TRANSCRIPTOME PROFILING REVEALS MECHANISMS FOR THE EVOLUTION OF INSECT SEASONALITY

Crista B. Wadsworth* and Erik B. Dopman*
Department of Biology, Tufts University
200 Boston Ave, Suite 4700
Medford, MA, 02155 USA

*Corresponding Authors:
Crista Wadsworth (crista.wadsworth@tufts.edu; 585-354-2260) and Erik Dopman (erik.dopman@tufts.edu; 617-627-4890)

Keywords: temporal isolation, dormancy, phenology, developmental timing, Lepidoptera

SUMMARY STATEMENT

Evolutionary changes to the timing of diapause will be key for insects to adapt to a globally warming climate. Here, we identify putative loci responsible for timing shifts.

ABSTRACT

Rapid evolutionary change in seasonal timing can facilitate ecological speciation and resilience to climate warming. However, the molecular mechanisms behind shifts in animal seasonality are still unclear. Evolved differences in seasonality occur in the European corn borer moth (*Ostrinia nubilalis*), in which early summer emergence in E-strain adults and later summer emergence in Z-strain adults is explained by a shift in the length of the termination phase of larval diapause. Here, we sample from the developmental time course of diapause in both strains and use transcriptome sequencing to profile regulatory and amino acid changes associated with timing divergence. Within a previously defined QTL, we nominate 48 candidate genes including several in the insulin signaling and circadian rhythm pathways. Genome-wide transcriptional activity is negligible during the extended Z-strain termination, whereas shorter E-strain termination is characterized by a rapid burst of regulatory changes involved in resumption of the cell cycle, hormone production, and stress response. Although gene expression during diapause termination in *Ostrinia* is similar to that found previously in flies, nominated genes for shifts in timing are species-specific. Hence, across distant relatives the evolution of insect seasonality appears to involve unique genetic switches that direct organisms into distinct phases of the diapause pathway through wholesale restructuring of conserved gene regulatory networks.

Journal of Experimental Biology • Advance article

INTRODUCTION

Defined as a state of environmentally induced developmental arrest, dormancy is a physiologically dynamic developmental trajectory and is a nearly ubiquitous life-history strategy for animals and plants to survive inhospitable climatic conditions in temperate and polar environments. Diapause, a widespread form of insect dormancy, is also viewed as the primary mechanism through which the annual rhythm of major life history phases such as reproduction, growth, development, and migration are synchronized with seasonally varying biotic and abiotic requirements (Danilevskii, 1965; Tauber and Tauber, 1981; Danks, 1987; Leopold and Lang, 1996). Diapause occurs at various life-stages in insects (egg, larval, pupal, adult), and proceeds through pre-diapause, diapause, and post-diapause developmental phases (Koštál, 2006**)**. Diapause itself is further decomposed into an initiation phase in which direct development ceases, a maintenance phase in which developmental arrest continues, and a termination phase in which a stimulus triggers an active potential for development although development may not be overtly visible (Tauber et al., 1986; Koštál, 2006).

Species and populations can reveal subtle shifts in the timing of the unique stages of diapause that can substantially alter annual life-history patterns. Whereas the onset of diapause commonly initiates seasonal migration or polyphenism at appropriate times, the termination of diapause synchronizes active growth and reproduction with favorable conditions (Tauber et al., 1986). Examples of rapid shifts in the timing of the onset or termination of diapause are common, with many believed to be adaptive responses to a range of contemporary environmental perturbations including changing climates, anthropogenic impacts, and species introductions (Bradshaw and Holzapfel, 2001, 2006, 2008; Bradshaw et al., 2004; Filchak et al., 2000; Mathias et al., 2005; Schmidt et al., 2005; Gomi et al., 2007).

To understand how organisms adapt and persist in novel seasonal environments like those experienced during global warming, we must understand the molecular bases for shifts in diapause timing. In several study systems, different populations or species show shifts to their seasonal dormancy timing (Picket and Neary, 1940; Tauber and Tauber, 1976; Bradshaw and Lounibos, 1977; Glover et al., 1992). However, complications in these systems arising from genetic architecture (e.g., polygenic, epistasis) or genomic architecture (e.g., chromosomal rearrangements), have confounded the identification of genes underlying variation in diapause timing (Tauber et al., 1977; Feder et al., 2002; Bradshaw, 2005; Mathias et al., 2007; Emerson et al., 2010; Wadsworth et al., 2015). Despite these difficulties, genes involved in the transitions of diapause phases have been identified. Cell cycling, DNA replication/transcription, and stress response genes have been nominated as important for

Journal of Experimental Biology • Advance article

diapause initiation in moths and mosquitos (Bin Bao and Xu, 2011; Poelchau et al., 2011). Additionally, genes in the Wnt and TOR signaling pathways have been nominated as strong regulatory candidates during diapause termination in *Rhagoletis* flies (Ragland et al., 2011). Perhaps the best single candidate thus far was identified for the egg diapause of the silkmoth (*Bombyx mori*), in which transgenerational diapause induction involved the temperature-sensitive *transient receptor potential A1* (*Trpa1*) gene (Sato et al., 2014). Nevertheless, in many of these systems it is still unclear how the regulation of diapause genes and pathways are altered to produce ecologically relevant variation in diapause timing.

A useful model for studying the molecular basis of shifts in seasonality is the European corn borer moth (*Ostrinia nubilalis*). Introduced to North American from Europe in the early 20th century, multiple corn borer ecotypes exist that differ by 20-30 days in the time needed for overwintering larvae in diapause to reinitiate development in spring (Glover et al., 1992; Dopman et al., 2005). Over a half century of research has connected variation in diapause timing to latitudinal differences in generation number (Beck and Apple, 1961; Palmer et al., 1985; Levy et al., 2015). More recently, differences in diapause timing have been emphasized because they contribute to asynchronous adult mating flights of two-generation E-strain borers and one-generation Z-strain borers in New York (Dopman et al., 2010). Thus, divergence in the timing of diapause appears relevant for both successful colonization and ecological speciation.

Two observations provide important insight about possible loci responsible for diapause timing change in European corn borers. First, the a major genetic factor (*Pdd*) has been shown to map to a region of the Z (sex) chromosome that may be rearranged between corn borer strains (Glover et al., 1992; Dopman et al., 2004, 2005; Wadsworth et al., 2015). Second, high temperatures and/or long-day photoperiod cues are known to stimulate diapause termination in corn borers (McLeod and Beck, 1963). Thus, pathways and genes on the sex chromosome that are sensitive to light and temperature represent strong candidate genes for *Pdd*.

In a prior study, metabolic rate trajectories were used to identify a shift in the timing of the end of the diapause termination phase as critical for explaining earlier emergence in the E-strain and later emergence in the Z-strain of the European corn borer moth (Fig. 1A) (Wadsworth et al., 2013). Within a week of exposure to diapause-breaking (long-day) conditions, E-strain moths rapidly progress through the diapause termination phase and enter post-diapause development, as indicated by increased respiratory metabolism. In contrast, Z-strain borers exhibit continued suppression of metabolic rate for nearly a month longer,

suggesting that timing divergence stems from an extended diapause termination phase. Here, we sample from the developmental time course of diapause in both strains and we profile regulatory and amino acid changes coincident with changes in timing. We then address two main questions. First, what candidate genes might underlie *Pdd* and therefore divergence in diapause timing? Second, what downstream pathways might direct insects to an earlier or later end of diapause?

RESULTS

Transcriptome. For Z and E-strain corn borers, we sampled three time points along the photoperiodically-initiated and terminated diapause developmental pathway (Fig. 1B). Larval diapause in corn borers is initiated under short-day photoperiod, even under temperatures that are permissive for development (MeLeod and Beck, 1963; Glover et al., 1992). Therefore, the first sampled time point occurred when both strains were in the diapause maintenance phase under short-day diapause inducing conditions (12:12 L/D). Exposure to long-day conditions stimulates corn borers to enter diapause termination and downstream pupal development (MeLeod and Beck, 1963; Glover et al., 1992). Therefore, the second sampled time point occurred on day 1 after long-day exposure (16:8 L/D), but before a physiological change in respiratory metabolism in either strain (hereafter, "day 1"). The final time point was on day 7 after long-day exposure and after physiological divergence between strains (hereafter, "day 7"). Within this sampling scheme, we expected changes on day 1 to be more reflective of strain specific differences in upstream causal factors. Day 7 in contrast might be more informative of downstream developmental consequences of *Pdd* such as the transition from diapause termination to post-diapause development (as indicated by a metabolic uptick, Fig. 1A). Nevertheless, within-strain variation in the length of the diapause termination phase exists (Wadsworth et al., 2013), and therefore patterns of gene expression on day 1 and day 7 likely represent a mixture of cause and effect.

As the most likely tissues involved in photoperiodic responses have been hypothesized to be located in the brain, eyes, or major neuroendocrine glands (Koštál, 2011), at each of the three time points we sampled larval head tissues (n=5 individuals for n=5 replicates per strain). A total of ~562 million 50-bp single-end reads were generated across 30 libraries. Each library had on average 19.46 ± 8.99 million reads. After removal of adapters, mtDNA, and rRNA contaminants, average library size decreased to 18.38 ± 8.24 million reads. Our *de novo* assembly consisted of 47,565 transcripts (mean length=686, N50=1,025).

Of these, 14,177 were annotated via TBLASTX to FlyBase and 23,077 were annotated via TBLASTX to the NCBI non-redundant (nr) database.

There were a total of 15,208 transcripts that were significantly differentially expressed in at least one comparison involving strains or time-points (Table 1). We removed some transcripts from our analysis to minimize the number of genes that were differentially expressed between strains because of population divergence (strains diverged approximately 100k years ago, Malausa et al., 2007) and not because of associations with differences in the timing of diapause termination. With E and Z-strain corn borers being physiologically indistinguishable under short-day diapause inducing conditions (Wadsworth et al., 2013), we considered differentially expressed transcripts between strains at this time-point as a control for regulatory divergence unrelated to photoperiodic response or divergence in length of the diapause termination phase. We therefore removed these transcripts from differential expression comparisons between strains at other time points (day 1 and day 7) to help control for false positives.

Gene Candidates. To nominate transcripts as candidates for the 30-day shift in the end of diapause, we initially considered genomic position. *Pdd* maps to a region of the European corn borer Z chromosome between *Ket* and *Ldh* (Dopman et al., 2005). Therefore, an initial candidate list was nominated by TBLASTX hits of corn borer transcripts to the macrosyntenic interval of the *Bombyx mori* Z chromosome containing these genes (6.5 Mb-17.4 Mb) (Kroemer et al., 2011). Of the corn borer transcripts that mapped to this interval, two forms of evidence were evaluated. First, the causal factor(s) might be regulatory. Therefore, patterns of differential expression either within strain from diapause maintenance to a later developmental time-point, or between strain on day 1 or day 7, were considered. Second, as differences in diapause timing could be due to structural mutation(s), we considered the presence of fixed amino-acid changing substitutions between strains.

Of the 563 (~1%) transcripts with putative locations between *Ket* and *Ldh*, a total of 48 were nominated as candidates for *Pdd*. 41 were nominated based on patterns of significant differential expression between strains, or significant differential expression within strains across the developmental time course (Fig. 2; Table 2). Six broad patterns of expression differences were observed between strains: (A) 5 genes were upregulated in the Z-strain on day 1 relative to the E-strain, (B) 3 genes were upregulated in the Z-strain on day 7 relative to the E-strain, (C) 3 genes were upregulated in the Z-strain on day 1 and day 7 relative to the E-strain, (D) 10 genes were upregulated in the E-strain on day 1 relative to the Z-strain, (E) 8 genes were upregulated in the E-strain on day 7 relative to the Z-strain, and (F) 4 genes were

upregulated in the E-strain on days 1 and 7 relative to the Z-strain. Within the E-strain there were two observed expression patterns across the time course: (G) 4 downregulated genes from diapause maintenance to day 1, and (H) 1 upregulated gene from diapause maintenance to day 7. Of the transcripts nominated by regulatory patterns, all had some associated annotation information (Table 2).

