to different combinations of

temperature and photoperiod, and performed histological analysis of their compound eyes.

To establish a baseline phenotype, we did the same analysis on the strain $Cry^{MB08319}$, which has a *Minos* transposon inserted in the intron of *Cry* that reduces its expression to less than 10% of the wild-type (Metaxakis *et al.* 2005; Kuraishi *et al.* 2011). The *Cry* hypomorph maintains integral corneal lens structure but the lenses are significantly thinner than the wild type ($p<0.0001$) (Figure 5A, 5B). We therefore use the lens thickness to characterize lens morphology.

The lens thickness of *Cry* mutant responds to environmental factors differently than the wild-type: lenses are thicker under 25°C compared to 11°C, but the lens thickness does not respond to photoperiod (Table 10.1). While for the wild-type, the lenses are thicker under short day as compared to under long day, but do not respond to temperature (Figure 5C, Table 10.1).

For the wild-type alleles $Cry^{536:T(S)}$ and $Cry^{536:C(W)}$, we found that genotype significantly interacts with temperature, photoperiod and temperature×photoperiod (Table 10.2). These significant interactions are mainly driven by individuals exposed to short day and high temperature, where the summer allele produces much thicker lens than the winter allele ($p=0.0064$) (Figure 5D). The summer allele is more responsive to photoperiod under high temperature, while the winter allele is more responsive to photoperiod under low temperature (Figure 5D).

Previous work on *Cry* (Janssens & Gehring 1999) as well as the modENCODE Anatomy RNAseq on all genes (Li *et al.* 2014) showed that *Cry* expression starts to become observable in the central nervous system of pupae, reaches the peak in late pupae, stays high in heads of 1-day adults, and decreases rapidly in the heads as the adult flies age. The expression pattern of *Cry* in the head suggests that the corneal lens is almost fully developed before metamorphosis, but since small amount of mRNA is still present in young adults, the lens synthesis may continue at low levels in young adults. Our data demonstrate that lens morphology is affected by environmental conditions that are imposed subsequent to eclosion. The response to temperature is likely due to the lens structure constituents other than *Cry*, as the *Cry* hypomorph is able to respond to temperature. Lens morphology is influenced by photoperiod only when the flies carry wild-type *Cry*, suggesting that a functional copy of *Cry* is required for the lens morphology to respond to photoperiod at the adult life stage. The two alleles of the seasonal SNP on *Cry* respond to environmental parameters differently in terms of lens morphology, with the summer allele being more responsive to both photoperiod and temperature.

Discussion

We have multiple lines of evidences showing that a genetic polymorphism in a particular gene is likely under seasonally varying selective pressures, from both population genomics screens of seasonally oscillating polymorphisms, and transcriptomic and proteomic analyses of the diapause phenotype. Our data show that this polymorphism

affects patterns of gene expression and two phenotypes that are associated with fitness in seasonal environments: diapause expression and desiccation tolerance. Furthermore, these patterns are affected by the two hallmarks of changing seasonal environments, temperature and photoperiod. Our results provide a rare example of functional validation of polymorphism identified from –omics level screens, that connect molecular variants, phenotypes, population dynamics and environmental parameters.

Allelic effects on diapause and desiccation resistance

Given the associations between diapause incidence and environmental heterogeneity in *D. melanogaster* (Schmidt *et al.* 2005a; Schmidt & Conde 2006), as well as the seasonal allele frequency change on candidate diapause gene (Cogni *et al.* 2013), the prediction is that any winter-associated allele would be associated with a higher diapause incidence. Here, we observe that the summer and winter alleles of Cry^{536} do not differ in their overall diapause propensity; rather, the allele affect diapause in a photoperiod-dependent manner, with the summer allele being more responsive in eliciting diapause as compared to the winter allele. Diapause incidence in *D. melanogaster* is previously shown to be affected by photoperiod, with short day facilitating the expression of diapause and long day promoting ovarian development (Saunders *et al.* 1989; Tauber *et al.* 2007) (but see (Emerson *et al.* 2009)). We speculate that the winter allele is less responsive to photoperiod because winter-adapted individuals may enter diapause regardless of photoperiodic conditions, while the summer-adapted

individuals need a better prediction of forthcoming environmental changes to decide the subsequent developmental pathways to achieve optimal fitness.

Desiccation tolerance is an important fitness component in Drosophilidae: it shows consistent clinal patterns in more than a dozen species, suggesting that it is one component of climatic adaptation (Hoffmann & Harshman 1999; Rajpurohit *et al.* 2013; Rajpurohit & Nedved 2013). In North America where the seasonally oscillating SNPs were identified and the inbred lines comprising the test populations originated, winter is the season associated with low humidity; thus, if *Cry* contributes to patterns of desiccation tolerance in the field, the prediction would be that the winter allele exhibits increased tolerance to desiccating conditions. Our observation is consistent with the prediction. The mechanism as to why *Cry* contributes to desiccation tolerance is less clear: although *Cry* has a short chitin-binding domain and shows sequence similarity to insect cuticular protein (Janssens & Gehring 1999), it is not present in the epicuticle (Komori *et al.* 1992) and therefore may not be directly responsible for the cuticle water permeability. However, because *Cry* is a constituent of the peritrophic matrix that lines the gut epithelium (Kuraishi *et al.* 2011), and gut epithelium plays critical role in insect osmobalance (Huang *et al.* 2015), it is possible that the alleles of *Cry* differentially regulate water loss from the inside of the fly instead from the outside.

We showed that in our test populations, the alleles of Cry^{536} are functionally distinct in a subset of fitness-related phenotypes. We would be cautious to extrapolate the allelic phenotypic effects to natural populations. Future work will be focused on

examining the parallels between laboratory based fitness assays and dynamics in a natural setting.

Allelic effects on gene expression levels

Cry^{536} is on the single intron located at the 5' end of *Cry*. Introns may contain regulatory sequences involved in transcriptional control (Rose 2008), especially the 5'-most "first" introns (Park *et al.* 2014). In the case of *Cry*, the intron sequence by itself gives specific expression patterns in imaginal discs regardless of its orientation, supporting that the intron contains enhancer activity(Janssens & Gehring 1999). However, recent effort in identifying *cis*-eQTLs with a subset of the DGRP lines did not identify any variants on the intron as associated with *Cry* expression levels, but identified a few *cis*-eQTLs of *Cry* ~800bp upstream of the transcriptional start site (Massouras *et al.* 2012). Because the *cis*-eQTLs identification focused on expression in adult whole bodies, it is possible that elements that fine-tune the gene expression at different life stages, in different parts of the body, or under different environmental conditions were not captured. We hypothesized that individuals homozygous for different alleles of the seasonal SNP have different transcriptional expression levels of *Cry*, and the differential expression may depend upon temperature and photoperiod. Consistent with the hypothesis, we showed that the alleles are functionally distinct in patterns of *Cry* expression, and this pattern is more pronounced under short day photoperiod.

Like diapause propensity, photoperiod plays a role in the allelic effect of gene expression of *Cry*. What is particularly interesting is that, we previously showed that *Cry* expression is down-regulated in heads of diapausing females as compared to non-diapausing females, and here we showed that the winter allele is associated with lower *Cry* expression under short day. Although this does not necessarily draw a direct link between the winter allele and the expression of diapause, we show that the populations fixed for the winter allele phenocopies the diapausing genotype in *Cry* expression.

The qPCR was conducted on whole body of adult females, therefore we do not know whether the differential expression of *Cry* is driven by the head, the midgut, or other parts of the body. The females used in this study were exposed to different environments shortly after eclosion to imitate the environmental conditions that the flies went through in the diapause assay, therefore we do not know how exposure to different environments at the larval and pupal stages would affect *Cry* expression. According to modENCODE Anatomy RNAseq and modENCODE Development RNAseq (Li *et al.* 2014), the expression level of *Cry* declines rapidly in the head after eclosion, but remains high in the digestive system, therefore it is possible that the pattern we saw was driven by the gut. modENCODE Treatment RNAseq shows that exposure to extended cold would increase *Cry* expression, but it is not clear which part of the body drives the elevation in *Cry* expression levels.

Allelic effects on lens morphology

We showed that a functioning copy of *Cry* is required for the corneal lens morphology to respond to photoperiod, and that the two alleles of Cry^{536} respond differently to environmental parameters: the summer allele is more responsive to both temperature and photoperiod, and the two alleles yield the most differentiated lens under short day and high temperature.

The allelic and environmental control of lens thickness is consistent with *Cry* expression in some but not all ways. Low expression of *Cry* in the hypomorph produces flies with a thin lens. Under short day and high temperature, the summer allele shows significantly higher *Cry* expression level, as well as significantly thicker lens compared to the winter allele. However, although *Cry* expression under high temperature is just a fraction of that under low temperature in the wild type, lens thickness is not regulated by temperature in the wild type. We speculate that temperature may affect the cornea lens morphology via other lens components, as *Cry* hypomorph would produce thicker lens under high temperature than under low temperature.

Here the assays were done on adult, when lens is already formed and *Cry* expression in the head has started to decline. We focused on adults because we were primarily interested in allelic effects on phenotypes that affect fitness in the adult stage (e.g. diapause, desiccation), instead of developmental plasticity. The results demonstrate that flies are still able to modulate aspects of morphology in response to environmental cues after eclosion. It would be interesting to expose flies to various environmental conditions at earlier life stages and test the allelic effects on lens morphology, where we expect to see greater magnitude of phenotypic plasticity.

While a change in the thickness of the lens may directly affect optical function such as light transmittance and refractive index, this was not addressed in our study; whether the observed morphological differences result in differential perception and downstream phenotype (i.e., diapause expression) requires further investigation.

The effects of photoperiod

In seasonal environments, exploiting the favorable season and avoiding or mitigating the harsh season by precisely timing processes such as migration, reproduction, hibernation and diapause is essential for evolutionary success. In temperate regions, environmental factors that determine organismal survival and reproduction such as temperature, humidity and food availability vary among seasons and also from year to year, while the seasonal change of photoperiod, as an anticipatory cue, provides a stable signal that enables the animals to time their seasonal activities (Tauber *et al.* 1986; Bradshaw & Holzapfel 2007).

Photoperiodic response in insects is generally thought to comprise a sequence of several events: 1) photoreception; 2) measurement of day or night length; 3) accumulation of photoperiod conditions by a "counter" mechanism; 4) downstream neuronal or endocrine regulation that triggers seasonal responses such as diapausing or non-diapausing developmental pathway (Goto 2013; Meuti & Denlinger 2013; Saunders 2014). It has been suggested (Bunning 1960) and also shown (Tauber *et al.* 2007; Goto 2013; Saunders 2014; Pegoraro *et al.* 2014) that elements of the circadian clocks are

recruited in producing photoperiodic responses. It has also been argued that the sensitivity of light receptors in the clock needs to be attenuate in higher latitude to keep accurate because of the increase in duration of the daily light (Pittendrigh & Takamura 1989; Pittendrigh *et al.* 1991). One potential mechanism to dampen circadian photoresponsiveness at high latitudes may be the filtering of light input into the photoreceptors (Sandrelli *et al.* 2007). Because *Cry* is a major constituent of the corneal lens and is enriched in the layers with higher refractive index (Komori *et al.* 1992), we speculate that natural variants of *Cry* respond differently to photoperiod by influencing the very first component of photoperiodic response: different alleles give distinct lens structure and function, result in different light input into the photoreceptors, and eventually gives rise to different phenotypic responses to photoperiod. We predict that the properties of light such as amplitude and wavelength should shift after penetrating the cornea lens, and the light properties change may be distinct between individuals homozygous for different alleles of the seasonal SNP.

Tables

Table 1. Mixed-model ANOVA for diapause propensity

		Seasonal SNP			Control SNP		
		df	*F*	*p*	*df*	*F*	*p*
Fixed Effects	Genotype	1	3.1169	0.2195	1	2.8717	0.2322
	Photoperiod	1	11.4172	0.0038*	1	16.6957	0.0009*
	Genotype×Photoperiod	1	5.8522	0.0278*	1	2.1225	0.1645
Random Effects		Var ratio	Pct total		Var ratio	Pct total	
	Cage[Genotype]	-0.1398	0.000		0.6874	17.450	
	Batch	0.2201	18.041		2.2519	57.165	
	Residual		81.959			23.385	

Table 2. Statistical results for diapause complementation test

ANOVA				
		df	*F*	*p*
Fixed Effects	Genotype	1	0.0000	0.9950
	Chromosome	1	194.8136	<0.0001*
	Genotype×Chromosome	1	0.0120	0.9162
	Photoperiod	1	25.9703	0.0022*
	Genotype×Photoperiod	1	6.9531	0.0386*
	Chromosome×Photoperiod	1	0.0568	0.8195
	Genotype×Chromosome×Photoperiod	1	10.1559	0.0188*
Random Effects		Var Ratio	Pct Total	
	Cage[Genotype]	0.0688	6.435	
	Residual		93.565	

Variance comparison				
	SS^2_{Bal}	SS^2_{def}	*F*	σ^2_{Bal} not greater than σ^2_{def}?
Genotype	0.0012	0.0043	0.2791	✓
Photoperiod	0.6907	0.4219	1.6371	✓
Genotype×Photoperiod	0.0073	0.7783	0.0094	✓

Table 3. Mixed-model ANOVA for desiccation resistance

		Seasonal SNP			Control SNP		
		df	*F*	*p*	*df*	*F*	*p*
Fixed Effects	Genotype	1	5.9789	0.1345	1	0.7010	0.4905
	Sex	1	530.08	<0.0001*	1	610.87	<0.0001*
	Genotype×Sex	1	8.5121	0.0037*	1	1.0580	0.3043
Random Effects		Var ratio	Pct total		Var ratio	Pct total	
	Cage[Genotype]	0.0561	5.313		0.0184	1.807	
	Residual		94.687			98.193	

Table 4. Statistical results for desiccation resistance complementation test

ANOVA

		df	*F*	*p*
Fixed Effects	Genotype	1	0.2070	0.6938
	Chromosome	1	109.1159	<0.0001*
	Genotype×Chromosome	1	1.5158	0.2186
	Sex	1	1110.271	<0.0001*
	Genotype×Sex	1	22.2454	<0.0001*
	Chromosome×Sex	1	10.4710	0.0013*
	Genotype×Chromosome×Sex	1	0.1191	0.7301
Random Effects		Var Ratio	Pct Total	
	Cage[Genotype]	0.1930	16.177	
	Residual		83.823	

Variance comparison

	SS^2_{Bal}	SS^2_{def}	*F*	σ^2_{Bal} not greater than σ^2_{def}?
Genotype	12.0390	74.1365	0.1624	✓
Sex	5761.32	3839.72	1.5005	✓
Genotype×Sex	82.3299	105.7803	0.7783	✓

Table 5. Mixed-model ANOVA for starvation resistance

		Seasonal SNP			Control SNP		
		df	*F*	*p*	*df*	*F*	*p*
Fixed Effects	Genotype	1	1.0378	0.4156	1	0.0282	0.8821
	Sex	1	164.26	<0.0001*	1	169.69	<0.0001*
	Genotype×Sex	1	0.8838	0.3479	1	0.0202	0.8870
Random Effects		Var ratio	Pct total		Var ratio	Pct total	
	Cage[Genotype]	0.4090	29.029		0.0734	6.835	
	Residual		70.971			93.165	

Table 6. Mixed-model ANOVA for chill coma recovery rate

		Seasonal SNP			Control SNP		
		df	*F*	*p*	*df*	*F*	*p*
Fixed Effects	Genotype	1	1.0067	0.4215	1	1.9367	0.2987
	Sex	1	316.27	<0.0001*	1	254.54	<0.0001*
	Genotype×Sex	1	0.5236	0.4696	1	0.3129	0.5761
Random Effects		Var ratio	Pct total		Var ratio	Pct total	
	Cage[Genotype]	0.0515	4.898		0.0155	1.525	
	Residual		95.102			98.475	

Table 7. Statistical results of early life fecundity

Mixed-model ANOVA for total fecundity of the first two weeks post eclosion							
		Seasonal SNP			Control SNP		
Fixed Effect		*df*	*F*	*p*	*df*	*F*	*p*
	Genotype	1	0.4420	0.5746	1	2.9589	0.2275
Random Effects		Var ratio	Pct total		Var ratio	Pct total	
	Cage[Genotype]	0.8831	46.897		-0.2309	0.000	
	Residual		53.103			100.000	
Repeated measure ANOVA for per capita fecundity							
		Seasonal SNP			Control SNP		
Fixed Effects		*df*	*F*	*p*	*df*	*F*	*p*
	Genotype	1	0.4504	0.5713	1	3.0129	0.2247
	Days since Eclosion	11	12.1889	<0.0001*	11	11.1190	<0.0001*
	Genotype×Days	11	0.7336	0.7049	11	0.6402	0.7922
Random Effects		Var ratio	Pct total		Var ratio	Pct total	
	Cage[Genotype]	0.2635	17.826		-0.2973	0.000	
	Bottle[Cage]	0.2149	14.533		1.2258	55.071	
	Residual		67.641			44.929	

Table 8. Proportional hazards analysis of lifespan

	Seasonal SNP			Control SNP		
	df	χ^2	*p*	*df*	χ^2	*p*
Genotype	1	17.8869	<0.0001*	1	57.7192	<0.0001*
Sex	1	0.1763	0.6746	1	1.1049	0.2932
Genotype×Sex	1	5.2072	0.0025*	1	20.0985	<0.0001*

Table 9. Mixed-model ANOVA for qPCR and within model contrasts

ANOVA				
Fixed Effects		*df*	*F*	*p*
	Genotype	1	10.5372	0.0832
	Temperature	1	664.5048	<0.0001*
	Genotype×Temperature	1	0.128	0.7219
	Photoperiod	1	0.0836	0.7736
	Genotype×Photoperiod	1	1.5263	0.222
	Temperature×Photoperiod	1	3.8542	0.0548
	Genotype×Temperature ×Photoperiod	1	1.8013	0.1852
Random Effects		Var Ratio	Pct Total	
	Cage[Genotype]	-0.0478	0.000	
	Residual		100.000	

Table 10. Statistical results for corneal lens thickness

Table 10.1 Responses of mutant and wild type *Cry* to environmental parameters

		Mutant *Cry*			Wild type *Cry*		
		df	*F*	*p*	*df*	*F*	*p*
Fixed Effects	Temperature	1	15.6405	<0.0001*	1	1.1430	0.2852
	Photoperiod	1	0.3908	0.5323	1	40.4215	<0.0001*
	Temperature×Photoperiod	1	0.0704	0.7908	1	1.1972	0.2741

Table 10.2 Allelic response to environmental parameters

ANOVA				
		df	*F*	*p*
Fixed Effects	Genotype	1	1.6181	0.3312
	Temperature	1	1.1883	0.2759
	Genotype×Temperature	1	11.3545	0.0008*
	Photoperiod	1	42.0258	<0.0001*
	Genotype×Photoperiod	1	8.3461	0.0039*
	Temperature×Photoperiod	1	1.2447	0.2647
	Genotype×Temperature ×Photoperiod	1	22.2590	<0.0001*
Random Effects		Var Ratio	Pct Total	
	Cage[Genotype]	0.0134	1.319	
	Residual		98.681	

Figures

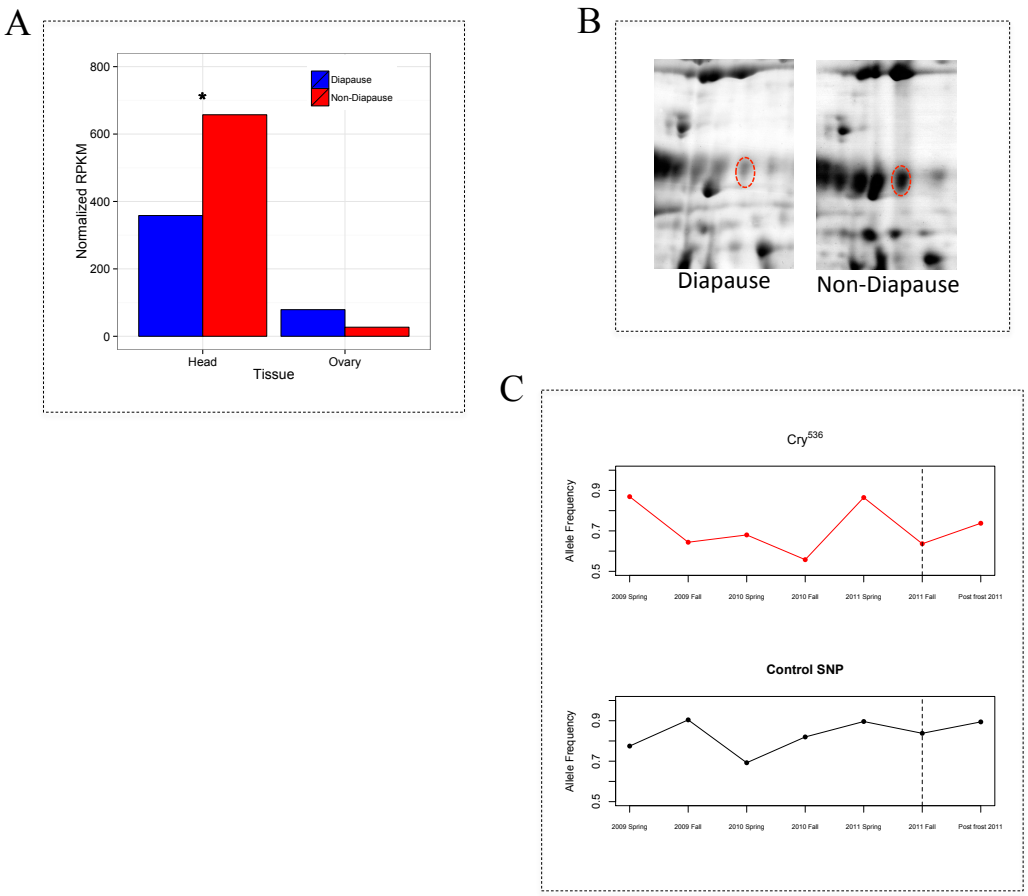

Figure 1. *Cry* is an intersection of several global screens related to seasonal adaptation.

A. *Cry* is significantly down-regulated in heads of diapausing individuals as compared to non-diapausing individuals in a global transcriptomic profile of diapause. B. *Cry* protein abundance is significantly down-regulated in heads of diapausing individuals as compared to non-diapausing individuals in a proteomic screen of diapause. There is a 4.3 fold increase in the non-diapause sample. C. One SNP (Cry^{536}) on *Cry* has allele frequency cycling with season in three consecutive years. The allele frequency became more "spring-like" after an acute frost in 2011. The control SNP used in the phenotypic assays does not cycle with season.