We next considered candidates in the *Ket—Ldh* interval that harbored fixed amino-acid changes between Z and E-strains. Of the total number of variable SNPs between strains (n=1,347,090), 250 were located in European corn borer transcripts with significant TBLASTX hits to the *Ket—Ldh* interval in *B. mori*. These were distributed among 114 unique transcripts, 11 of which contained at least one predicted amino acid change that was fixed between strains. Eight of the 11 transcripts had annotation information (Table 3). Four amino acid changes were switches in polarity, two were switches in polarity and charge, and one was a switch in charge.

Global Transcriptional Patterns. To characterize regulatory mechanisms that might represent consequences of differences at *Pdd*, we focused on differential expression of transcripts between strains after exposure to diapause-breaking cues (Table 1). For Gene Ontology (GO) and KEGG enrichment (*P*-value cutoff ≤ 0.001) of differentially expressed gene lists, on day 1 there were 14 enriched "biological process" GO terms, 4 "molecular function" terms, and 5 KEGG pathways (Fig. 3; supplementary material Table S1). Two terms associated with Wnt signaling were also enriched ($P \leq 0.05$). Between strains on day 7 there were 24 enriched "biological process" terms, 9 "molecular function" terms, and 1 KEGG pathway (Fig. 3; supplementary material Table S1).

To identify pathways under the control of similar regulatory mechanisms, for each strain we evaluated correlated changes in gene expression across the three time points. Normalized expression data from the diapause maintenance time point were randomly coupled to later developmental time points and then $\log_2$ normalized to simulate microarray M values. Short Time-Series Expression Miner (STEM) (Ernst et al., 2005) was then used to discover 5 significant gene expression profiles in the E-strain and 2 in the Z-strain ($P \leq 0.01$) (Fig. 4). Pattern A had 105 transcripts assigned to it, pattern B had 109, pattern C had 108, pattern D had 67, pattern E had 64, pattern F had 38, and patterns G had 32. The most extreme expression changes were seen from diapause maintenance to day 1. Groups of genes with similar expression patterns across the time course in the E-strain were enriched for GO categories including development, metabolism, protein catabolism, and oxidoreductase activity (supplementary material Table S2). In contrast, there were no significantly enriched

GO terms for the two temporal expression profiles in the Z-strain. Time series expression profiles were also plotted for individual transcripts in pathways of interest. Pathways included cell cycling, ecdysone, circadian rhythms, heat shock, Wnt signaling, and insulin signaling (Fig. 5; supplementary material Fig. S1). At least one transcript within each of these pathways showed significant differential expression between the strains (supplementary material Table S3).

DISCUSSION

In insects, divergence in seasonal life-cycle timing may commonly involve changes in the duration of the diapause termination phase in spring. Z and E-strain European corn borers differ by ~30 days in the length of the diapause termination phase as indicated by a shift in the timing of metabolic release coincident with the end of diapause (Wadsworth et al., 2013) (Fig. 1A). Although strains are physiologically indistinguishable in early stages of diapause termination, RNA sequencing shows this to be a period of intense transcriptional activity in the E-strain, as indicated by a large number of differentially expressed transcripts from diapause maintenance to day 1 (Table 1). Conversely, the Z-strain shows negligible change from diapause maintenance to day 1 or 7 (Table 1). Thus, the major genetic cause of altered seasonal diapause timing, *Pdd,* appears to act as a genetic switch operating shortly after the beginning of the growing season in the E-strain, allowing for rapid development, while being delayed in the Z-strain.

Gene Candidates. Specific gene candidates for shifts in insect seasonal timing have been difficult to accumulate via QTL mapping due to complex genetic bases or genomic architectures between populations (Tauber and Tauber, 1977; Feder et al., 2002; Bradshaw et al., 2005; Mathias et al., 2007; Wadsworth et al., 2015). In such systems, transcriptome profiling can be a useful alternative to QTL mapping to nominate candidate genes. For example, in the pitcher-plant mosquito (*Wyeomyia smithii*), Northern and Southern populations differ in the critical day length required for diapause termination, but the genetic basis is complex involving many loci and epistasis (Bradshaw et al., 2005; Mathias et al., 2007). Using microarrays, a total of 29 genes were nominated as candidates with one, a cuticular protein with cornea-specific expression, falling within a previously defined QTL for altered termination response (Emerson et al., 2010).

In European corn borers, early results implicated a major Mendelian factor underlying differences in diapause termination timing (Glover et al., 1992; Dopman et al., 2004, 2005). However, recent findings show that *Pdd* and roughly 20% of the sex chromosome may be

caught up in a large inversion (Wadsworth et al., 2015). As inversions can cause large swaths of genes to be inherited together and thus act as "supergenes", there is potential for *Pdd* to be composed of multiple linked loci. Of the genes that are predicted to be within the chromosomal interval containing *Pdd*, 48 have fixed amino acid changes between strains or experience a change in expression through diapause termination (Table 2, 3). Candidate gene lists did not overlap between *W. smithii* and European corn borers (Emerson et al., 2010). Within our nominated gene lists, transcripts involved in insulin signaling and circadian rhythms are some of the most promising candidates for *Pdd* because both pathways are sensitive to light and temperature (Suttie et al., 1991; Chan et al., 1999; Shingleton et al., 2005; Taylor et al., 2005; Luckenbach et al., 2007; Sim and Denlinger, 2009), environmental factors that are known to stimulate diapause termination in corn borers (McLeod and Beck, 1963; Glover et al., 1992). These two pathways are also proposed regulators of dormancy or diapause in other species (Ikeno et al., 2010; Hahn and Denlinger, 2011). Of the 48 candidates, we focus on four with direct involvement in these pathways.

The γ subunit (*Snf4Aγ,* 15.6 Mb in *B. mori*) of AMP-activated protein kinase (AMPK) is a highly conserved gene that makes up the energy-sensing component of the AMPK protein (Kahn et al., 2005). AMPK is a member of the insulin, target of rapamycin (TOR), and cell death pathways (Kahn et al., 2005; Lippai et al., 2008), all of which are known to be important in the progression of development and molting in Lepidoptera (Lippai et al., 2008; Gu et al., 2009; Smith et al., 2014). In our transcriptome, *Snf4Aγ* was significantly downregulated in the later-emerging Z-strain compared to the earlier-emerging E-strain (FDR < 0.001 on day 7) (Table 2; Fig. 2). In *Drosophila*, loss of *Snf4Aγ* expression has been shown to inhibit the ability of 20-hydroxyecdysone (20HE) to promote cell death, a major component in metamorphosis in insects (Lippai et al., 2008). Thus, perhaps suppression of *Snf4Aγ* in Z-strain moths promotes a state of delayed development through inhibition of cell death and turnover, processes that occur ~20-30 days earlier in the E-strain.

Two members of the insulin growth factor (IGF) pathway were identified as candidate genes. The first, *IGF-II mRNA binding protein* (*Imp*, 11.4 Mb in *B. mori*), is associated with growth and development, and like the insulin signaling pathway, the IGF pathway commonly shows sensitivity to both temperature and photoperiod (Suttie et al., 1991; Dahl et al., 1997; Luckenbach et al., 2007). *Imp* was found to be downregulated (day 1 and 7) in the later-emerging Z-strain larvae (FDR < 0.001) (Table 2; Fig. 2). This result is notable because suppression of the IMP homolog in mice prevents IMP from binding to insulin growth factors, resulting in a reduction in growth (Hansen et al., 2004). In both mice and nematodes,

disruption of other genes within the IGF pathway result in phenotypes reminiscent of longer dormancy such as growth reduction and extend lifespan (Liu et al., 1993; Kenyon, 2010). Hence, downregulation of *Imp* in corn borers may promote later emergence from diapause in the Z-strain through decreased IGF signaling.

The second IGF candidate, *Receptor for activated C kinase 1* (*Rack1*, 10.9 Mb in *B. mori*), interacts with the IGF pathway through the insulin-like growth factor I receptor (IGF-IR). In corn borers, *Rack1* was upregulated in the later-emerging Z-strain on day 7 (FDR < 0.001) (Table 2; Fig. 2). *Rack1* shows a similar association with diapause maintenance in the Asian tiger mosquito (*Aedes albopictus*), where it was one of 314 transcripts to be upregulated upon exposure to short-day conditions and entrance into diapause (Poelchau et al., 2011). Overexpression of RACK1 protein has also been shown to lead to reduced cell growth in mammals (Hermanto et al., 2002). With the upregulation of *Rack1* in the Z-strain, and known upregulation in other species delaying growth or promoting diapause, this gene is an interesting candidate for *Pdd*.

A final gene candidate for *Pdd* is the clock gene *Period* (*Per*, 13.0 Mb in *B. mori*). *Per* shows a predicted glycine to arginine amino acid change between Z and E strain insects (Table 3) and is modestly upregulated in the E-strain on both day 1 and day 7 (FDR = 0.58 and 0.89 respectively) (Figs 5A, 5B). Evidence is accumulating that clock genes like *Per* are important for photoperiodic responses and diapause transitions. Within the Z-strain of corn borers, a different amino-acid substitution in *Per* varies with diapause emergence timing and latitudinal changes in generation number (Levy et al., 2015). In *D. melanogaster*, geographic variation in a threonine-glycine repeat within *Per* associates with diapause incidence (Kyriacou et al., 2008). Laboratory studies also support a role for *Per* in diapause. RNA interference of *Per* in the bean bug (*Riptortus pedestris*) disrupts both circadian rhythms and photoperiodic response (Ikeno et al., 2010), and artificial selection for enhanced diapause in grey flesh fly (*Sarcophaga bullata*) is associated with a deletion at the *Per* locus that is 33 amino acids long (Han and Denlinger, 2009).

Global Transcriptional Patterns. Our results show that the diapause termination phase and the transition into post-diapause development involves stress response, hormone processing, cell cycling, developmental patterning, and metabolic pathways (Figs 3, 4). Similar results were found in flies (Emerson et al., 2010; Poelchau et al., 2011; Ragland et al., 2010, 2011). This speaks to a conserved transcriptional basis for the end of the diapause developmental pathway, despite ~200 million years of divergence (Pringle et al., 2007), and distinct diapausing life-stages (larval *vs*. pupal) in different environments (host plant *vs*. soil).

Release From Developmental Arrest. Previous results in flies suggest that a common characteristic of diapause is cell-cycle arrest in the G0/G1 or G2 phases, with transition into cell cycle progression being indicative of the end of diapause (Koštál et al., 2009; Ragland et al., 2010, 2011). Gene expression profiles in corn borers support a transition into active cell cycling upon exposure to long-day cues in the E-strain but not the Z-strain (Figs 5C, 5D), suggesting E-strain borers are rapidly exiting diapause. For example, *Proliferating cell nuclear antigen* (*Pcna*), a DNA replication related protein that is active in the S phase of the cell cycle, is upregulated during the rapid diapause termination phase in E-strain European corn borers and also diapause termination of *Rhagoletis* flies (Ragland et al., 2011), but it is downregulated in Z-strain borers (FDR < 0.00001 on day 1). Differentially expressed genes between Z and E-strains were highly enriched for cell cycling GO terms (Fig. 3; supplementary material Table S1). Additionally, genes involved in the progression of the cell cycle were upregulated in the E-strain, whereas they were downregulated in Z-strain (e.g., *CycA*, *CycB* [FDR < 0.001 days 1 and 7]; *Cdk4*, *Polo* [FDR < 0.01 on day 1]). Overall, patterns suggest that active cell cycling occurs within days of exposure to long-day conditions in the E-strain allowing for release from diapause-induced metabolic depression and progression into post-diapause development, whereas cellular development remains suppressed for ~30 additional days in the Z-strain.

Transcripts involved in development showed congruent patterns to those involved in mitotic progression. Wnt signaling pathways promote proper developmental patterning in insects (e.g., metamorphosis) through cellular communication, resumption of the cell cycle, and also may regulate growth through the canonical growth-regulatory pathways (Logan and Nusse, 2004; Gokhale and Shingleton, 2015). The Wnt signaling and canonical Wnt signaling pathways showed enrichment of GO terms on day 1 between strains (Fig. 3; supplementary material Table S1). Many transcripts involved in these pathways were downregulated in the Z-strain and upregulated in the E-strain (e.g., *Fz3*, *Arm*, *Smo*, *Wg*, and *Pan* [FDR < 0.01 on day 1]) (supplementary material Figs S1A, S1B; Table S3). Coupled with the regulation of members of the cell cycling pathway, upregulation of members of Wnt signaling pathways in the E-strain indicate a reversal of arrest on both the cellular and developmental levels.

Stress Response. Studies in flies have documented shifts from anaerobic to aerobic metabolism during the diapause termination phase as the hypoxic stress of winter is lifted (Emerson et al., 2010; Ragland et al., 2010, 2011). In moths, we find evidence for an equivalent shift in the E-strain but not in the Z-strain. Members of the anaerobic pyruvate metabolic pathway experienced a ~2-fold decrease in expression within the E-strain from

diapause maintenance to day 1 or day 7 (e.g., *Aldh, CG6084*, *CG9629*, *Men, CG31674*, *Hex-t2* [FDR < 0.001]), and the pathway was upregulated in the Z-strain compared to the E-strain during diapause termination (e.g., *Aldh, CG9629, CG31075* [FDR < 0.01 on day 1]). These results suggest that a switch to aerobic metabolism is not associated with the diapause termination phase in moths *per se*, but is primarily associated with the metabolic uptick that marks the end of diapause termination phase.

Diapause is commonly associated with cold-hardiness across many insect species, yet it is unclear if the cold-hardy state persists through all of the phases of diapause. Our results suggest that in moths the reduction in cold-hardiness out of winter is primarily a function of a change in photoperiod (short to long-days) and entrance into diapause termination phase rather than the end of diapause. One class of proteins associated with cold-hardiness and diapause are heat shock proteins (HSPs) that act as molecular chaperones (Rinehart et al., 2007; Koštál and Tollarová-Borovanská, 2009; Ragland et al., 2010, 2011). Proteins within the HSP70 family are essential to recovery after chill shock treatments with the silencing of these proteins increasing mortality in flesh flies and linden bugs (Rinehart et al., 2007; Koštál and Tollarová-Borovanská, 2009). Our results show significant downregulation of genes in this family within the E-strain, both in the diapause termination phase and at the end of diapause (e.g., *Hsp70AB*, *Hsp70BC* [FDR < 0.01 on day 1 and 7]). Although not significant, a similar pattern is seen for *Hsp70BC* within the Z-strain, even though they remain in the diapause termination phase for much longer (Figs 5E, 5F). Consistent downregulation of HSPs in both strains prior to the end of diapause implies that cold-hardiness is a property of diapause maintenance in moths, rather than the entire diapause developmental pathway.

Encocrine. Once hormones are released from neurosecretory cells in the brain, downstream development and metamorphosis become inevitable. One important hormone, 20-hydroxyecdysone (20HE), plays a role in inducing insect molting and metamorphosis at appropriate developmental stages (Denlinger, 2002). 20HE commonly breaks diapause in many insect species when injected, fed, or topically applied (Žďárek and Denlinger, 1975; Gadenne et al., 1990; Kidokoro et al., 2006; Yamamoto et al., 2008). Therefore, extension of the diapause termination phase should be correlated with suppressed transcription of ecdysone-related genes through decreased availability of 20HE or the inability to bind it. Indeed, *Neverland* (*Nvd*), *Spook* (*Spo*), and *Phantom* (*Phm*) are involved in 20HE production and were upregulated in the E-strain compared to the Z-strain (e.g., *Nvd* [FDR < 0.01 day 1 and 7]; *Spo*, *Phm* [FDR < 0.0001 on day 1 and 7]) (supplementary material Figs S1C, S1D; Table S3). Loss of function of *Nvd* is known to cause developmental arrest and reduced

growth in *D. melanogaster* through reduced ecdysteroid production (Yoshiyama et al., 2006). 20HE receptors were also upregulated in the E-strain (e.g., *EcR* [FDR < 0.01 on day 1]; *Ftz-f1* [FDR < 0.01 on day 7]). Surprisingly, one upregulated receptor, *Ftz transcription factor 1* (*Ftz-f1*), controls stage-specific responses to 20HE (Broadus et al., 1999) and was located within the QTL for *Pdd* (Table 2).

The Circadian Clock. It has long been hypothesized that the circadian clock is intimately connected to photoperiodic responses like diapause (Bünning, 1936). Supporting this idea, many latitudinal clines show co-variation between clock genes and diapause response (Mathias et al., 2005; Schmidt et al., 2008; Tauber et al., 2007; Cogni et al., 2013; Levy et al., 2015). In addition to the possible involvement of *Per*, we found the circadian gene *PAR-domain protein 1* (*Pdp1*) to be significantly differently expressed between the strains (e.g., FDR < 0.01 on day 1) (Figs 5A, 5B; supplementary material Table S3). However, although predicted to be Z-linked, *Pdp1* is not located in the QTL (21.9 Mb in *B. mori*). In *Drosophila*, *Pdp1* encodes a transcription factor that promotes the transcription of the gene *Clock* (*Clk*) (Cyran et al., 2003). In the apple maggot fly, *Pdp1* is significantly downregulated from diapause maintenance to the diapause termination phase (Ragland et al., 2011). Similarly, in the E-strain *Pdp1* is downregulated from diapause maintenance to day 1 (Fig. 5A). Although the functional significance of this expression change is unclear, conserved expression of this gene between species may be indicative of an important role for *Pdp1* in the transition to the end of diapause.

Conclusions. Although diapause is deployed at various life stages in insect taxa, studies in flies and moths now point to deep conservation of transcriptional pathways during the phases of diapause (Fig. 3; supplementary material Table S3) (Emerson et al., 2010; Ragland et al., 2010, 2011). However, there appears to be little evidence in these organisms for shared genetic control of natural variation in diapause timing (Table 2, 3) (Emerson et al., 2010). The situation might be different within the Lepidoptera, such as *Ostrinia* moths and swallowtail butterflies, in which variation in diapause termination timing is associated with a common region of the sex chromosome (Hagen and Scriber, 1989; Ording et al., 2010; Dopman et al., 2005; Wadsworth et al., 2015). Hence, independent physiological and genetic mechanisms appear to govern the evolution of diapause timing across distantly related taxa, whereas a common origin may characterize more closely related lineages.

Elucidating the regulatory mechanisms of diapause evolution has long been challenging because of a lack of anatomical landmarks and a complex genetic basis. Our results demonstrate the combined power of high-throughput transcriptome and physiological

screens for understanding the evolution of diapause. We anticipate that expanding this framework to other diapause phases (i.e., diapause induction and maintenance), taxonomic groups, and at various stages of evolutionary divergence will provide a foundation to understand how animals adapt to climate change and synchronize major life-history events with seasonally varying environments.

MATERIALS AND METHODS

Sample preparation. Diapause was induced in larvae using short-day conditions (12:12 LD at 23°C). After 35 days of exposure to these conditions, diapausing borers remain at the 5th instar whereas direct developing insects will have already pupated or enclosed as adults (Glover et al., 1991; Dopman et al., 2005; Wadsworth et al., 2013). Diapause was terminated in 36 day old larvae using long-day conditions (16:8 LD at 26°C) (Glover et al., 1991; Dopman et al., 2005; Wadsworth et al., 2013). Alternative developmental genetic pathways underlying differences in diapause termination timing seem to be triggered by *Pdd* within the first 7 days after insects experience diapause-breaking cues, as indicated by elevated respiratory metabolism in the E-strain but not the Z-strain (Wadsworth et al., 2013). Therefore, we collected tissues at three time points from both strains within this interval (Fig. 1B): in diapause maintenance prior to a long-day (diapause breaking) cue, 1 day after transfer to long-day conditions when strains are physiologically indistinguishable but have entered the diapause termination phase, and on day 7 after transfer when the Z-strain remains in diapause termination (previously referred to as "diapause maintenance" in Wadsworth et al., 2013) and most E-strain moths have entered post-diapause development. Head tissues were sampled to capture changes in major neuroendocrine glands and the primary endogenous clock (Tauber et al., 1986; Koštál, 2011). Five individuals were pooled for each of 5 replicates per strain per time point. All samples were collected at the same time of day (14:00-16:00 hrs.) to avoid spurious effects of circadian rhythmicity. Tissues were ground in liquid nitrogen and total RNA was prepared for multiplexed Illumina sequencing using the Qiagen RNeasy kit (Qiagen Inc., Germantown, MD) followed by the TruSeq RNA Prep kit (Illumina Inc., San Diego, CA). Six libraries were multiplexed per lane and subjected to single-end 50 bp sequencing on an Illumina HiSeq 2000 (Illumina Inc., San Diego, CA) at Tufts University Core Facility.

Data processing. Low quality reads and adapter sequences were removed using Trimmomatic v.0.27 (Lohse et al., 2012). Retained reads were ≥ 36bp after eliminating reads with an average quality of < 15, and trimming off ends with a quality of < 5 in leading and

trailing bases. Mitochondrial (NC_003367.1) and ribosomal contaminants (all *Ostrinia* ribosomal sequences in NCBI) were removed using Bowtie2 v.2.1.0 (Langmead and Salzberg, 2012). Library quality was subsequently assessed using FastQC v.0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Prior to *de novo* assembly, duplicate reads were collapsed to reduce library complexity using FASTX-Toolkit v.0.0.13 (Gordon and Hannon, 2010). Trinity was chosen for *de novo* assembly because it has been shown to recover more full length genes than other single-kmer *de novo* assembly programs (Grabherr et al., 2011; Zhao et al., 2011). It was run at default parameters (kmer=25, minimum contig length=48) on all libraries and the longest transcript per transcript cluster was used for subsequent analyses. Annotation of genes was by TBLASTX search to FlyBase and the NCBI nr database and retention of transcripts with e-values $\leq$ $10^{-}5$. Putative locations for corn borer transcripts was determined using TBLASTX (E-value $\leq 10^{-5}$) and the genome of *B. mori* (KAIKObase v.3.2.2) (Shimomura et al., 2009), which shows conserved macrosynteny with *O. nubilalis* (Kroemer et al., 2011). The *Pdd* QTL is known to be between *Ket* and *Ldh* (Dopman et al., 2005) or between 6.5 and 17.3 Mb on the *B. mori* Z chromosome. We used Bowtie 2 v.2.1.0 with *--end-to-end* and *--very-sensitive* options, which resulted in a higher percentage of uniquely mapped reads than any other combination of preset options. Count data were generated using the single best alignment for each read. To help mitigate G-C bias during PCR amplification, within and between sample data were corrected for G-C content using EDAseq (supplementary material Fig. S2) (Risso et al., 2011).