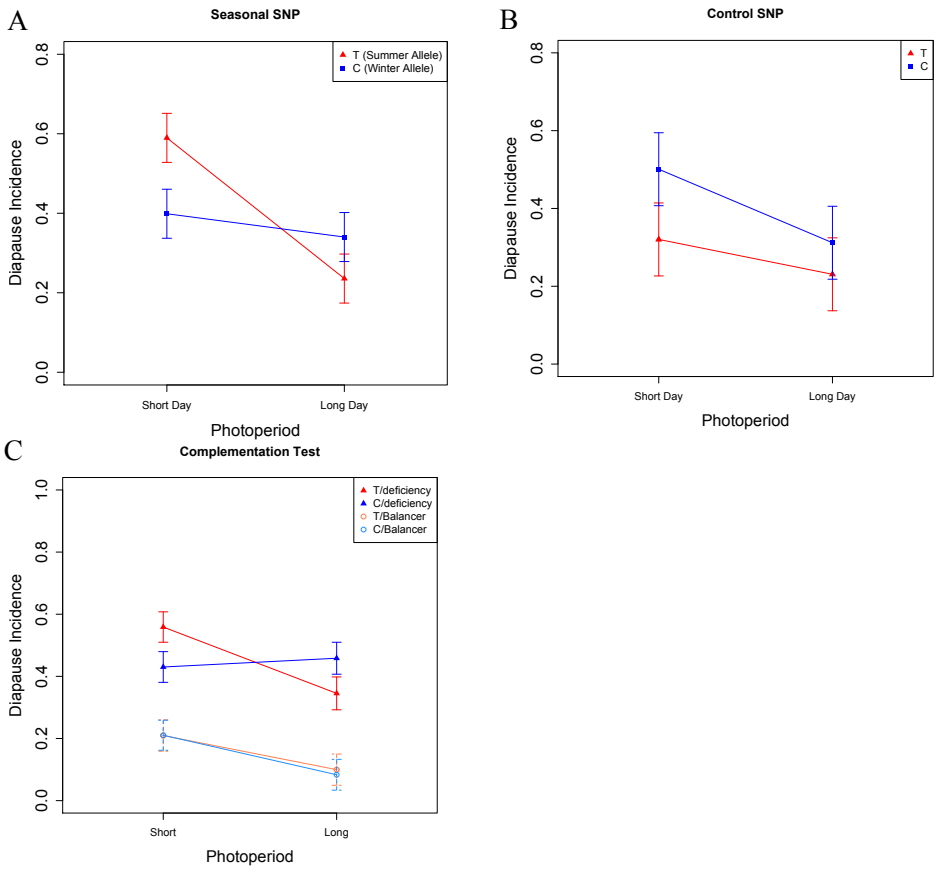

Figure 2. Allele state of the seasonal SNP on *Cry* affects diapause incidence in a photoperiod-specific manner.

A. The alleles of the seasonal SNP respond differently to photoperiod in eliciting diapause, with the summer allele being more responsive to photoperiod. B. The allele state of the control SNP does not affect diapause incidence. C. Quantitative complementation test showed failure to complement, and the variance in the deficiency background is grater than that in the balancer background, consist with the results from tests on populations fixed for either allele.

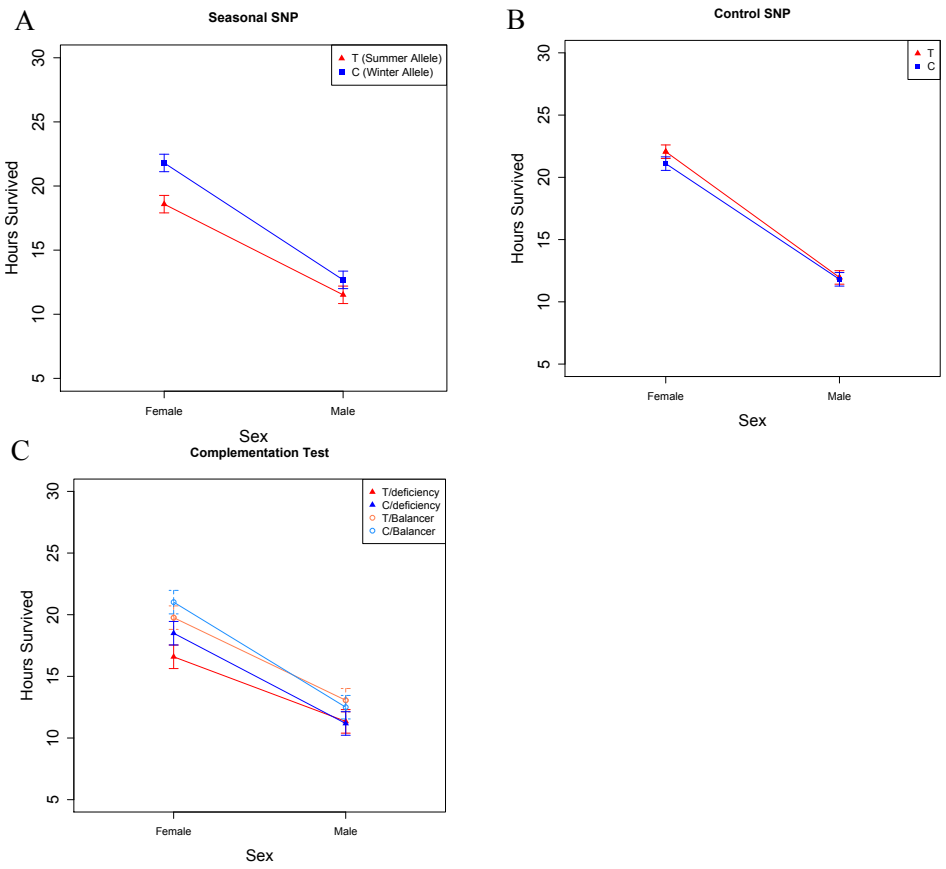

Figure 3. Allele state of the seasonal SNP on *Cry* affects desiccation resistance in a sex-specific manner.

A. The winter allele is more resistant to desiccation stress compared to the summer allele, and this effect is only seen in females. B. The allele state of the control SNP does not affect desiccation tolerance. C. Quantitative complementation test show that females carrying the winter allele were more resistant to desiccation than those carrying the summer allele, and the difference between alleles was greater in the deficiency background than in the wild-type balancer background, although the differences were not statistically significant.

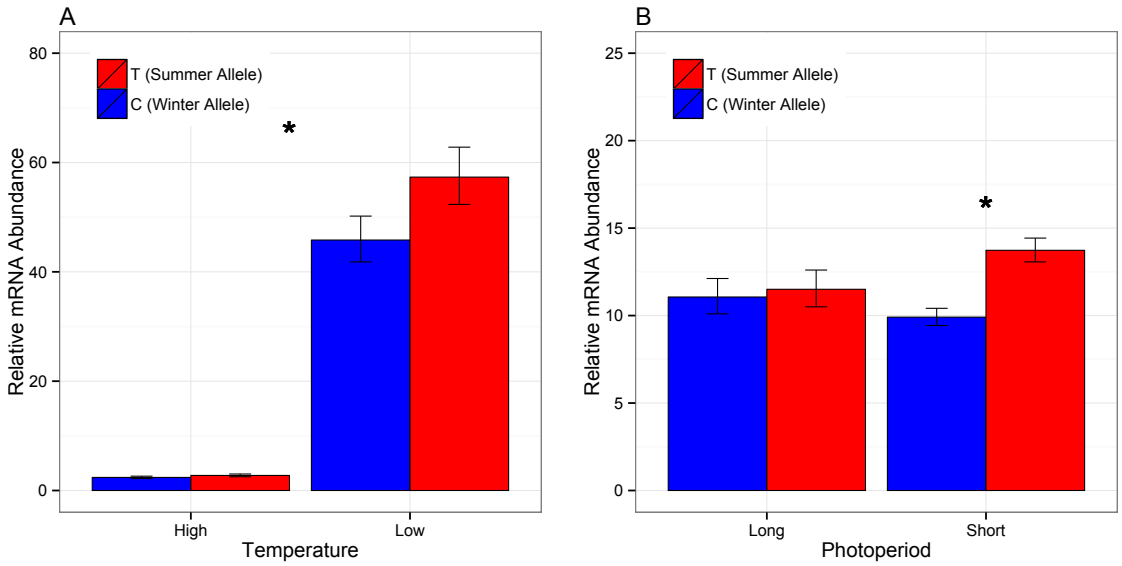

Figure 4. Allele state of the seasonal SNP on *Cry* affects *Cry* expression in a environment-specific manner.

A. The expression of *Cry* is significantly higher under low temperature (11°C) as compared to under high temperature (25°C). The two alleles do not different in their expression level under either temperature. B. The expression of *Cry* is significantly higher in individuals homozygous for the summer allele as compared to the winter allele under short day photoperiod (9h dark: 15h light), but not under long day photoperiod (15h dark: 9h light).

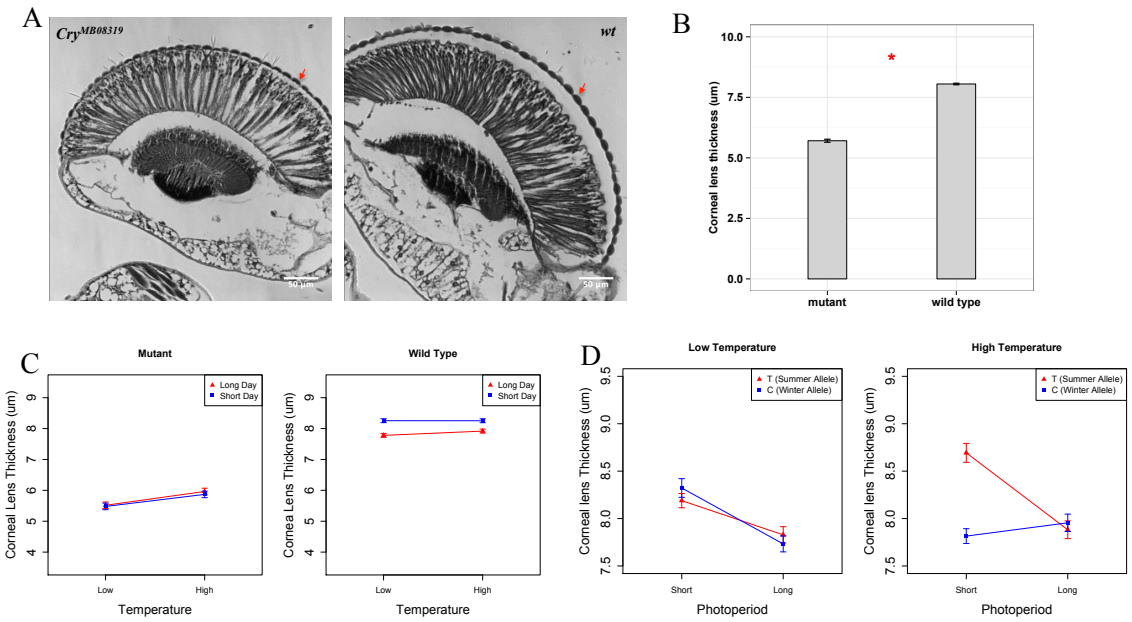

Figure 5. *Cry* affects the corneal lens morphology.

A. Longitudinal sections showing the corneal lens of flies carrying mutant (left) and wild type (right) copy of *Cry*. B. The corneal lenses are significantly thinner in flies carrying *Cry* hypomorph, as compared to wild type. C. *Cry* mutant and wild type respond to temperature and photoperiod differently: flies carrying the hypomorph of *Cry* show thicker lens under high temperature compared to under low temperature, but do not differ in lens thickness under different photoperiod conditions; flies carrying wild type copies of *Cry* show thicker lenses under short day compared to under long day, but do not differ under different temperatures. D. Flies homozygous for the summer allele give thicker lens compared to those homozygous for the winter allele under short day and high temperature condition.

Chapter Three

Genomic and transcriptomic analyses of diapause in *Drosophila melanogaster* in the context of adaptation to spatial and temporal environmental heterogeneity

Introduction

Understanding the genetic basis of adaptation remains one major goal of ecological and evolutionary biology. Genetic analyses of fitness-related traits with simple genetic architecture has successfully identified genes and polymorphisms responsible for the variation in adaptive phenotypes (Nachman *et al.* 2003; Swallow 2003; Shapiro *et al.* 2004; Tishkoff *et al.* 2007). The genetic basis of many complex traits however, especially those underlying differential life-history strategies, remains unclear despite extensive investigation (Denlinger 2002; Ehrenreich *et al.* 2009). With the development of genomic, transcriptomic, proteomic and metabolomic technologies, depicting the phenotype-genotype maps of complex traits has become possible.

Diapause provides an ideal model to study the genetic mechanism and evolutionary dynamics of complex traits that are intimately associated with fitness. It is a genetically determined syndrome cued by photoperiod and/or temperature that results in lifespan extension, delayed senescence, increased stress tolerance, and reproductive quiescence (Tauber *et al.* 1986). It is the primary adaptation in invertebrates to survive unfavorable seasons (Leather *et al.* 1995). Unlike quiescence, which is induced passively and directly by environmental stress, diapause is a dynamic alternative developmental pathway that animals enter usually in response to a number of environmental stimuli that

precede unfavorable conditions (Denlinger 2002). As such, this phenotype represents a logical system for the study of the genetic control of environmental sensing, biological rhythms, and the aging process. However, while diapause has been extensively characterized at the physiological level over the past 60 years, the genetic and molecular mechanisms that sense environmental change to induce diapause expression and stimulate the associated phenotypic cascade are largely unknown (Ragland *et al.* 2010).

The genetic model system *Drosophila melanogaster* has an adult reproductive diapause. It is induced by moderately low temperature and shortened photoperiod, and results in extreme lifespan extension, negligible senescence and increased stress tolerance (Saunders *et al.* 1989; Saunders & Gilbert 1990; Tatar & Yin 2001; Schmidt *et al.* 2005a; b). Unlike many other invertebrate systems, diapause expression is highly variable within and among *D. melanogaster* populations: when exposed to the appropriate cues, some genotypes tend to become reproductively quiescent whereas others tend to proceed with vitellogenesis and reproductive development (Williams & Sokolowski 1993). The natural variation of diapause propensity provides an opportunity to identify genetic polymorphisms co-segregating with the phenotype, which are presumably in linkage disequilibrium with the causal loci. We can also take advantage of the natural variation to identify genes and transcripts expressed at different levels as a function of the diapause phenotype. By combining the genome-wide association study and the transcriptional profile of diapause, it is possible to reveal the pathways controlling diapause induction, progression, and termination, and identify the downstream targets that may ultimately result in the observed phenotypes.

Several lines of evidences supports that diapause propensity underlies adaptation to spatially and temporally varying selective pressures in temperate regions. The incidence of reproductive diapause exhibits a strong latitudinal cline in eastern North American populations: greater than 90% in temperate areas and ~30% in neotropical habitats (Schmidt *et al.* 2005a). Diapause incidence also varies with season, with high incidence (~90%) in spring, and lower incidence (40%-50%) in fall (Schmidt & Conde 2006). Similarly, genetic polymorphisms in candidate diapause genes also vary clinally (Fabian *et al.* 2012; Bergland *et al.* 2014) and seasonally (Cogni *et al.* 2013). Furthermore, diapause incidence is shown to be genetically linked to a suite of other traits associated with organismal fitness: even without exposure to diapause-inducing conditions, genotypes that show a high diapause rate are constitutively longer-lived, less fecund, and more stress resistant than the lines that show a low diapause incidence (Schmidt *et al.* 2005b; 2008). These correlated traits have also been shown to be clinal (e.g. Hoffmann et al. 2002; Schmidt et al. 2005b) as well as seasonal (Behrman *et al.* 2015). The clinal and seasonal patterns of diapause incidence and its association with other life-history traits may represent trade-off between somatic maintenance and reproduction: under stressful and unfavorable conditions, hardiness is favored despite being associated with lower fecundity; while under benign conditions, high reproductive rate is favored at the cost of aspects of survival. We hypothesize that diapause is one major determinant of adaptation to spatial and temporal environmental heterogeneity in this taxon, therefore genetic polymorphism co-segregating with diapause propensity, as well as genes differentially regulated as a function of the diapause phenotype are likely

under spatially and/or temporally varying selectively pressure, thus sequence polymorphisms on the involved genes may show clinal and seasonal patterns.

Here we perform a genome-wide association study as well as a transcriptional profile of diapause using natural populations of *Drosophila melanogaster*. The populations used in this study were collected throughout the growing season from mid-Atlantic region of North America, which represents the midpoint of the previously observed latitudinal cline of diapause propensity, thus the variation of the diapause phenotype, as well as the variation in allele frequencies of genetic polymorphisms underlying the phenotype is presumably maximized. The heads and ovaries, two tissue types associated with neuroendocrine signaling, of diapausing and non-diapausing populations were used in the transcriptional profile, and the carcasses of the same diapausing and non-diapausing populations were used in the genome-wide association study. The fact that the DNA and RNA are from the same populations makes it possible to integrate the genomic and transcriptomic level screens. We show that diapause is a multi-genic trait that is actively regulated at the transcriptional level. The genetic polymorphisms co-segregating with diapause propensity show signatures of clinal and seasonal adaptation. We also have evidence that a subset of diapause SNPs may affect the phenotype by altering expression levels of candidate genes. Functional validation of the candidates, as well as more rigorous statistical tests are needed to establish the link between sequence variation, gene expression, phenotype and fitness.

Materials and Methods

Fly populations

D. melanogaster were collected using banana and yeast traps and by direct aspiration on wind fallen fruit from wild populations at the Linvilla Orchard in Media, PA (39.9°N, 75.4°W). The populations were collected throughout the growing season of 2011 in June, July, September and November. Isofemale lines were established from individual females shortly subsequent to collection. Within six months after the flies were collected from the wild, 241 isofemale lines were used to construct an outbred population. Ten females and ten males of a single age cohort from each isofemale line were released into a population cage and were allowed to interbreed for six generations. Four cages were created from the F1 of the initial cohort, and maintained as biological replicates.

Diapause assay

Flies in the population cages were allowed to lay eggs on the surfaces of standard cornmeal-molasses medium. Eggs were reared at low density (<40 eggs per vial) at 25°C, 12 hour Light: 12 hour Dark. Females were collected within 2 hours post-eclosion and were immediately exposed to diapausing-inducing conditions (11°C at a photoperiod of 9h light: 15h dark). After being exposed to low temperature and short days for four weeks, all experimental females were dissected and the developmental status of the ovaries was assessed (King 1970). A female was scored as diapausing (D) if the most advanced oocyte was arrested before vitellogenesis (before stage 8); a female was scored as non-diapausing (ND) if vitellogenin was observed in either ovary (stage 8 or later).

Genomic DNA library construction, cDNA library construction and sequencing

Pooled genomic DNA libraries and cDNA libraries were prepared for sequencing on the Illumina HiSeq 2000 platform. The heads and ovaries of either diapausing or non-diapausing females from each of the four biological replicates were pooled and preserved in RNAlater reagent (Ambion) for total RNA extraction. The carcasses of the same population pools were preserved in RNAlater for genomic DNA extraction. The numbers of individuals of each pooled sample can be found from Table 1.

Genomic DNA was extracted and libraries for Illumina sequencing were prepared as previously described (Bergland *et al.* 2014). Libraries were sequenced at the Sequencing Service Center at the Stanford Center for Genomics and Personalized Medicine on a HiSeq 2000 platform using standard protocol for 200bp paired-end read sequencing.

Total RNA was extracted using TRIzol reagent following manufacture's protocol (Life Technologies). cDNA libraries were constructed using Illumina TruSeq RNA Library Preparation Kit v2 following manufacture's protocol (Illumina). Libraries were sequenced on a HiSeq 2000 platform using standard protocol for 100bp single-end read sequencing.

Pooled genomic DNA sequencing: Read alignment, SNP identification and filtering, association test and enrichment test of identified SNPs

Raw paired-end sequence reads for the pooled genomic sequencing were mapped to the *D. melanogaster* genome version 5.57 using the *bwa* aligner version 0.7.10 with

the BWA-MEM algorithm and default options (Li *et al.* 2009). PCR duplicates were marked using the "MarkDuplicates" utility of Picard Tools version 1.127 (http://broadinstitute.github.io/picard/). Local realignment around indels was performed using the "RealignerTargetCreator" and "IndelRealigner" utilities of GATK tools version 3.3-0 (McKenna *et al.* 2010). We used VarScan version 2.3.7 (Koboldt *et al.* 2012) to call SNPs, and included all reads with read quality over 10. SNPs mapped to repetitive regions (obtained from the UCSC genome browser: https://genome.ucsc.edu/) were excluded from the subsequent analysis. SNPs that are not present in the DGRP freeze 2 (Mackay *et al.* 2012) dataset were excluded to ensure that SNPs analyzed were not sequencing artifacts. Only SNPs with minimum read depth greater than or equal to 5 and maximum read depth smaller than or equal to 200 in all populations were included, and only SNPs with the minor allele frequency greater than or equal to 5% across all samples were included in subsequent analysis. Around 1,900,000 SNPs were initially identified, and about 1,400,000 were left after applying these filters (Table 2). SNPs were annotated using SnpEff version 4.1 (Cingolani *et al.* 2012) and SnpEff database BDGP5.78.

SNPs that have different allele frequencies between diapausing and non-diapausing populations were identified using the Cochran-Mantel-Haenszel test for repeated tests of independence (http://www.biostathandbook.com/cmh.html). Benjamini-Hochberg procedures (Benjamini & Hochberg 1995) were applied to control for false discovery rate. These tests were implemented in R version 3.2.0 (http://cran.r-project.org/).

Gene Ontology enrichment tests of the identified SNPs were performed using the SNP mode of Gowinda (Kofler & Schlotterer 2012), which is designed specifically for genome wide association studies to account for gene lengths and spatially overlapping genes. The false discovery rate threshold was set to be 0.05.

Overlap of diapause SNPs and seasonal/clinal SNPs

Seasonal and clinal SNPs were identified based on the results of Bergland et al. 2014, where SNPs with seasonally-oscillating frequencies and SNPs with frequencies that vary with the latitude of geographical origins of populations were identified. SNPs with a seasonal q value of 0.35 or less were considered seasonal, and SNPs with a clinal q value of 0.1 or less were considered clinal, as per Bergland et al. 2014. Genes with at least one seasonal/clinal SNP on them, or within 5Kb upstream or downstream of an identified SNP were considered seasonal/clinal genes, respectively.