Gene Ontology Analysis. Annotation of genes was conducted by TBLASTX search to both the FlyBase and NCBI nr databases, retaining annotations with e-values $\leq 10^{-5}$ as putative orthologs. GOstats v.2.12 conditional hypergeometric test (Falcon and Gentleman, 2007) was used to test for "Biological Process", "Molecular Function", and "KEGG pathway" enrichment using the ontology terms from the Bioconductor package org.Dm.eg.db v.2.10.1 with the *P*-value cutoff of ≤ 0.001. The Short Time-series Expression Miner v.1.3.8 (Ernst et al., 2005) was used to look for groups of genes that showed coordinated expression change from diapause maintaining to diapause breaking conditions. Normalized expression data were randomly coupled from diapause maintenance to later developmental time-points and then $\log_2$ normalized to simulate microarray M-values. Multiple transcripts with duplicate annotations were averaged. STEM Miner determines all possible iterations gene expression change through time, with the diapause maintenance time-point normalized to 0, increasing or decreasing up to a user defined step size (=2). Each gene in the dataset is then assigned to a

given model based on its correlation coefficient, which was defined as 1. *P*-values were determined by calculating the observed number of genes assigned to each profile compared to the expected number, which is calculated by permuting the fit of each gene to the model 1000 times ($P < 0.01$). Significantly enriched Gene Ontology terms for each time-series was determined using Bonferroni corrected *P*-values (≤ 0.0001).

Differential expression. Differential expression analyses used edgeR (Robinson and Oshlack, 2010), which employs an empirical Bayes method to estimate gene-specific biological variation. A multi-factor analysis was used to account for both day and strain factors. Per-gene dispersions were estimated using the Cox-Reid profile-adjusted method and a negative binomial generalized log-linear model was fit to identify significantly differentially expressed transcripts (supplementary material Figs. S3, S4). Subsequently, gene lists were corrected for multiple testing using a false discovery rate (FDR) (cutoff of ≤ 0.01). Candidates within the predicted QTL interval were selected if they showed no differential expression between strains during diapause maintenance, were differentially expressed between strains on day 1 or 7, or were differentially expressed within strain through time.

SNP Analysis. We identified fixed SNPs between strains by pooling all libraries by strain, mapping reads, and estimating SNP frequencies using Popoolation2 v1.201 (Kofler et al., 2011). This method has been shown to be effective at recovering the majority of true allele frequencies in populations (Konczal et al., 2014). A minimum coverage of 8 reads per population and a pool size of 75 was used to calculate F_{ST} for every SNP. To determine if fixed SNPs coded for an amino acid change in the subset of transcripts whose putative locations were within the QTL, ESTScan was used to predict the reading frame (Iseli et al., 1999).

ACKNOWLEDGMENTS

We would like to thank the DopChewLew laboratory group for early discussion of this project, and Genevieve Kozak for providing feedback on early drafts of this manuscript.

DATA ARCHIVING

Libraries are archived in Genbank (Bioproject ID: PRJNA294976; for accession numbers see supplementary material Table S4).

COMPETING INTERESTS

No competing interests declared.

AUTHOR CONTRIBUTIONS

Listed authors EBD and CBW both contributed to conception of the project, analysis of the data, and writing the final manuscript. CBW conducted data collection.

FUNDING

The study was supported by the United States Department of Agriculture (2010-65106-20610 to EBD) and the National Science Foundation (DEB-1257251 to EBD; 2011116050 to CBW).

REFERENCES

Beck, S. D. and Apple, J. W. (1961). Effects of temperature and photoperiod on voltinism of geographical populations of the European Corn Borer, *Pyrausta nubilalis*. *J. Econ. Entomol.* **54**, 550–558.

Bin Bao and Xu, W.-H. (2011). Identification of gene expression changes associated with the initiation of diapause in the brain of the cotton bollworm, *Helicoverpa*. *BMC Genomics* **12**, 224.

Bradshaw, W. E., and Lounibos, L. P. (1977). Evolution of dormancy and its photoperiodic control in

pitcher-plant mosquitoes. Evolution **31**, 546–567.

Bradshaw, W. E. and Holzapfel, C. M. (2001). Genetic shift in photoperiodic response correlated with global warming. *P. Natl. Acad. Sci.* **98**, 14509–14511.

Bradshaw, W. E. and Holzapfel, C. M. (2006). Climate Change: Evolutionary response to rapid climate change. *Science* **312**, 1477–1478.

Bradshaw, W. E. and Holzapfel, C. M. (2008). Genetic response to rapid climate change: it's seasonal timing that matters. *Mol. Ecol.* **17**, 157–166.

Bradshaw, W. E., Zani, P. A. and Holzapfel, C. M. (2004). Adaptation to temperate climates. *Evolution* **58**, 1748–1762.

Bradshaw, W. E., Haggerty, B. P. and Holzapfel, C. M. (2005). Epistasis underlying a fitness trait within a natural population of the pitcher-plant mosquito, *Wyeomyia smithii*. *Genetics* **169**, 485–488.

Bünning, E. (1936). Endogenous daily rhythms as the basis of photoperiodism. *Ber Deut Bot Ges* **54**, 590–607.

Chan, E. Y., Stang, S. L., Bottorff, D. A. and Stone, J. C. (1999). Hypothermic stress leads to activation of Ras-Erk signaling. *J. Clin. Invest.* **103**, 1337–1344.

Cogni, R., Kuczynski, C., Koury, S., Lavington, E., Behrman, E. L., O'Brien, K. R., Schmidt, P. S. and Eanes, W. F. (2013). The intensity of selection acting on the *couch potato* gene-spatial-temporal variation in a diapause cline. *Evolution* **68**, 538–548.

Danilevskii, A.S. (1965) *Photoperiodism and Seasonal Development of Insects*. London, UK: Oliver & Boyd Ltd.

Danks, H.V. (1987). Insect dormancy: An ecological perspective. *Biological Survey of Canada (Terrestrial Arthropods)* **1**, 1–439.

Dahl, G. E., Elsasser, T. H., Capuco, A. V., Erdman, R. A. and Peters, R. R. (1997). Effects of a long daily photoperiod on milk yield and circulating concentrations of insulin-like growth factor-I. *J. Dairy Sci.* **80**, 2784-2789.

Denlinger, D. L. (2002). Regulation of Diapause. *Annu. Rev. Entomol.* **47**, 93–122.

Dopman, E. B., Bogdanowicz, S. M. and Harrison, R. G. (2004). Genetic Mapping of Sexual Isolation Between E and Z Pheromone Strains of the European corn borer (*Ostrinia nubilalis*). *Genetics* **167**, 301–309.

Dopman, E. B., Perez, L., Bogdanowicz, S. M. and Harrison, R. G. (2005). Consequences of reproductive barriers for genealogical discordance in the European corn borer. *P. Natl. Acad. Sci.* **102**, 14706–14711.

Dopman, E. B., Robbins, P.S. and Seaman, A. (2010). Components of reproductive isolation between North American pheromone strains of the European corn borer. *Evolution* **64**, 881-902.

Emerson, K. J., Bradshaw, W. E. and Holzapfel, C. M. (2010). Microarrays reveal early transcriptional events during the termination of larval diapause in natural populations of the mosquito, *Wyeomyia smithii. PLOS ONE* **5**, e9574.

Ernst, J., Nau, G. J. and Bar-Joseph, Z. (2005). Clustering short time series gene expression data. *Bioinformatics* **21**, i159–i168.

Falcon, S. and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258.

Feder, J. L., Roethele, J. B., Filchak, K., Niedbalski, J. and Romero-Severson, J. (2002). Evidence for inversion polymorphism related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* **163**, 939–953.

Filchak, K., Roethele, J. B. and Feder, J. L. (2000). Natural selection and sympatric divergence in the apple maggot fly, *Rhagoletis pomonella*. *Nature* **407**, 739–742.

Gadenne, C., Varjas, L. and Maucahmp, B. (1990). Effects of the non-steriodal ecdysone mimic, RH-5849, on diapause and non-diapause larvae of the European corn borer, *Ostrinia nubilalis*. *J. Insect Physiol.* **36**, 555–559.

Glover, T. J., Robbins, P. S., Eckenrode, C. J. and Roelofs, W. L. (1992). Genetic control of voltinism characteristics in European corn borer races assessed with a marker gene. *Arch. Insect Biochem.* **20**, 107–117.

Gokhale, R. H. and Shingleton, A. W. (2015). Size control: the developmental physiology of body and organ size regulation. *WIREs Dev. Biol*.

Gomi, T., Nagasaka, M., Fukuda, T. and Hagihara, H. (2007). Shifting of the life cycle and life-history traits of the fall webworm in relation to climate change. *Entomol. Exper. Applic.* **125**, 179–184.

Gordon, A. and Hannon, G. J. (2010). FASTX-Toolkit: FASTQ/A short-reads pre-processing tools. Retrieved from: http://www.hannonlab.cshl.edu/fastx_toolkit.html

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652.

Gu, S.-H., Lin, J.-L., Lin, P.-L. and Chen, C.-H. (2009). Insulin stimulates ecdysteroidogenesis by prothoracic glands in the silkworm, *Bombyx mori*. *Insect Biochem. Molec.* **39**, 171–179.

Hagen, R. H. and Scriber, J. M. (1989). Sex-linked diapause, color, and allozyme loci in *Papilio glaucus*: linkage analysis and significance in a hybrid zone. *J. Hered*. **80**, 179-185.

Hahn, D. A. and Denlinger, D. L. (2011). Energetics of Insect Diapause. *Annu. Rev. Entomol.* **56**, 103–121.

Han, B. and Denlinger, D. L. (2009). Length variation in a specific region of the *period* gene correlates with differences in pupal diapause incidence in the flesh fly, *Sarcophaga bullata*. *J. Insect Physiol.* **55**, 415–418.

Hermanto, U., Zong, C. S., Li, W. and Wang, L. H. (2002). RACK1, an insulin-like growth factor I (IGF-I) receptor-interacting protein, modulates IGF-I-dependent integrin signaling and promotes cell spreading and contact with extracellular matrix. *Mol. Cell. Biol.* **22**, 2345–2365.

Ikeno, T., Tanaka, S. I., Numata, H. and Goto, S. (2010). Photoperiodic diapause under the control of circadian clock genes in an insect. *BMC Biol.* **8**, 116.

Iseli, C., Jongeneel, C. V. and Bucher, P. (1999). ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in ESTsequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **99**, 138–148.

Kahn, B. B., Alquier, T., Carling, D. and Hardie, D. G. (2005). AMP-activated protein kinase: Ancient energy gauge provides clues to modern understanding of metabolism. *Cell. Metab.* **1**, 15–25.

Kenyon, C. J. (2010). The genetics of ageing. *Nature* **464**, 504–512.

Kidokoro, K., Iwata, K.-I., Fujiwara, Y. and Takeda, M. (2006). Effects of juvenile hormone analogs and 20-hydroxyecdysone on diapause termination in eggs of *Locusta migratoria* and *Oxya yezoensis*. *J. Insect Physiol.* **52**, 473–479.

Kofler, R., Pandey, R. V. and Schlotterer, C. (2011). PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435–3436.

Konczal, M., Koteja, P., Stuglik, M. T., Radwan, J. and Babik, W. (2014). Accuracy of allele frequency estimation using pooled RNA-Seq. *Mole. Ecol. Resour.* **14**, 381–392.

Koštál, V. (2006). Eco-physiological phases of insect diapause. *J Insect Physiol* **52**, 113-127.

Koštál, V. (2011). Insect photoperiodic calendar and circadian clock: independence, cooperation, or unity? *J. Insect Physiol.* **57**, 538-556.

Koštál, V. and Tollarová-Borovanská, M. (2009). The 70 kda heat shock protein assists during the repair of chilling injury in the insect, *Pyrrhocoris apterus*. *PLOS ONE* **4**, e4546.