We identified SNPs that segregate as a function of diapause and were also identified as being clinal. If the overlapping SNPs show differentiated allele frequencies along the latitudinal gradient because of their association with different diapause propensity, we would expect that the alleles associated with high diapause propensity to be at higher frequency in populations from high latitude. We checked if the majority of the overlapping alleles show the expected pattern.

Because the number of seasonal SNPs is small we did not find appreciable number of seasonal SNPs that also segregate with the diapause phenotype; therefore, we conducted the overlap analysis at gene level. We identified genes that harbor diapause

SNPs, and genes that harbor seasonal SNPs, and determined if diapause genes are enriched in the seasonal gene set using a G test assuming all genes in the genome have equal chances to be called seasonal, and all seasonal genes have equal chance to harbor diapause SNPs.

Pooled mRNA sequencing: Read alignment, count estimation, differential expression and enrichment test

Raw single-end sequence reads were mapped to the *D. melanogaster* reference transcriptome, and expression levels of genes and isoforms were estimated following the RSEM pipeline version 1.2.19 (Li & Dewey 2011). The reference transcriptome was constructed using the *D. melanogaster* reference genome version 5.57 and the Ensembl annotation version 5.78.

EBSeq version 1.8.0 (Leng *et al.* 2013) was used to call for genes that show differential expression levels between the diapause and non-diapause samples in heads and ovaries respectively. Library sizes were normalized using the median normalization approach. Genes and isoforms that had adjusted *p*-values (FDR) smaller than 0.05 were considered differentially expressed (DE) as a function of the diapause phenotype.

Gene ontology (GO) and Kyoto encyclopedia of gene and genomes (KEGG) pathway enrichment analysis was performed using the GOseq package (Young *et al.* 2010) that accounts for transcript length bias associated with RNA-seq data. Only genes that had RPKM values greater than one in at least one of the D or ND samples in the respective tissues were used as the background gene list for that tissue. Genes that were

of significantly high and low expression in diapausing head/ovary samples were separately tested to identify which categories of genes are enhanced and suppressed as a function of diapause. The *p* values were corrected using the Benjamini and Hochberg procedure (Benjamini & Hochberg 1995), and the false discovery rate threshold was set to be 0.05.

Results and Discussion

Genome-wide association study of diapause

SNPs that have different allele frequencies in diapausing (D) and non-diapausing (ND) populations were found throughout all major chromosomes (Figure 1), demonstrating that the variation of diapause propensity is caused by multiple genetic polymorphisms with small effects. We did not identify SNPs that are alternatively fixed in D and ND populations. The largest allele frequency difference observed between populations was ~ 40%. With a false discovery rate (FDR) threshold of 0.35, we identified 1140 SNPs segregating between the two phenotypic pools. The relatively high level of FDR was justified because given the number of replicates, the sequencing depths and the potential epistatic interactions, our power to identify segregating SNPs was limited.

SNPs with different allele frequencies in D and ND were found enriched for multiple Gene Ontology (GO) enrichment terms (Supplementary Table 1), including metabolic processes (fatty acid, nucleic acid, icosanoid, monocarboxylic acid etc.), organ

formation (eye, heart, neuron etc.), regulation (crystal cell differentiation, hemocyte differentiation, DNA replication, eclosion etc.) and defense response.

Notably, 2 SNPs (3R: 13802036, 3R: 13825936) on the intronic region of the candidate diapause gene *couch potato* (*cpo*) were found segregating between D and ND. The other two candidate diapause genes, *timeless* (*tim*) and *Dp110* were not found involved given the chosen FDR cut-off.

Diapause SNPs vs. clinal and seasonal SNPs

135 SNPs were clinal and also segregate as a function of the diapause phenotype, of which 97 had allele frequencies consistent with the hypothesis that high diapause propensity is associated with temperate habitats: the allele that showed higher frequency in the diapausing populations also showed higher frequency in populations from high latitude as compared to those from low latitude. While by chance alone, we would expect to see about half of the 135 SNPs showing this pattern. The significant departure from random suggests that the latitudinal allele frequency gradients may in part be shaped by differential selection on diapause propensity or genetically linked traits. Some SNPs show latitudinal clines because of demographic processes such as population sub-division and secondary contact. It is difficult to disentangle the clines shaped by local adaptation and those shaped by demography based on population genomics screen; however, anchoring the population dynamics to a fitness-related phenotype might make it possible.

1855 genes were found harboring diapause SNPs, of which 812 were found harboring seasonal SNPs as well. Considering the total number of seasonal genes (4156)

and the total number of genes in the *D. melanogaster* genome (15682), we had a *p* value close to 0 to reject the hypothesis that diapause genes are not more likely to be seasonal genes; in other words, genes harboring diapause SNPs are significantly enriched for seasonal genes. Our result is consistent with the hypothesis that diapause is a primary candidate phenotype of adaptation to seasonality in *D. melanogaster*.

We showed preliminary evidence that natural variation of diapause propensity underlies adaptation to spatially and temporally varying selective pressures, based on the allele frequencies of diapause SNPs and clinal SNPs, and the substantial overlaps of diapause genes and seasonal genes. However, more rigorous tests that control for linkage disequilibrium among identified SNPs, allele frequency, recombination rates and inversion status are needed to test if SNPs segregating between D and ND are more likely than expected by chance to be seasonal or clinal.

Transcriptional profile of diapause

Our data demonstrated that diapause is an actively regulated process at the transcriptional level in both heads and ovaries (Figure 2, Supplementary Table 2), as the numbers of genes that were up- and down-regulated in diapausing individuals were approximately the same in both tissues. The expression patterns were distinct in the two tissue types with respect to dispersion: genes showed much more drastic fold changes between D and ND in the ovaries as compared to in the heads. The less drastic fold changes in head might reflect that transcriptional regulation plays a less important role in head in initiating and maintaining diapuase, but a more plausible explanation is that head is a complicated structure comprising multiple parts with distinct functions, therefore

patterns of transcriptional regulation is highly heterogeneous within head, and a signal of differential expression needs to be highly robust and consistent across different structures to be seen when the head is evaluated as a whole. The ovaries, on the other hand, is a single tissue type, and because the most significant difference between diapausing and non-diapausing individuals are the development status of the ovaries, it is not surprising that many genes are differentially expressed in the ovary. With an FDR of 0.05, 187 genes were found differentially expressed in heads between diapausing and non-diapausing individuals, of which 47 were up-regulated in D and 140 were down regulated. 5339 genes were differentially expressed in the ovaries of diapausing and non-diapausing individuals, of which 2424 were up-regulated in D and 2915 were down-regulated in D.

Differentially expressed genes were found enriched for multiple GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Table 3, Supplementary Table 3). In the heads, several metabolic pathways were down-regulated in diapausing individuals, and fatty acid degredation was shut down, suggesting a shift towards storage in diapausing flies. The up-regulation of ribosome biogenesis in diapausing ovaries may suggest similar responses to hypoxia occur in ovaries of diapausing flies (Zhao *et al.* under review).

None of the three candidate diapause genes were differentially expressed in the head, while both *cpo* and *tim* are up-regulated in diapausing ovaries. Neither *cpo* or *tim* is known for its function in the ovary. The differential expression of these genes in the ovary may suggest pleiotropic effects that were not previously studied.

As much of the genetic variation underlying observed phenotypic variation is due to modification of gene regulation (Gilad *et al.* 2008; Cookson *et al.* 2009), we were interested in whether some of the SNPs segregating with diapause propensity are found in the regulatory regions of genes differentially expressed as a function of diapause. We obtained a list of known expression quantitative trait loci (eQTL) of adult female head from King et al. 2014 (King *et al.* 2014). Of the 187 genes differentially expressed in heads of D and ND individuals, 87 have known eQTL regions, of which 67 had diapause SNPs falling in their eQTL regions. We have preliminary evidence that some diapause SNPs may affect the diapause phenotype through modification of gene regulation. More rigorous tests that control for gene lengths, expression levels, linkage disequilibrium among identified SNPs, allele frequency, recombination rates and spatial overlaps of genes are needed to test if diapause SNPs are more likely to be found on eQTL regions of the differentially expressed genes then would be expected by chance. Moreover, functional validation of individual polymorphisms is needed to establish the link between sequence polymorphism, expression regulation, diapause phenotype and ultimately organismal fitnesses in various environments.

Tables

Table 1. Number of individuals pooled in each biological replicate

Replicate	A	B	C	D	Total
Diapause	49	49	31	44	173
Non-Diapause	63	54	80	60	257

Table 2. SNP statistics of the genome-wide association study of diapause

	Number of SNPs remaining after filter
Total identified	1,934,409
Exclude repetitive regions	1,767,126
Polymorphic in DGRP	1,670,493
Read depth ≤ 10X and ≥ 200X	1,514,279
Minor allele frequency > 5%	1,392,214
Total used in analysis	**1,391,214**

Table 3. Enriched KEGG pathways of genes differentially expressed as a function of the diapause phenotype

Tissue	Pathway	FDR	No. DE in category	Total in category
Head	Up-regulated in D			
	No enrichment			
	Down-regulated in D			
	fatty acid degradation	3.48E-09	10	28
	valine, leucine and isoleucine degradation	2.20E-08	10	32
	tryptophan metabolism	8.32E-06	6	18
	metabolic pathways	1.35E-05	43	852
	pyruvate metabolism	2.02E-05	8	41
	propanoate metabolism	3.75E-05	6	22
	beta-Alanine metabolism	9.52E-05	5	17
	butanoate metabolism	1.81E-04	5	18
	lysine degradation	3.18E-04	5	20
	arginine and proline metabolism	7.50E-04	7	50
	glycerolipid metabolism	1.34E-03	6	42
	one carbon pool by folate	2.94E-03	3	11
Ovary	Up-regulated in D			
	ribosome	2.67E-50	68	80
	ribosome biogenesis in eukaryotes	3.09E-12	40	63
	pyrimidine metabolism	4.05E-04	27	71
	RNA polymerase	7.96E-04	12	24
	Down-regulated in D			
	proteasome	5.91E-10	31	40
	protein processing in endoplasmic reticulum	7.45E-09	62	109
	DNA replication	5.83E-07	23	31
	ubiquitin mediated proteolysis	1.44E-06	46	82
	SNARE interactions in vesicular transport	1.60E-05	15	19
	RNA degradation	1.29E-03	25	49
	fatty acid degradation	1.36E-03	17	28
	mRNA surveillance pathway	1.60E-03	26	50
	basal transcription factors	3.37E-03	16	29
	mismatch repair	3.67E-03	11	17

Figures

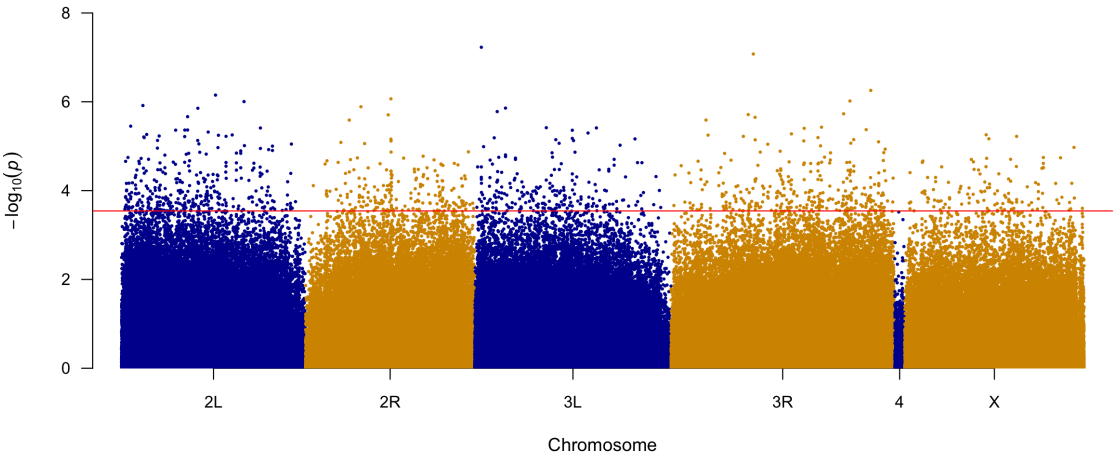

Figure 1. Manhattan plot of the genome-wide association study of diapause.
p-values are calculated using Cochran-Mantel-Haenszel test for repeated tests of independence, and small *p* values corresponds to SNPs that consistently differ in allele frequency between D and ND populations across biological replicates. The horizontal line represents *p*-value of 0.0002868, which corresponds to an FDR of 0.35. Dots above the line represent SNPs that have different allele frequencies in diapausing and non-diapausing populations.