Koštál, V., Simuringnkov, A. P., Kobelkov, A.A. and Shimada, K. (2009). Cell cycle arrest as a hallmark of insect diapause: Changes in gene transcription during diapause induction in the drosophilid fly, *Chymomyza costata*. *Insect Biochem. Molec.* **39**, 875–883.

Kroemer, J. A., Coates, B. S., Nusawardani, T., Rider, S. D., Fraser, L. M. and Hellmich, R. L. (2011). A rearrangement of the Z chromosome topology influences the sex-linked gene display in the European corn borer, *Ostrinia nubilalis*. *Mol. Genet. Genomics* **286**, 37–56.

Kyriacou, C., Peixoto, A.A., Sandrelli, F., Costa, R. and Tauber, E. (2008). Clines in clock genes: fine-tuning circadian rhythms to the environment. *Trends Genet.* **24**, 124-132.

Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Meth.* **9**, 357–359.

Leopold, A.C. and Lang, G.A. (1996). Natural history of seed dormancy. In *Plant Dormancy: Physiology, Biochemistry and Molecular Biology*, pp. 3–16. Wallingford,UK: CAB International.

Levy, R. C., Kozak, G. M., Wadsworth, C. B., Coates, B. S. and B, D. E. (2015). Explaining the sawtooth: latitudinal periodicity in a circadian gene correlates with shifts in generation number. *J Evolution. Biol.* **28**, 40–53.

Lippai, M., Csikós, G., Maróy, P., Lukácsovich, T., Juhász, G. and Sass, M. (2008). SNF4Aγ, the Drosophila AMPK γ subunit is required for regulation of developmental and stress-induced autophagy. *Autophagy* **4**, 476–486.

Liu, J.-P., Baker, J., Perkins, A. S., Robertson, E. J. and Efstrtiadis, A. (1993). Mice carrying null mutations of the genes encoding insulin-like growth factor I (*Igf-1*) and type 1 IGF receptor (*Igf1r*). *Cell* **75**, 59–72.

Logan, C. Y. and Nusse, R. (2004). The Wnt signaling pathways in development and disease. *Ann. Rev Cell. Dev. Biol.* **20**, 781–810.

Lohse, M., Bolger, A. M., Nagel, A., Fernie, A. R., Lunn, J. E., Stitt, M. and Usadel, B. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic Acids Res.* **40**, W622–W627.

Luckenbach, J. A., Murashige, R., Daniels, H. V., Godwin, J. and Borski, R. J. (2007). Temperature affects insulin-like growth factor I and growth of juvenile southern flounder, *Paralichthys lethostigma*. *Comp. Biochem. Phys. A* **146**, 95–104.

Malausa, T., Leniaud, L., Martin, J-F., Audiot, P., Bourguet, D., Ponsard, S., Lee, S.F., Harrison,

R.G. and Dopman, E. (2007). Molecular differentiation at nuclear loci in French host races of the European corn borer (*Ostrinia nubilalis*). Genetics **176**, 2343–2355.

Mathias, D., Jacky, L., Bradshaw, W. E. and Holzapfel, C. M. (2005). Geographic and developmental variation in expression of the circadian rhythm gene, *timeless*, in the pitcher-plant mosquito, *Wyeomyia smithii*. *J. Insect Physiol.* **51**, 661–667.

Mathias, D., Jacky, L., Bradshaw, W. E. and Holzapfel, C. M. (2007). Quantitative trait loci associated with photoperiodic response and stage of diapause in the pitcher-plant mosquito, *Wyeomyia smithii*. *Genetics* **176**, 391–402.

McLeod, D. and Beck, S. (1963). Photoperiodic termination of diapause in an insect. *Biol. Bull.* **124**, 84–96.

Ording, G. J., J, M. R., Aardema, M. L. and Scriber, J. M. (2010). Allochronic isolation and incipient hybrid speciation in tiger swallowtail butterflies. *Oecologia* **162**, 523–531.

Palmer, D. F., Schenk, T. C. and Chiang, H. (1985). Dispersal and voltinism adaptation of the European corn Borer in North America, 1917-1977. *Station Bulletin Agricultural Experiment Station University of Minnesota*, 1–36.

Pickett, A. D. and Neary, M. E. (1940). Further studies on *Rhagoletis pomonella* (Walsh). *Sci. Agr.* **20**, 551-556.

Poelchau, M. F., Reynolds, J. A., Denlinger, D. L., Elsik, C. G. and Armbruster, P. A. (2011). A *de novo* transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. *BMC Genomics* **12**, 619.

Pringle, E. G., Baxter, S. W., Webster, C. L., Papanicolaou, A., Lee, S.-F. and Jiggins, C. D. (2007). Synteny and chromosome evolution in the lepidoptera: Evidence from mapping in *Heliconius melpomene*. *Genetics* **177**, 417–426.

Ragland, G. J., Fuller, J., Feder, J. L., & Hahn, D. A. (2009). Biphasic metabolic rate trajectory of pupal diapause termination and post-diapause development in a tephritid fly. *J. Insect Physiol.* **55**, 344-350.

Ragland, G. J., Denlinger, D. L. and Hahn, D. A. (2010). Mechanisms of suspended animation are revealed by transcript profiling of diapause in the flesh fly. *P. Natl. Acad. Sci.* **107**, 14909–14914.

Ragland, G. J., Egan, S. P., Feder, J. L., Berlocher, S. H. and Hahn, D. A. (2011). Developmental trajectories of gene expression reveal candidates for diapause termination: a key life-history transition in the apple maggot fly *Rhagoletis pomonella*. *J. Exp. Biol.* **214**, 3948–3960.

Rinehart, J. P., Li, A., Yocum, G. D., Robich, R. M., Hayward, S. A. L. and Denlinger, D. L. (2007). Up-regulation of heat shock proteins is essential for cold survival during insect diapause. *Proc. Natl. Acad. Sci.* **104**, 11130–11137.

Risso, D., Schwartz, K., Sherlock, G. and Dudoit, S. (2011). GC-Content Normalization for RNA-Seq Data. *BMC Bioinformatics* **12**, 480.

Robinson, M. D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25.

Sato, A., Sokabe, T., Kashio, M., Yasukochi, Y., Tominaga, M. and Shiomi, K. (2014). Embryonic thermosensitive TRPA1 determines transgenerational diapause phenotype of the silkworm, *Bombyx mori*. *P. Natl. Acad. Sci.* **111**, E1249–E1255.

Schmidt, P. S., Matzkin, L., Ipplito, M. and Eanes, W. F. (2005). Geographic variation in diapause incidence, life-history traits, and climatic adaptation in *Drosophila melanogaster*. *Evolution* **59**, 1721–1732.

Schmidt, P. S., Zhu, C. T., Das, J., Batavia, M., Yang, L. and Eanes, W. F. (2008). An amino acid polymorphism in the *couch potato* gene forms the basis for climatic adaptation in *Drosophila melanogaster*. *P. Natl. Acad. Sci.* **105**, 16207–16211.

Shimomura, M., Minami, H., Suetsugu, Y., Ohyanagi, H., Satoh, C., Antonio, B., Nagamura, Y., Kadono-Okuda, K., Kajiwara, H., Sezutsu, H., et al. (2009). KAIKObase: An integrated silkworm genome database and data mining tool. *BMC Genomics* **10**, 486.

Shingleton, A. W., Das, J., Vinicius, L. and Stern, D. L. (2005). The temporal requirements for insulin signaling during development in *Drosophila*. *PLOS Biol.* **3**, e289.

Sim, C. and Denlinger, D. L. (2009). A shut-down in expression of an insulin-like peptide, ILP-1, halts ovarian maturation during the overwintering diapause of the mosquito *Culex pipiens*. *Insect Mol. Biol.* **18**, 325–332.

Smith, W. A., Lamattina, A. and Collins, M. (2014). Insulin signaling pathways in lepidopteran ecdysone secretion. *Frontiers Physiol.* **5**, 19.

Suttie, J. M., White, R. G., Breier, B. H. and Gluckman, P. D. (1991). Photoperiod associated changes in insulin-like growth factor-I in reindeer. *Endocrinology* **129**, 679–682.

Tauber, C. and Tauber, M. (1981). Insect seasonal cycles: Genetics and evolution. *Annu. Rev. Ecol. Syst.* **12**, 281–308.

Tauber, C., Tauber, M. and Nechols, J.R. (1977) Two genes control seasonal isolation in sibling species. *Science* **197**, 592–593.

Tauber, E., Zordan, M., Sandrelli, F., Pegoraro, M., Osterwalder, M., Breda, C., Daga, A., Selmin, A., Monger, K., Benna, C., et al. (2007). Natural selection favors a newly derived *timeless* allele in *Drosophila melanogaster*. *Science* **316**, 1895–1898.

Tauber, M. J. and Tauber, C. A. (1976). Environmental control of univoltinism and its evolution in an
insect species. *Can. J. Zool.* **54**, 260-265.

Tauber, M. J., Tauber, C. and Masaki, S. (1986). *Seasonal Adaptations of Insects*. Oxford, UK: Oxford University Press.

Taylor, J. F., Migaud, H., Porter, M. J. R. and Bromage, N. R. (2005). Photoperiod influences growth rate and plasma insulin-like growth factor-I levels in juvenile rainbow trout, *Oncorhynchus mykiss*. *Gen. Comp. Endocr.* **142**, 169–185.

Wadsworth, C. B., Woods, W. A., Jr, Hahn, D. A. and B, D. E. (2013). One phase of the dormancy developmental pathway is critical for the evolution of insect seasonality. *J. Evolution Biol.* **26**, 2359–2368.

Wadsworth, C. B., Li, X. and Dopman, E. B. (2015). A recombination suppressor contributes to ecological speciation in *Ostrinia* moths. *Heredity* ***114,*** 593-600.

Yamamoto, K., Nohata, J., Kadono-Okuda, K., Narukawa, J., Sasanuma, M., Sasanuma, S.-I.,

Minami, H., Shimomura, M., Suetsugu, Y., Banno, Y., et al. (2008). A BAC-based integrated linkage map of the silkworm *Bombyx mori*. *Genome Biol.* **9**, R21.

Yoshiyama, T., Namiki, T., Mita, K., Kataoka, H. and Niwa, R. (2006). *Neverland* is an evolutionally conserved Rieske-domain protein that is essential for ecdysone synthesis and insect growth. *Development* **133**, 2565–2574.

Zhao, Q.-Y., Wang, Y., Kong, Y.-M., Da Luo, Li, X. and Hao, P. (2011). Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* **12**, S2.

Žďárek, J. and Denlinger, D. L. (1975). Action of ecdysoids, juvenoids, and non-hormonal agents on the termination of pupal diapause in the flesh fly. *J. Insect Physiol.* **21**, 1193–1202.

FIGURES

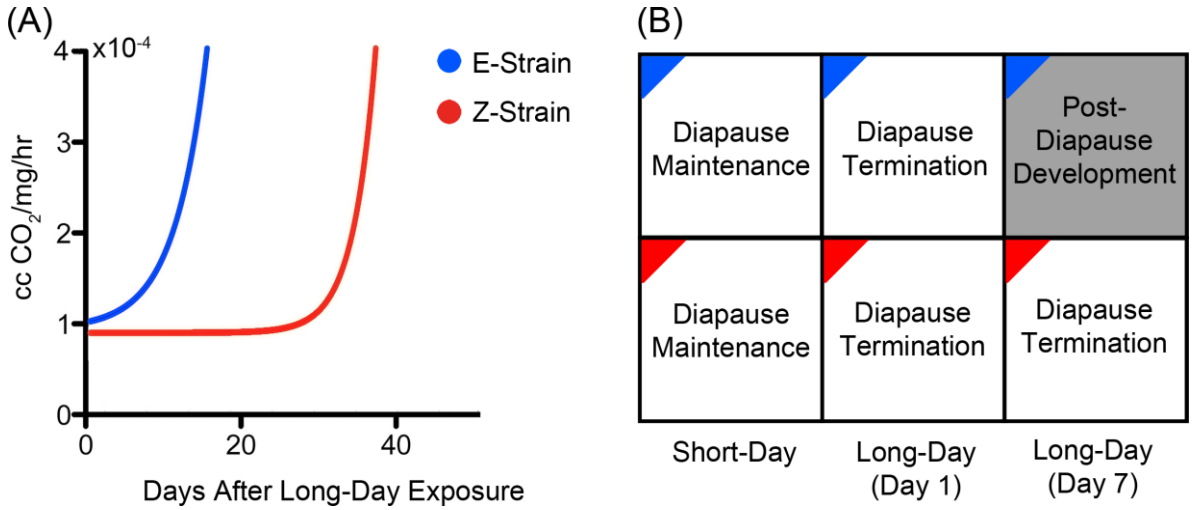

Figure 1. Divergence in the timing of photoperiodically-terminated diapause in the European corn borer moth. (A) Non-linear regression of median metabolic trajectories for earlier-emerging E-strain (blue) and later-emerging Z-strain (red) borers after transfer from short-day (diapause inducing) to long-day (diapause breaking) conditions (Wadsworth et al., 2013). Physiological models suggest that the end of diapause, which coincides with metabolic uptick, is shifted by ~30 days in Z-strain borers. (B) Experimental design for transcriptome profiling in which borers were sampled before and after experiencing long-day cues.

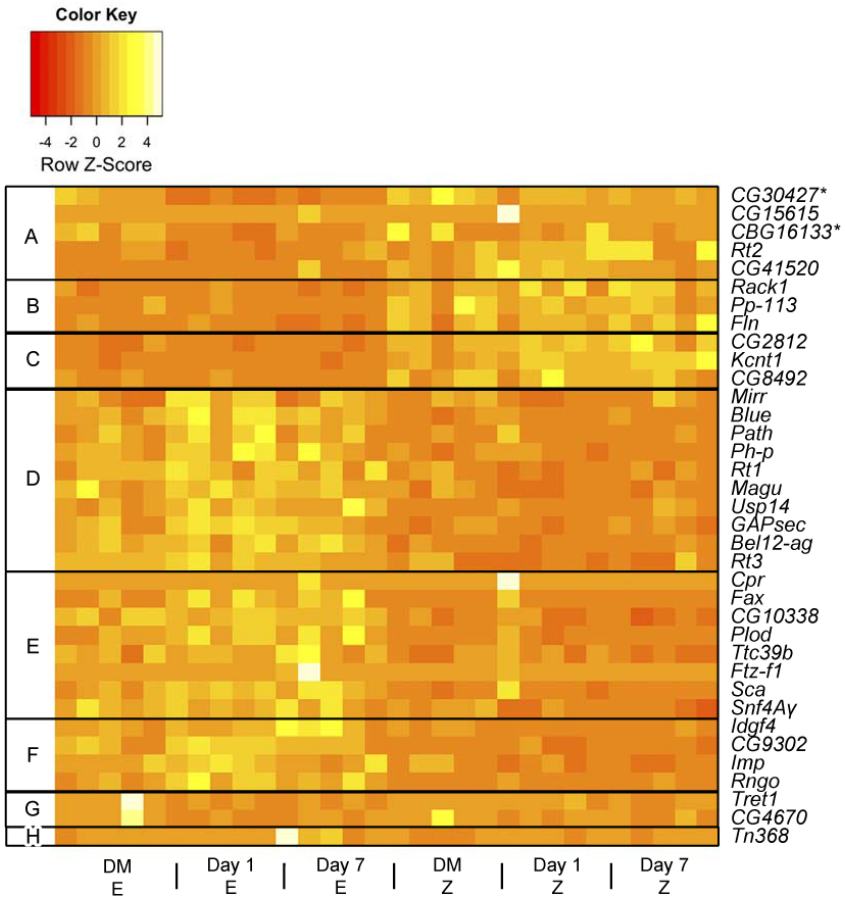

Figure 2. Heatmap of differentially expressed genes within the *Ket-Ldh* genomic interval. (A) 5 genes upregulated in the Z-strain on day 1 relative to the E-strain; (B) 3 genes upregulated in the Z-strain on day 7 relative to the E-strain; (C) 3 genes upregulated in the Z-strain on days 1 and 7 relative to the E-strain; (D) 10 genes upregulated in the E-strain on day 1 relative to the Z-strain; (E) 8 genes upregulated in the E-strain on day 7 relative to the Z-strain; and (F) 4 genes upregulated in the E-strain on days 1 and 7 relative to the Z-strain. Within the E-strain there were two observed expression patterns: (G) 4 downregulated genes from diapause maintenance to day 1 (*including two genes in A) and (H) 1 upregulated gene from diapause maintenance to day 7. Duplicate annotations were averaged for expression. All were FDR < 0.01.

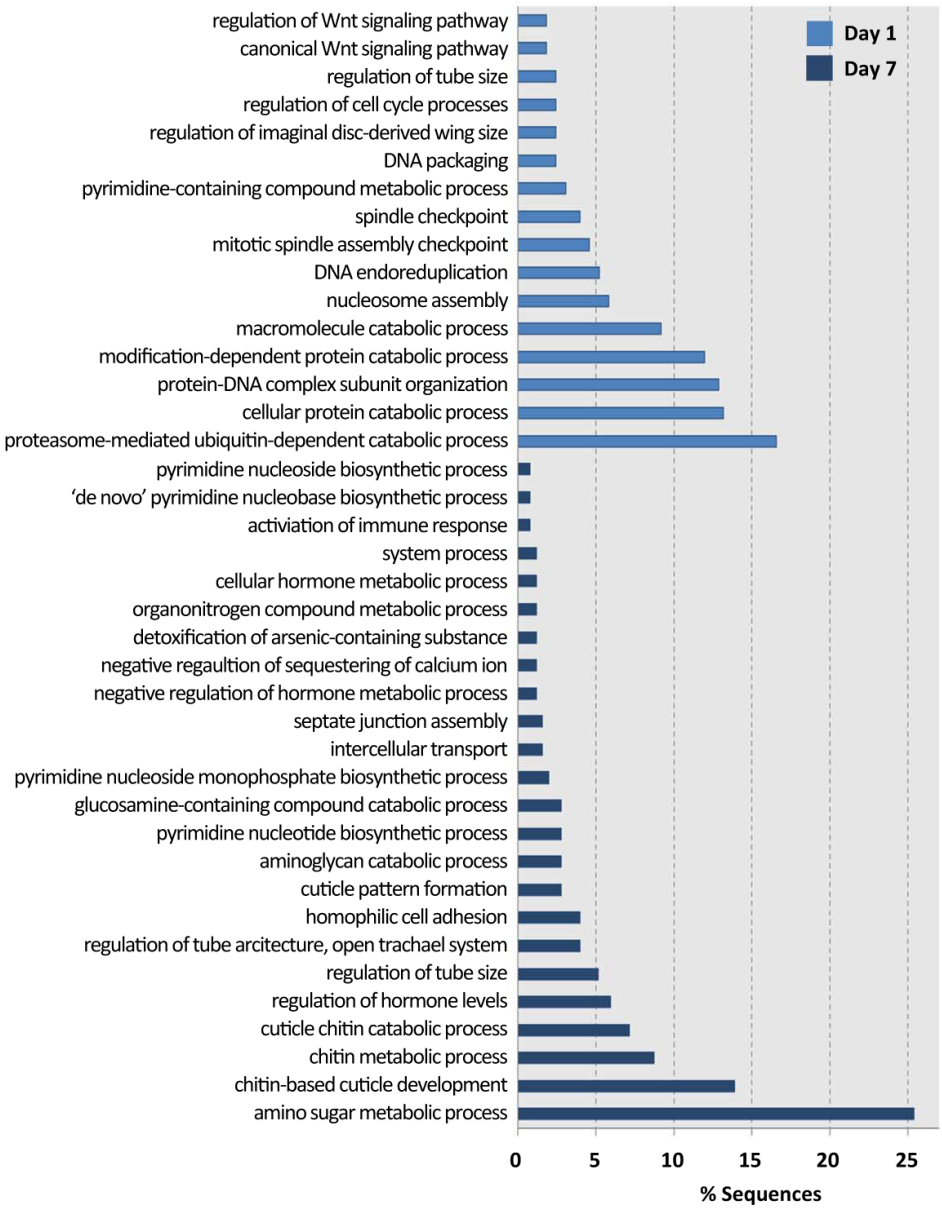

Figure 3. Enriched Biological Process GO terms for differentially expressed genes on day 1 (light blue) and day 7 (dark blue) after exposure to long-day conditions. *P*-value cutoff < 0.001 or < 0.05 (*).

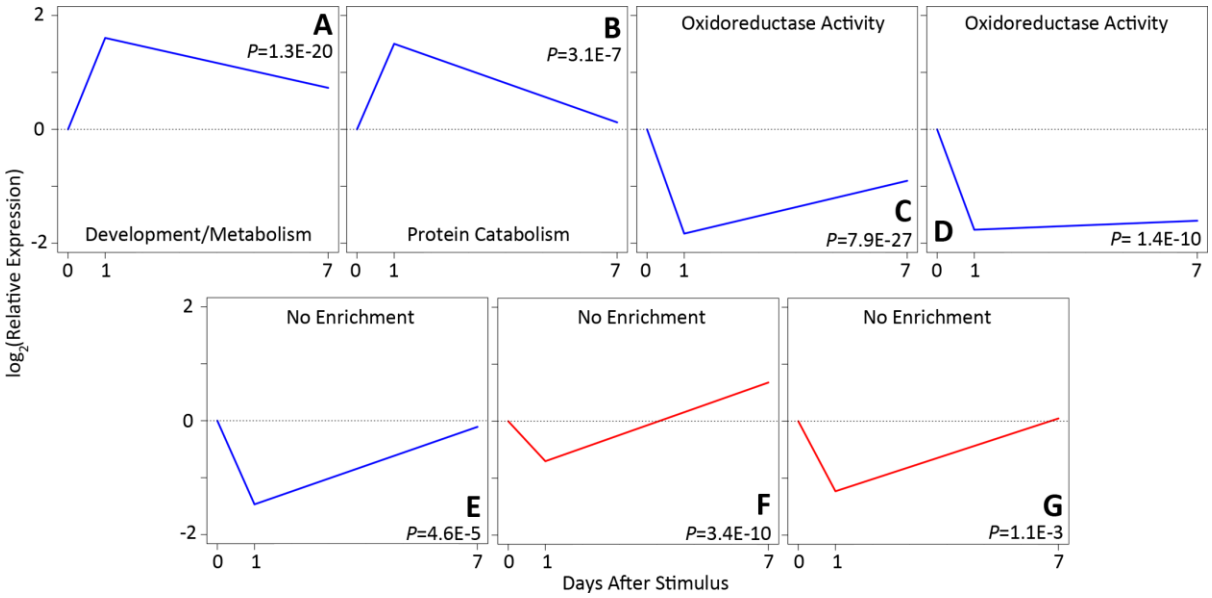

Figure 4. Pathways with correlated changes in gene expression during the diapause developmental pathway. Gene expression was $\log_2$ normalized to the diapause maintenance time-point for each strain. (A-E) are significant ($P \leq 0.01$) trajectories identified in the E-strain (blue) and (F-G) are significant ($P \leq 0.01$) in the Z-strain (red). Labels represent enriched GO terms for each profile (supplementary material Table S2).