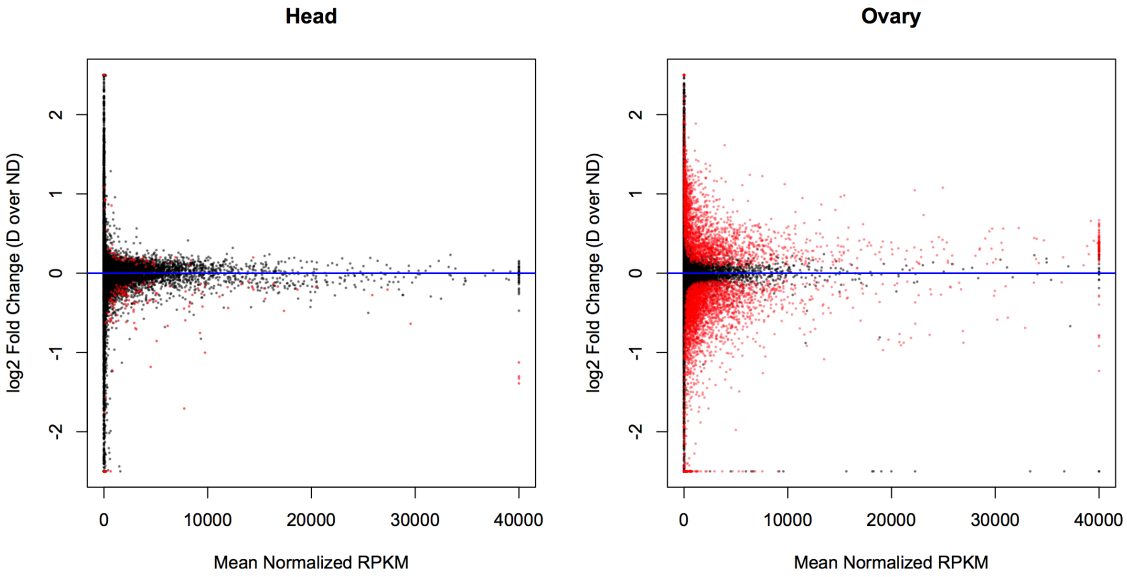

Figure 2. Log2 fold change of expression levels versus mean normalized RPKM values in head and ovary. Genes that are significantly differentially expressed (FDR<0.05) in D and ND are shown in red. Genes with relatively higher expression levels in diapausing samples as compared to non-diapausing samples fall above the horizontal line of y=0.

References

Abramoff MD, Magalhaes PJ, Ram SJ (2004) Image Processing with ImageJ. **11**, 36–42.

Ahmad K, Henikoff S (2002) The Histone Variant H3.3 Marks Active Chromatin by Replication-Independent Nucleosome Assembly. *Molecular Cell*, **9**, 1191–1200.

Apfeld J, Kenyon C (1998) Cell nonautonomy of C. elegans daf-2 function in the regulation of diapause and life span. *Cell*, **95**, 199–210.

Aronesty E (2013) Comparison of Sequencing Utility Programs. *The Open Bioinformatics Journal*, **7**, 1–8.

Azevedo RBR, James AC, McCabe J, Partridge L (1998) Latitudinal Variation of Wing: Thorax Size Ratio and Wing-Aspect Ratio in Drosophila melanogaster. *Evolution*, **52**, 1353.

Baker DA, Russell S (2009) Gene expression during Drosophila melanogaster egg development before and after reproductive diapause. *BMC Genomics*, **10**, 242.

Begun DJ, Holloway AK, Stevens K *et al.* (2007) Population Genomics: Whole-Genome Analysis of Polymorphism and Divergence in Drosophila simulans. *PLoS Biology*, **5**, e310.

Behrman EL, Watson SS, O'Brien KR, Heshel SM, Schmidt PS (2015) Seasonal life history adaptation in two species of *Drosophila*. *Journal of Evolutionary Biology*.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, **57**, 289–300.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA (2014) Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila (D Bolnick, Ed,). *PLoS Genetics*, **10**, e1004775.

Bergland AO, Tobler R, Gonzalez J, Schmidt P, Petrov D (2015) Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in Drosophila melanogaster. *bioRxiv*, 009084.

Berry A, Kreitman M (1993) Molecular Analysis of an Allozyme Cline: Alcohol Dehydrogenase in Drosophila Melanogaster on the East Coast of North America.

Genetics, **134**, 869–20130143.

Blanckenhorn WU, Fairbairn DJ (1995) Life history adaptation along a latitudinal cline in the water strider Aquarius remigis (Heteroptera: Gerridae). *Journal of Evolutionary Biology*, **8**, 21–41.

Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI (2002) Large clusters of co-expressed genes in the Drosophila genome. *Nature*, **420**, 666–669.

Böhni R, Riesgo-Escovar J, Oldham S *et al.* (1999) Autonomous control of cell and organ size by CHICO, a Drosophila homolog of vertebrate IRS1-4. *Cell*, **97**, 865–875.

Bradshaw WE, Holzapfel CM (2007) Evolution of Animal Photoperiodism. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 1–25.

Brown CR, Brown MB, Roche EA (2013) Fluctuating viability selection on morphology of cliff swallows is driven by climate. *Journal of Evolutionary Biology*, **26**, 1129–1142.

Bunning E (1960) Circadian Rhythms and the Time Measurement in Photoperiodism. *Cold Spring Harbor Symposia on Quantitative Biology*, **25**, 249–256.

Caicedo AL, Stinchcombe JR, Olsen KM, Schmitt J, Purugganan MD (2004) Epistatic interaction between Arabidopsis FRI and FLC flowering time genes generates a latitudinal cline in a life history trait. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 15670–15675.

Capy P, Pla E, David JR (1993) Phenotypic and genetic variability of morphometrical traits in natural populations of Drosophila melanogaster and D simulans. I. Geographic variations. *Genetics Selection Evolution*, **25**, 1–20.

Cingolani P, Platts A, Wang LL *et al.* (2012) A program for annotatin and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly*, **6**, 80–92.

Clancy DJ, Gems D, Harshman LG *et al.* (2001) Extension of life-span by loss of CHICO, a Drosophila insulin receptor substrate protein. *Science (New York, N.Y.)*, **292**, 104–106.

Clapham DH, Dormling I, Ekberg L *et al.* (1998) Latitudinal cline of requirement for far-red light for the photoperiodic control of budset and extension growth in Picea abies (Norway spruce). *Physiologia Plantarum*, **102**, 71–78.

Cogni R, Kuczynski C, Koury S *et al.* (2013) The intensity of selection acting on *couch potato* gene-spatial-temporal variation in a diapause cline. *Evolution*, **68**, 538–548.

Cogni R, Kuczynski K, Lavington E *et al.* (2014) Variation in Drosophila melanogaster central metabolic genes appears driven by natural selection both within and between populations. *Proceedings of the Royal Society of London B: Biological Sciences*, **282**, 20142688–20142688.

Cookson W, Liang L, Abecasis G, Moffatt M, Lathrop M (2009) Mapping complex disease traits with global gene expression. *Nature Reviews Genetics*, **10**, 184–194.

Craig TL, Denlinger DL (2000) Sequence and transcription patterns of 60S ribosomal protein P0, a diapause-regulated AP endonuclease in the flesh fly, Sarcophaga crassipalpis. *Gene*, **255**, 381–388.

David J, Capy P (1988) Genetic variation of Drosophila melanogaster natural populations. *Trends in Genetics*, **4**, 106–111.

Denlinger DL (2002) Regulation of diapause. *Annual review of entomology*, **47**, 93–122.

Ehrenreich IM, Hanzawa Y, Chou L *et al.* (2009) Candidate gene association mapping of Arabidopsis flowering time. *Genetics*, **183**, 325–335.

Elliott R, Li F, Dragomir I *et al.* (2013) Analysis of the Host Transcriptome from Demyelinating Spinal Cord of Murine Coronavirus-Infected Mice (S Jacobson, Ed,). *PLoS ONE*, **8**, e75346.

Emerson KJ, Uyemura AM, McDaniel KL *et al.* (2009) Environmental control of ovarian dormancy in natural populations of Drosophila melanogaster. *Journal of Comparative Physiology A*, **195**, 825–829.

Enattah NS, Sahi T, Savilahti E *et al.* (2002) Identification of a variant associated with adult-type hypolactasia. *Nature genetics*, **30**, 233–237.

Engelmann K, Purugganan M (2006) The Molecular Evolutionary Ecology of Plant Development: Flowering Time in Arabidopsis thaliana. In: *Developmental Genetics*

of the Flower Advances in Botanical Research. pp. 507–526. Elsevier.

Fabian DK, Kapun M, Nolte V *et al.* (2012) Genome-wide patterns of latitudinal differentiation among populations of Drosophila melanogaster from North America. *Molecular ecology*, **21**, 4748–4769.

Fabian DK, Lack JB, Mathur V *et al.* (2015) Spatially varying selection shape life history clines among populations of *Drosophila melanogaster* from sub-Saharan Africa. *Journal of Evolutionary Biology*.

Flannagan RD, Tammariello SP, Joplin KH *et al.* (1998) Diapause-specific gene expression in pupae of the flesh fly Sarcophaga crassipalpis. *Proceedings of the National Academy of Sciences*, **95**, 5616–5620.

Folguera G, Ceballos S, Spezzi L, Fanara JJ, Hasson E (2008) Clinal variation in developmental time and viability, and the response to thermal treatments in two species of Drosophila. *Biological Journal of the Linnean Society*, **95**, 233–245.

Frydenberg J, Hoffmann AA, Loeschcke V (2003) DNA sequence variation and latitudinal associations in hsp23, hsp26 and hsp27 from natural populations of Drosophila melanogaster. *Molecular ecology*, **12**, 2025–2032.

García Gil MR, Mikkonen M, Savolainen O (2003) Nucleotide diversity at two phytochrome loci along a latitudinal cline in Pinus sylvestris. *Molecular ecology*, **12**, 1195–1206.

Geiger-Thornsberry GL, Mackay TFC (2004) Quantitative trait loci affecting natural variation in Drosophila longevity. *Mechanisms of Ageing and Development*, **125**, 179–189.

Giannakou ME, Goss M, Jünger MA *et al.* (2004) Long-lived Drosophila with overexpressed dFOXO in adult fat body. *Science (New York, N.Y.)*, **305**, 361–361.

Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics*, **24**, 408–415.

Gillespie J (1973) Polymorphism in random environments. *Theoretical Population Biology*, **4**, 193–195.

Gillespie JH, Langley CH (1974) A General Model to Account for Enzyme Variation in

Natural Populations. *Genetics*, **76**, 837–143.

Goenaga J, Fanara JJ, Hasson E (2013) Latitudinal Variation in Starvation Resistance is Explained by Lipid Content in Natural Populations of Drosophila melanogaster. *Evolutionary Biology*, **40**, 601–612.

Goto SG (2013) Roles of circadian clock genes in insect photoperiodism. *Entomological Science*, **16**, 1–16.

Grant PR, Grant BR (2002) Unpredictable evolution in a 30-year study of Darwin's finches. *Science (New York, N.Y.)*, **296**, 707–711.

Grönke S, Clarke D-F, Broughton S, Andrews TD, Partridge L (2010) Molecular Evolution and Functional Characterization of Drosophila Insulin-Like Peptides (E Rulifson, Ed,). *PLoS Genetics*, **6**, e1000857.

Hairston NG Jr., Dillon TA (1990) Fluctuating Selection and Response in a Population of Freshwater Copepods. *Evolution*, **44**, 1796.

Haldane JBS, Jayakar SD (1963) Polymorphism due to selection of varying direction. *Journal of Genetics*, **58**, 237–242.

Hendry AP (2013) Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity*, **111**, 456–466.

Henrich VC, Denlinger DL (1982) A maternal effect that eliminates pupal diapause in progeny of the flesh fly, Sarcophaga bullata. *Journal of Insect Physiology*, **28**, 881–884.

Hoekstra HE, Drumm KE, Nachman MW (2004) Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. *Evolution*, **58**, 1329–1341.

Hoffmann AA, Harshman LG (1999) Desiccation and starvation resistance in Drosophila: patterns of variation at the species, population and intrapopulation levels. *Heredity*, **83**, 637–643.

Hoffmann AA, Anderson A, Hallas R (2002) Opposing clines for high and low temperature resistance in Drosophila melanogaster. *Ecology Letters*, **5**, 614–618.

Hsu H-J, LaFever L, Drummond-Barbosa D (2008) Diet controls normal and tumorous

germline stem cells via insulin-dependent and -independent mechanisms in Drosophila. *Developmental Biology*, **313**, 700–712.

Huang J-H, Jing X, Douglas AE (2015) The multi-tasking gut epithelium of insects. *Insect Biochemistry and Molecular Biology*.

Huang W, Massouras A, Inoue Y *et al.* (2014) Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines. *Genome Research*, **24**, 1193–1208.

Huey RB, Gilchrist GW, Carlson ML, Berrigan D, Serra L (2000) Rapid evolution of a geographic cline in size in an introduced fly. *Science (New York, N.Y.)*, **287**, 308–309.

Hwangbo DS, Gersham B, Tu M-P, Palmer M, Tatar M (2004) Drosophila dFOXO controls lifespan and regulates insulin signaling in brain and fat body. *Nature*, **429**, 562–566.

James AC, Azevedo R, Partridge L (1997) Genetic and Environmental Responses to Temperature of Drosophila Melanogaster from a Latitudinal Cline. *Genetics*, **146**, 881–20130143.

Janssens H, Gehring WJ (1999) Isolation and characterization of drosocrystallin, a lens crystallin gene of Drosophila melanogaster. *Developmental Biology*, **207**, 204–214.

Kalmykova AI, Nurminsky DI, Ryzhov DV, Shevelyov YY (2005) Regulated chromatin domain comprising cluster of co-expressed genes in Drosophila melanogaster. *Nucleic acids research*, **33**, 1435–1444.

Kao JY, Zubair A, Salomon MP, Nuzhdin SV, Campo D (2015) Population genomic analysis uncovers African and European admixture in Drosophila melanogaster populations from the south‐eastern United States and Caribbean Islands. *Molecular ecology*, **24**, 1499–1509.

Kapun M, Schalkwyk H, McAllister B, Flatt T, Schlotterer C (2014) Inference of chromosomal inversion dynamics from Pool‐Seq data in natural and laboratory populations of Drosophila melanogaster. *Molecular ecology*, **23**, 1813–1827.

Karan D, Dahiya N, Munjal AK *et al.* (1998) Desiccation and Starvation Tolerance of

Adult Drosophila: Opposite Latitudinal Clines in Natural Populations of Three Different Species. *Evolution*, **52**, 825.

Kimura KD, Tissenbaum HA, Liu Y, Ruvkun G (1997) daf-2, an insulin receptor-like gene that regulates longevity and diapause in Caenorhabditis elegans. *Science (New York, N.Y.)*, **277**, 942–946.

King EG, Sanderson BJ, McNeil CL, Long AD, Macdonald SJ (2014) Genetic Dissection of the Drosophila melanogaster Female Head Transcriptome Reveals Widespread Allelic Heterogeneity. *PLoS Genetics*, **10**, e1004322.

King RC (1970) *Ovarian development in Drosophila melanogaster*. Academic Press, New York.

Knibb WR (1982) Chromosome inversion polymorphisms in Drosophila melanogaster II. Geographic clines and climatic associations in Australasia, North America and Asia. *Genetica*, **58**, 213–221.

Koboldt DC, Zhang Q, Larson DE *et al.* (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*, **22**, 568–576.

Kofler R, Schlotterer C (2012) Gowinda: unbiased analysis of gene set enrichment for genome-wide association studies. *Bioinformatics*, **28**, 2084–2085.

Komori N, Usukura J, Matsumoto H (1992) Drosocrystallin, a major 52 kDa glycoprotein of the Drosophila melanogaster corneal lens. Purification, biochemical characterization, and subcellular localization. *Journal of cell science*, **102 (Pt 2)**, 191–201.

Korves TM, Schmid KJ, Caicedo AL *et al.* (2007) Fitness Effects Associated with the Major Flowering Time Gene FRIGIDA in Arabidopsis thaliana in the Field. *The American Naturalist*, **169**, E141–E157.

Kuraishi T, Binggeli O, Opota O, Buchon N, Lemaitre B (2011) Genetic evidence for a protective role of the peritrophic matrix against intestinal bacterial infection in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 15966–15971.

Lack JB, Cardeno CM, Crepeau MW *et al.* (2015) The Drosophila Genome Nexus: A Population Genomic Resource of 623 Drosophila melanogaster Genomes, Including 197 from a Single Ancestral Range Population. *Genetics*, **199**, genetics.115.174664–1241.

LaFever L, Drummond-Barbosa D (2005) Direct control of germline stem cell division and cyst growth by neural insulin in Drosophila. *Science (New York, N.Y.)*, **309**, 1071–1073.

Leather SR, Walters KFA, Bale JS (1995) *The Ecology of Insect Overwintering*. Cambridge University Press.

Lee K-S, Kwon O-Y, Lee JH *et al.* (2008) Drosophila short neuropeptide F signalling regulates growth by ERK-mediated insulin signalling. *Nature Cell Biology*, **10**, 468–475.

Lee SF, Eyre-Walker YC, Rane RV *et al.* (2013) Polymorphism in the neurofibromin gene, Nf1, is associated with antagonistic selection on wing size and development time in Drosophila melanogaster. *Molecular ecology*, **22**, 2716–2725.

Lehane MJ (1997) Peritrophic matrix structure and function. *Annual review of entomology*, **42**, 525–550.

Leng N, Dawson JA, Thomson JA *et al.* (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, **29**, 2073–2073.

Levine MT, Eckert ML, Begun DJ (2010) Whole-genome expression plasticity across tropical and temperate *Drosophila melanogaster* populations from eastern Australia. *Molecular Biology and Evolution*, **28**, 249–256.

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li JJ, Huang H, Bickel PJ, Brenner SE (2014) Comparison of D. melanogaster and C. elegans developmental stages, tissues, and cells by modENCODE RNA-seq data. *Genome Research*, **24**, 1086–1101.

Liu X, Lorenz L, Yu QN, Hall JC, Rosbash M (1988) Spatial and temporal expression of the period gene in Drosophila melanogaster. *Genes & Development*, **2**, 228–238.

Loeschcke V, Bundgaard J, Barker JS (2000) Variation in body size and life history traits in Drosophila aldrichi and D. buzzatii from a latitudinal cline in eastern Australia. *Heredity*, **85 Pt 5**, 423–433.

Mackay TFC (2001) Quantitative trait loci in Drosophila. *Nature Reviews Genetics*, **2**, 11–20.

Mackay TFC, Richards S, Stone EA *et al.* (2012) The Drosophila melanogaster Genetic Reference Panel. *Nature*, **482**, 173–178.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10.

Massouras A, Waszak SM, Albarca-Aguilera M *et al.* (2012) Genomic Variation and Its Impact on Gene Expression in Drosophila melanogaster (B Oliver, Ed,). *PLoS Genetics*, **8**, e1003055.

McKechnie SW, Blacket MJ, Song SV *et al.* (2010) A clinally varying promoter polymorphism associated with adaptive variation in wing size in *Drosophila*. *Molecular ecology*, **19**, 775–784.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Metaxakis A, Oehler S, Klinakis A, Savakis C (2005) Minos as a genetic and genomic tool in Drosophila melanogaster. *Genetics*, **171**, 571–581.

Meuti ME, Denlinger DL (2013) Evolutionary links between circadian clocks and photoperiodic diapause in insects. *Integrative and Comparative Biology*, **53**, 131–143.

Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics*, **91**, 243–248.

Mitrovski P, Hoffmann AA (2001) Postponed reproduction as an adaptation to winter conditions in Drosophila melanogaster: evidence for clinal variation under semi-

natural conditions. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 2163–2168.

Moran PAP (1950) Notes on Continuous Stochastic Phenomena. *Biometrika*, **37**, 17.

Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5268–5273.

Neumüller RA, Richter C, Fischer A *et al.* (2011) Genome-wide analysis of self-renewal in *Drosophila* neural stem cells by transcgenic RNAi. *Cell Stem Cell*, **8**, 580–593.

Nilsson D-E (1989) Optics and Evolution of the Compound Eye. In: *Facets of vision*, pp. 30–73. Springer Berlin Heidelberg, Berlin, Heidelberg.

Oku K, Yano S, Takafuji A (2003) Different maternal effects on diapause induction of tetranychid mites, Tetranychus urticae and T. kanzawai (Acari: Tetranychidae). *Applied Entomology and Zoology*, **38**, 267–270.

Oliver B, Parisi M, Clark D (2002) Gene expression neighborhoods. *Journal of Biology*, **1**, 4.

Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, **6**, 119–127.

Osley MA (1991) The Regulation of Histone Synthesis in the Cell Cycle. *Annual Review of Biochemistry*, **60**, 827–861.

Paaby AB, Schmidt PS (2008) Functional Significance of Allelic Variation at methuselah , an Aging Gene in Drosophila (MW Hahn, Ed,). *PLoS ONE*, **3**, e1987.

Paaby AB, Bergland AO, Behrman EL, Schmidt PS (2014) A highly pleiotropic amino acid polymorphism in the Drosophila insulin receptor contributes to life‐history adaptation. *Evolution*, **68**, 3395–3409.

Paaby AB, Blacket MJ, Hoffmann AA, Schmidt PS (2010) Identification of a candidate adaptive polymorphism for *Drosophila* life history by parallel independent clines on two continents. *Molecular ecology*, **19**, 760–774.