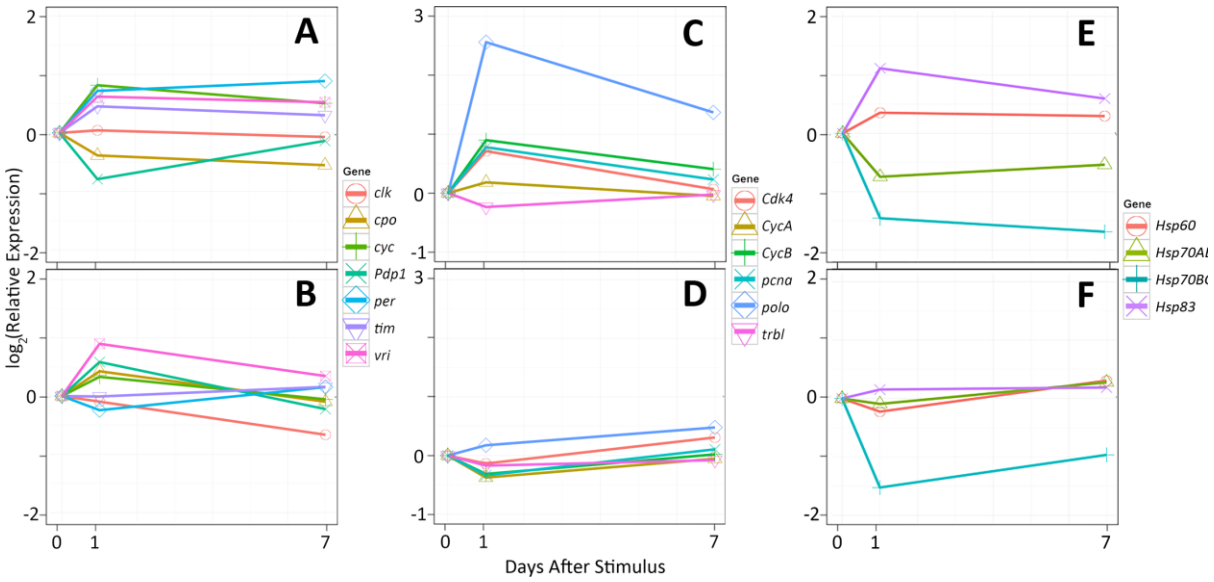

Figure 5. Trajectories of gene expression during the diapause developmental pathway. Shown are circadian genes in (A) the E-strain and (B) the Z-strain, cell cycle genes in (C) the E-strain and (D) the Z-strain, and heat shock genes in (E) the E-strain and (F) the Z-strain. Gene expression data were normalized to begin at zero during diapause maintenance where day 1 and day 7 time-points show simulated M values representing the $\log_2$ (diapause promoting/diapause breaking conditions).

TABLES

Table 1. Number of significantly differentially expressed (DE) transcripts between and within strain (FDR <0.01)

Comparison	Total DE	DE Corrected for Population Divergence
Diapause E vs Z	3077 (71)	-
Day 1 E vs Z	6615 (135)	4160 (77)
Day 7 E vs Z	4633 (114)	2337 (60)
Within Z Diapause vs Day 1	23 (0)	20 (0)
Within Z Diapause vs Day 7	1 (0)	1 (0)
Within E Diapause vs Day 1	764 (9)	625 (6)
Within E Diapause vs Day 7	95 (1)	77 (1)

Values in parentheses show the subset of DE genes within each comparison that are putatively Z-linked

Table 2. Significantly differentially expressed transcripts (FDR < 0.01) located in the *Ket-Ldh* interval

Transcript ID	Description	Gene Symbol	Functional Category	FlyBase ID	NCBI ID	Location B.mori (Mb)	FDR	Expression Pattern*
Between Strains Day 1								
comp12218_c0_seq1	*Mirror*	*Mirr*	Morphogenesis	FBgn0014343	XP_004929953	7.8	2.30E-03	D
comp34828_c0_seq1	*Fatty acyl-CoA reductase CG30427*	*CG30427*	Regulation of Lifespan	FBgn0043792	EHJ78784	9.0	1.66E-05	A
comp9910_c0_seq1	*Ferritin-like CG15615*	*CG15615*	Stress Response	FBgn0034159	XP_004929844	9.2	1.83E-04	A
comp7706_c0_seq1	*Bluestreak*	*Blue*	Development	FBgn0041161	XP_004929978	10.6	2.52E-04	D
comp26862_c0_seq1	*Pathetic*	*Path*	Growth	FBgn0036007	XP_004929913	11.6	7.78E-03	D
comp7044_c0_seq1	*Polyhomeotic proximal*	*Ph-p*	Development	FBgn0004861	NP_001188364	11.8	7.61E-03	D
comp23880_c0_seq1	*Reverse transcriptase*	*Rt1*	Reverse Transcription	-	AAQ09229	11.9	3.16E-03	D
comp19186_c0_seq1	*CBG16133*	*CBG16133*	-	-	XP_004929918	12.0	3.76E-03	A
comp47212_c0_seq1; comp25161_c0_seq1	*Magu*	*Magu*	Regulation of Lifespan	FBgn0262169	XP_004929925	12.4	8.66E-04	D
comp27958_c0_seq1	*Ubiquitin carboxyl-terminal hydrolase 14*	*Usp14*	Protein Modification	FBgn0032216	XP_004927226	12.4	7.85E-03	D
comp6478_c0_seq1; comp9449_c0_seq1	*GTPase activating protein, SECIS-dependent read-through*	*GAPsec*	Translation	FBgn0035916	XP_004927187	14.2	4.56E-03	D
comp4829_c0_seq1	*Reverse transcriptase*	*Rt2*	Reverse Transcription	-	AAQ09229	14.4	3.23E-05	A
comp25302_c0_seq1	*BEL12-AG transposon*	*Bel12-ag*	DNA Integration	-	XP_004927201	14.5	1.26E-04	D
comp25983_c0_seq1	*CG41520*	*CG41520*	-	FBgn0087011	XP_004933130	15.4	2.85E-04	A
comp41343_c0_seq1	*Endonuclease and reverse transcriptase*	*Rt3*	Reverse Transcription	-	CAX36787	16.4	4.12E-06	D
Between Strains Day 7								
comp48442_c0_seq1	*Cuticular protein*	*Cpr*	Cuticular Structure	-	AK401102.1	9.1	4.36E-03	E
comp9204_c0_seq1	*Failed axon connections*	*Fax*	Development	FBgn0014163	XP_004929846	9.3	9.64E-04	E
comp19707_c0_seq1	*CG10338*	*CG10338*	-	FBgn0032700	XM_004929795.1	9.4	1.22E-03	E
comp81670_c0_seq1	*Receptor of activated protein kinase C 1*	*Rack1*	Development	-	EHJ71933	10.9	4.27E-03	B
comp18687_c0_seq1	*Procollagen lysyl hydroxylase*	*Plod*	Development	FBgn0036147	XP_004929891	10.9	4.79E-03	E
comp117543_c0_seq1	*Tetratricopeptide repeat protein 39B*	*Ttc39b*	-	-	XP_004929983	11.1	2.86E-03	E
comp30814_c0_seq5	*Unknown secreted protein*	*Pp-113*	-	-	AK405478.1	11.5	4.38E-03	B
comp224242_c0_seq1	*Nuclear hormone receptor betaFTZ-F1*	*Ftz-f1*	Development	-	AAL50351	11.7	1.32E-03	E

comp37732_c0_seq1	*Flightin*	*Fln*	Development	FBgn0005633	NP_001130045	12.9	3.67E-05	B
comp78006_c0_seq1	*Scabrous*	*Sca*	Morphogenesis	FBgn0003326	XP_004930012	13.5	4.79E-03	E
comp26437_c0_seq1	*SNF4/AMP-activated protein kinase γ subunit*	*Snf4Aγ*	Growth/Stress	-	NM_001126248.1	15.6	3.11E-03	E
Between Strains Day 1 and Day 7								
comp9814_c0_seq1; comp35628_c0_seq1	*WD repeat-containing protein*	*CG2812*	-	FBgn0034931	EHJ79015	7.1	4.66E-03	C
comp32160_c0_seq1	*Imaginal disc growth factor 4*	*Idgf4*	Metamorphosis	FBgn0013763	ACW82749	8.5	6.44E-04	F
comp43490_c0_seq1; comp27799_c0_seq1	*Protein disulfide-isomerase*	*CG9302*	Cell Homeostasis	FBgn0032514	XP_004929835	9.0	4.21E-03	F
comp226203_c0_seq1	*Potassium channel subfamily T member 1*	*Kcnt1*	Ion Transport	-	XP_004929982	11.0	8.29E-05	C
comp12869_c0_seq1	*IGF-II mRNA-binding protein*	*Imp*	Response to Stress	FBgn0262735	EHJ69667	11.4	5.42E-03	F
comp9302_c0_seq1	*Rings lost*	*Rngo*	Development	FBgn0030753	NM_001046849.1	16.6	2.28E-03	F
comp18783_c0_seq1	*Lysozyme*	*CG8492*	Hydrolase Activity	FBgn0035813	ADA67927	16.9	1.80E-03	C
Diapause Maintenance to Later Time-point Within E Strain								
comp34828_c0_seq1	*Fatty acyl-CoA reductase CG30427*	*CG30427*	Regulation of Lifespan	FBgn0043792	EHJ78784	9.0	2.75E-04	G
comp11796_c0_seq1	*Facilitated trehalose transporter 1*	*Tret1*	Transmembrane Transport	FBgn0051100	XP_004929980	10.8	5.35E-03	G
comp19186_c0_seq1	*CBG16133*	*CBG16133*	-	-	XP_004929918	12.0	3.22E-03	G
comp9625_c0_seq1	*Sulfhydryl oxidase*	*CG4670*	Metabolic Process	FBgn0033814	XP_004929937	13.4	2.94E-03	G
comp166238_c0_seq1	*Retrotransposon TED*	*Tn368*	Reverse Transcription	-	M32662.1	14.5	5.81E-03	H

* Expression pattern refers to Figure 2.