Palmieri N, Nolte V, Chen J, Schlotterer C (2015) Genome assembly and annotation of a Drosophila simulans strain from Madagascar. *Molecular Ecology Resources*, **15**,

372–381.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, **20**, 289–290.

Park SG, Hannenhalli S, Choi SS (2014) Conservation in first introns is positively associated with the number of exons within genes and the presence of regulatory epigenetic signals. *BMC Genomics*, **15**, 526.

Pegoraro M, Gesto JS, Kyriacou CP, Tauber E (2014) Role for Circadian Clock Genes in Seasonal Timing: Testing the Bünning Hypothesis (P Emery, Ed,). *PLoS Genetics*, **10**, e1004603.

Pittendrigh CS, Takamura T (1989) Latitudinal clines in the properties of a circadian pacemaker. *Journal of biological rhythms*, **4**, 217–235.

Pittendrigh CS, Kyner WT, Takamura T (1991) The Amplitude of Circadian Oscillations: Temperature Dependence, Latitudinal Clines, and the Photoperiodic Time Measurement. *Journal of biological rhythms*, **6**, 299–313.

Plautz JD, Kaneko M, Hall JC, Kay SA (1997) Independent photoreceptive circadian clocks throughout Drosophila. *Science (New York, N.Y.)*, **278**, 1632–1635.

Poelchau MF, Reynolds JA, Elsik CG, Denlinger DL, Armbruster PA (2013a) Deep sequencing reveals complex mechanisms of diapause preparation in the invasive mosquito, Aedes albopictus. *Proceedings. Biological sciences / The Royal Society*, **280**, 20130143–20130143.

Poelchau MF, Reynolds JA, Elsik CG, Denlinger DL, Armbruster PA (2013b) RNA-Seq reveals early distinctions and late convergence of gene expression between diapause and quiescence in the Asian tiger mosquito, Aedes albopictus. *The Journal of experimental biology*, **216**, 4082–4090.

Pool JE, Corbett-Detig RB, Sugino RP *et al.* (2012) Population Genomics of Sub-Saharan Drosophila melanogaster : African Diversity and Non-African Admixture (HS Malik, Ed,). *PLoS Genetics*, **8**, e1003080.

Quijano C, Tomancak P, Lopez-Marti J *et al.* (2008) Selective maintenance of Drosophila tandemly arranged duplicated genes during evolution. *Genome Biology*,

9, R176.

Ragland GJ, Denlinger DL, Hahn DA (2010) Mechanisms of suspended animation are revealed by transcript profiling of diapause in the flesh fly. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 14909–14914.

Rajpurohit S, Nedved O (2013) Clinal variation in fitness related traits in tropical drosophilids of the Indian subcontinent. *Journal of Thermal Biology*, **38**, 345–354.

Rajpurohit S, Nedved O, Gibbs AG (2013) Meta-analysis of geographical clines in desiccation tolerance of Indian drosophilids. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **164**, 391–398.

Rajpurohit S, Parkash R, Ramniwas S (2008a) Body melanization and its adaptive role in thermoregulation and tolerance against desiccating conditions in drosophilids. *Entomological Research*, **38**, 49–60.

Rajpurohit S, Parkash R, Ramniwas S, Singh S (2008b) Variations in body melanisation, ovariole number and fecundity in highland and lowland populations of Drosophila melanogaster from the Indian subcontinent. *Insect Science*, **15**, 553–561.

Reinhardt JA, Kolaczkowski B, Jones CD, Begun DJ, Kern AD (2014) Parallel Geographic Variation in Drosophila melanogaster. *Genetics*, **197**, genetics.114.161463–373.

Reynolds JA, Clark J, Diakoff SJ, Denlinger DL (2013) Transcriptional evidence for small RNA regulation of pupal diapause in the flesh fly, Sarcophaga bullata. *Insect Biochemistry and Molecular Biology*, **43**, 982–989.

Richard DS, Watkins NL, Serafin RB, Gilbert LI (1998) Ecdysteroids regulate yolk protein uptake by Drosophila melanogaster oocytes. *Journal of Insect Physiology*, **44**, 637–644.

Roberts A, Pachter L (2011) RNA-Seq and find: entering the RNA deep field. *Genome Medicine*, **3**, 74.

Robinson SJ, Zwaan B, Partridge L (2000) Starvation resistance and adult body composition in a latitudinal cline of Drosophila melanogaster. *Evolution*, **54**, 1819–1824.

Rockey SJ, Miller BB, Denlinger DL (1989) A diapause maternal effect in the flesh fly, Sarcophaga bullata: Transfer of information from mother to progeny. *Journal of Insect Physiology*, **35**, 553–558.

Rockman MV (2012) The QTN program adn the alleles that matter for evolution: all that's gold dos not glitter. *Evolution*, **66**, 1–17.

Rogers RL, Cridland JM, Shao L *et al.* (2014) Landscape of standing variation for tandem duplications in Drosophila yakuba and Drosophila simulans. *Molecular Biology and Evolution*, **31**, 1750–1766.

Rose AB (2008) Intron-Mediated Regulation of Gene Expression. In: *Nuclear pre-mRNA Processing in Plants* Current Topics in Microbiology and Immunology. pp. 277–290. Springer Berlin Heidelberg, Berlin, Heidelberg.

Sandrelli F, Tauber E, Pegoraro M *et al.* (2007) A molecular basis for natural selection at the timeless locus in Drosophila melanogaster. *Science (New York, N.Y.)*, **316**, 1898–1900.

Saunders DS (1987) Maternal influence on the incidence and duration of larval diapause in Calliphora vicina. *Physiological Entomology*, **12**, 331–338.

Saunders DS (2013) Insect photoperiodism: Measuring the night. *Journal of Insect Physiology*, **59**, 1–10.

Saunders DS (2014) Insect photoperiodism: effects of temperature on the induction of insect diapause and diverse roles for the circadian system in the photoperiodic response. *Entomological Science*, **17**, 25–40.

Saunders DS, Gilbert LI (1990) Regulation of ovarian diapause in Drosophila melanogaster by photoperiod and moderately low temperature. *Journal of Insect Physiology*, **36**, 195–200.

Saunders DS, Henrich VC, Gilbert LI (1989) Induction of diapause in Drosophila melanogaster: photoperiodic regulation and the impact of arrhythmic clock mutations on time measurement. *Proceedings of the National Academy of Sciences*, **86**, 3748–3752.

Schmidt P (2011) Evolution and mechanisms of insect reproductive diapause: A plastic

and pleiotropic life history syndrome. In: *The Genetics and Physiology of Life History Traits and Trade-Offs*, pp. 230–242. Oxford University Press.

Schmidt PS, Conde DR (2006) Environmental heterogeneity and the maintenance of genetic variation for reproductive diapause in Drosophila melanogaster. *Evolution*, **60**, 1602–1611.

Schmidt PS, Matzkin L, Ippolito M, Eanes WF (2005a) Geographic variation in diapause incidence, life-history traits, and climatic adaptation in Drosophila melanogaster. *Evolution*, **59**, 1721–1732.

Schmidt PS, Paaby AB, Heschel MS (2005b) Genetic variance for diapause expression and associated life histories in Drosophila melanogaster. *Evolution*, **59**, 2616–2625.

Schmidt PS, Zhu C-T, Das J *et al.* (2008) An amino acid polymorphism in the couch potato gene forms the basis for climatic adaptation in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 16207–16211.

Sezgin E, Duvernell DD, Matzkin LM *et al.* (2004) Single-locus latitudinal clines and their relationship to temperate adaptation in metabolic genes and derived alleles in Drosophila melanogaster. *Genetics*, **168**, 923–931.

Shapiro MD, Marks ME, Peichel CL *et al.* (2004) Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, **428**, 717–723.

Sim C, Denlinger DL (2013) Insulin signaling and the regulation of insect diapause. *Frontiers in Physiology*, **4**.

Spellman PT, Rubin GM (2002) Evidence for large domains of similarly expressed genes in the Drosophila genome. *Journal of Biology*, **1**, 5.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.

Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

Stinchcombe JR, Weinig C, Ungerer M *et al.* (2004) A latitudinal cline in flowering time

in Arabidopsis thaliana modulated by the flowering time gene FRIGIDA. *Proceedings of the National Academy of Sciences*, **101**, 4712–4717.

Swallow DM (2003) Genetics of lactase persistence and lactose intolerance. *Annual Review of Genetics*, **37**, 197–219.

Tarwater CE, Beissinger SR (2013) Opposing selection and environmental variation modify optimal timing of breeding. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 15365–15370.

Tatar M, Yin C (2001) Slow aging during insect reproductive diapause: why butterflies, grasshoppers and flies are like worms. *Experimental gerontology*, **36**, 723–738.

Tatar M, Chien SA, Priest NK (2001a) Negligible Senescence during Reproductive Dormancy in Drosophila melanogaster. *The American Naturalist*, **158**, 248–258.

Tatar M, Kopelman A, Epstein D *et al.* (2001b) A mutant Drosophila insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science (New York, N.Y.)*, **292**, 107–110.

Tauber E, Zordan M, Sandrelli F *et al.* (2007) Natural selection favors a newly derived timeless allele in Drosophila melanogaster. *Science (New York, N.Y.)*, **316**, 1895–1898.

Tauber MJ, Tauber CA, Masaki S (1986) *Seasonal Adaptations of Insects*. Oxford University Press.

Tishkoff SA, Reed FA, Ranciaro A *et al.* (2007) Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics*, **39**, 31–40.

Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics*, **179**, 455–473.

Umina PA, Weeks AR, Kearney MR, McKechnie SW, Hoffmann AA (2005) A rapid shift in a classic clinal pattern in Drosophila reflecting climate change. *Science (New York, N.Y.)*, **308**, 691–693.

Verrelli BC, Eanes WF (2001) Clinal variation for amino acid polymorphisms at the Pgm locus in Drosophila melanogaster. *Genetics*, **157**, 1649–1663.

Weinig C, Ungerer MC, Dorn LA *et al.* (2002) Novel loci control variation in

reproductive timing in Arabidopsis thaliana in natural environments. *Genetics*, **162**, 1875–1884.

Williams KD, Sokolowski MB (1993) Diapause in Drosophila melanogaster females: a genetic analysis. *Heredity*, **71**, 312–317.

Williams KD, Busto M, Suster ML *et al.* (2006) Natural variation in Drosophila melanogaster diapause due to the insulin-regulated PI3-kinase. *Proceedings of the National Academy of Sciences*, **103**, 15911–15915.

Wullschleger S, Loewith R, Hall MN (2006) TOR Signaling in Growth and Metabolism. *Cell*, **124**, 471–484.

Yamamoto R, Bai H, Dolezal AG, Amdam G, Tatar M (2013) Juvenile hormone regulation of Drosophila aging. *BMC Biology*, **11**, 85.

Yocum GD, Rinehart JP, Horvath DP *et al.* (2015) Key molecular processes of the diapause to post‐diapause quiescence transition in the alfalfa leafcutting bee Megachile rotundata identified by comparative transcriptome analysis. *Physiological Entomology*, **40**, 103–112.

Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, **11**, R14.

Zhang Q, Li H, Li R *et al.* (2008) Association of the circadian rhythmic expression of GmCRY1a with a latitudinal cline in photoperiodic flowering of soybean. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 21028–21033.

Zhao X, Bergland AO, Behrman EL *et al.* Global transcriptional profiling of diapuase adn climatic adaptation in Drosophila melanogaster. In revision at *Molecular Biology and Evolution*.