Table 3. All fixed SNPs resulting in predicted amino acid changes between E and Z-strain borers within the *Ket-Ldh* interval

Transcirpt ID	Description	Gene Symbol	Functional Category	FlyBase ID	Location B.mori (Mb)	SNP Z/E	Position Transcript	Predicted AA inChange Z/E	Predicted Polarity Change
comp34828_c0_seq1	*Fatty acyl-CoA reductase CG30427*	*CG30427*	Regulation of Lifespan	FBgn0043792	9.0	C/T	1378	Ala/Thr	Nonpolar -> Polar
comp25898_c0_seq1	*Belphegor*	*Bor*	ATP Binding	FBgn0040237	9.9	T/G	1004	Ser/Ala	Polar -> Nonpolar
comp117726_c0_seq1	-	-	-	-	10.2	A/G	280	Cys/Arg	Polar -> Basic
comp67276_c0_seq1	-	-		-	10.6	A/C	486	Ile/Cys	Nonpolar -> Polar
comp67276_c0_seq1	-	-	-	-	10.6	T/A	487	Ile/Cys	Nonpolar -> Polar
comp18031_c0_seq1	*Neprilysin 4*	*Nep4*	Metabolic Process	FBgn0038818	10.6	T/A	210	Phe/Leu	Nonpolar -> Nonpolar
comp26697_c0_seq1	*Pico*	*Pico*	Development/Growth	FBgn0261811	10.7	A/G	1072	Ile/Val	Nonpolar -> Nonpolar
comp122129_c0_seq1	*Acetyl Coenzyme A synthase*	*AcCoAS*	Response to Stimulus	FBgn0012034	11.9	G/T	258	Ala/Glu	Nonpolar -> Acidic
comp29232_c0_seq2	*CG10226*	*CG10226*	Response to Stimulus	FBgn0035695	12.2	C/T	1014	Ser/Asn	Polar -> Polar
comp18621_c0_seq1	*CG10226*	*CG10226*	Response to Stimulus	FBgn0035695	12.2	G/A	233	Ser/Asn	Polar -> Polar
comp97089_c0_seq1	*Period*	*Per*	Circadian Rhythm	FBgn0003068	13.0	C/G	232	Gly/Arg	Nonpolar -> Basic
comp81055_c0_seq1	*Activated Cdc42 kinase-like*	*Ack-like*	Cell Death	FBgn0263998	16.2	T/G	384	Ser/Ala	Polar -> Nonpolar

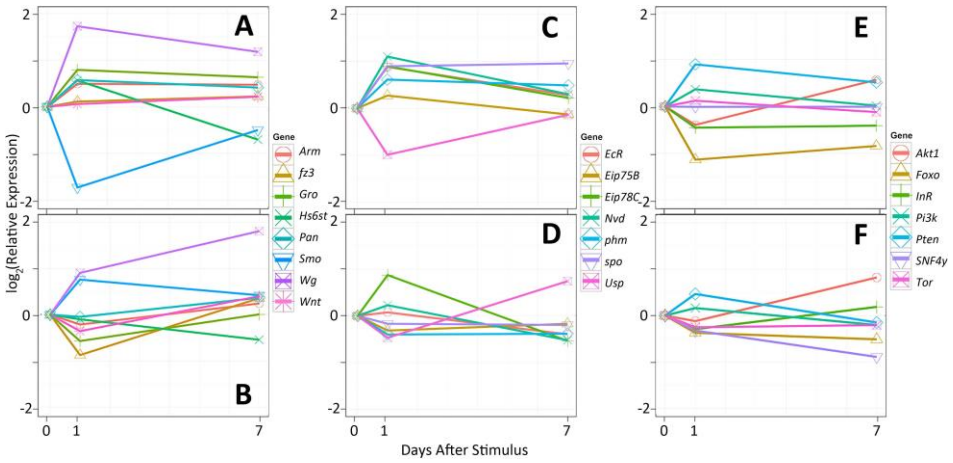

Fig. S1. Gene expression trajectories from short-day to long-day conditions for (A) E-strain Wnt signaling genes, (B) Z-strain Wnt signaling genes, (C) E-strain ecdysone-related genes, (D) Z-strain ecdysone-related genes, (E) E-strain insulin signaling genes, and (F) Z-strain insulin signaling genes.

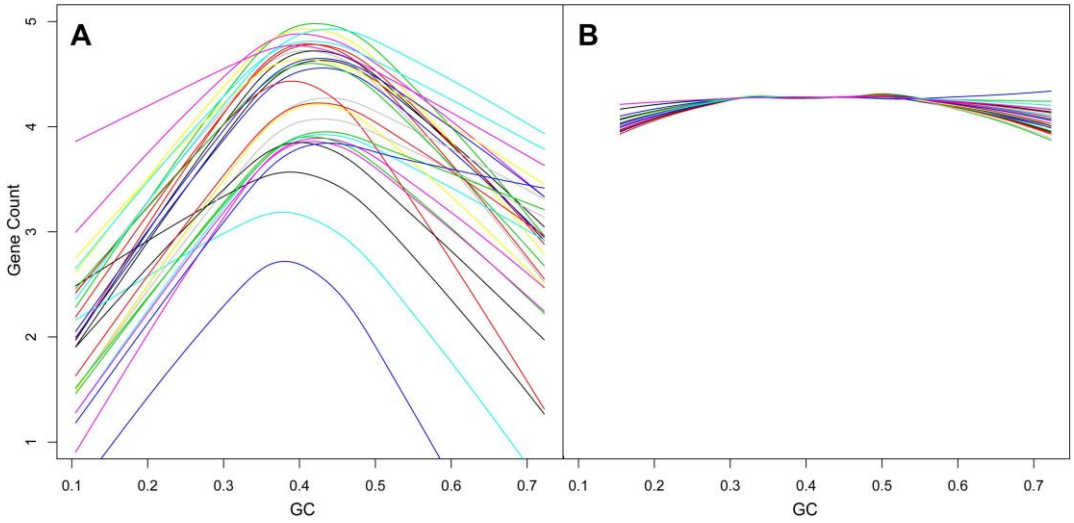

Fig. S2. GC normalization in EDAseq. (A) Lowess regression of log gene-level counts on GC content for each library. (B) Full quantile within and between lane normalization for each library.

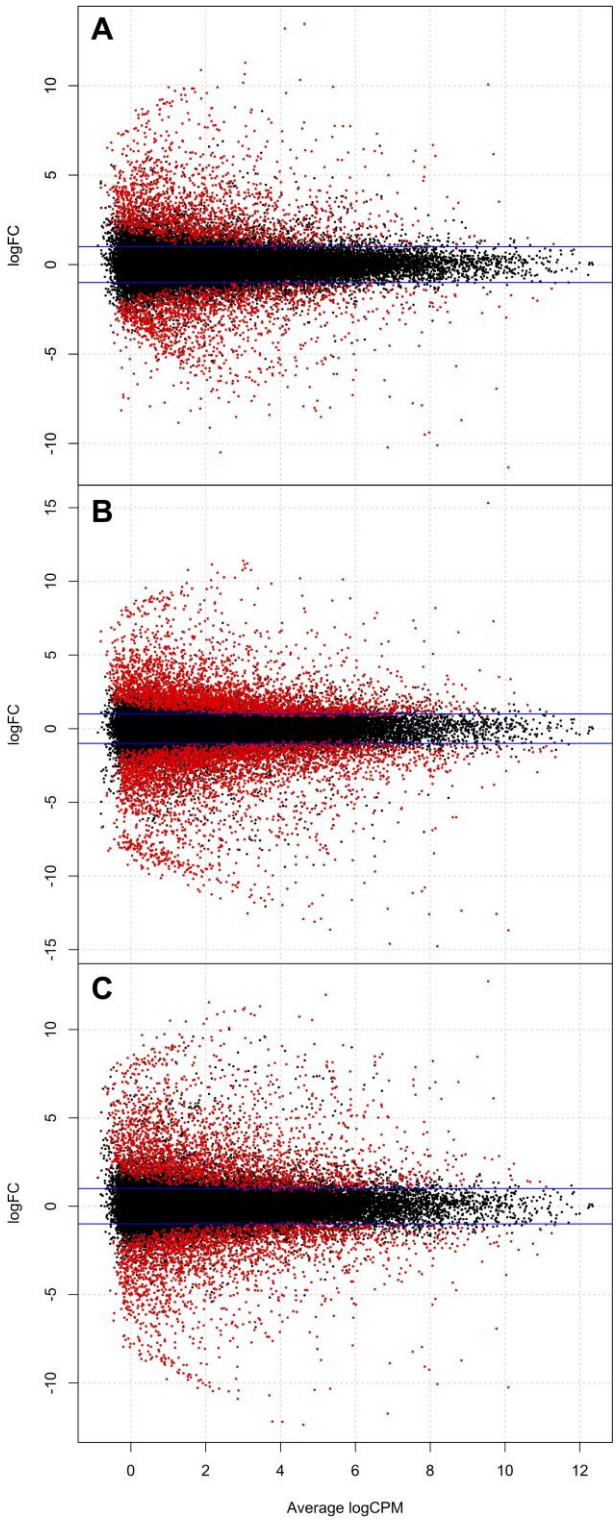

Fig. S3. MA plot of differential expression within strain shown as log-fold change of expression versus average log counts per million. (A) E-strain diapause maintenance to day 1, (B) E-strain diapause maintenance to day 7, (C) Z-strain diapause maintenance to day 1, (D) Z-strain diapause maintenance to day 7. Red dots are significantly differentially expressed at FDR < 0.01.

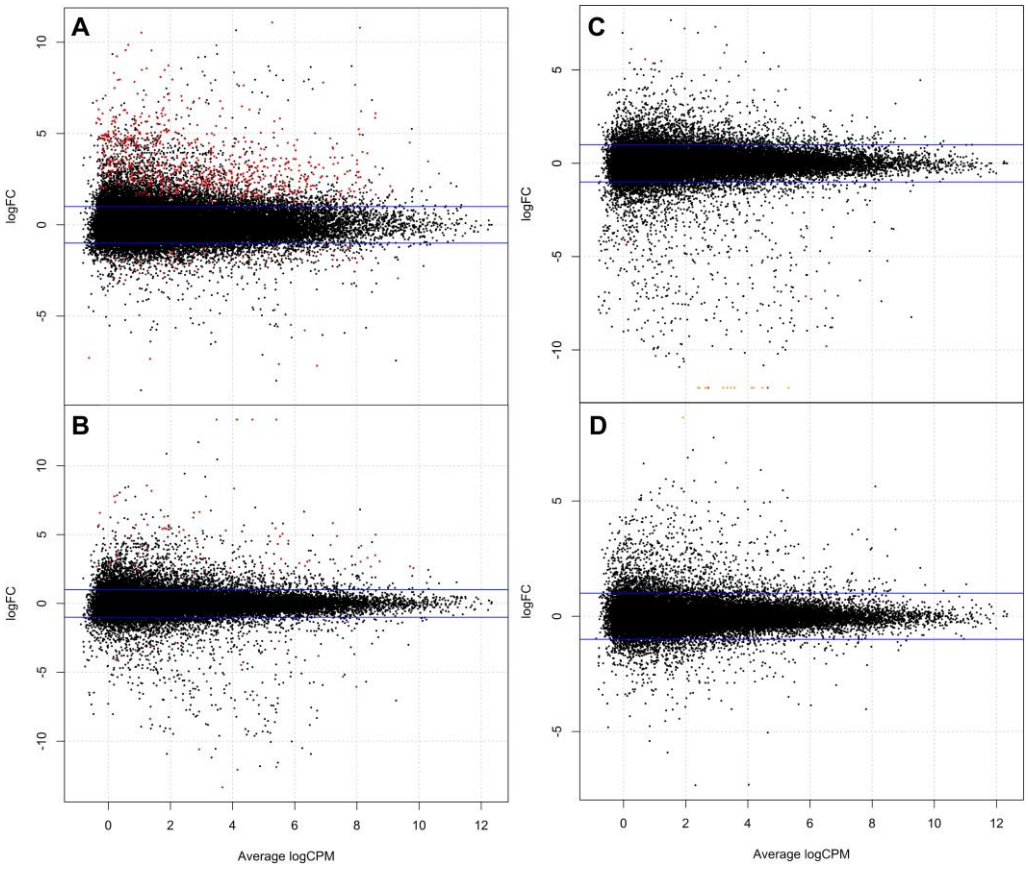

Fig. S4. MA plot of differential expression between strains (A) in diapause maintenance, (B) on day 1, and (C) on day 7, shown as log-fold change of expression versus average log counts per million. Red dots are significantly differentially expressed at FDR < 0.01.

Table S1. Enriched GO Terms and KEGG Pathways for Differentially Expressed Genes Between E and Z Strains (*P*-value < 0.001)

Category	Term	Count	Size	*P*-value
Day1				
GO: biological process	proteasome-mediated ubiquitin-dependent protein catabolic process	30	83	8.39E-06
GO: biological process	cellular protein catabolic process	43	146	4.24E-05
GO: biological process	protein-DNA complex subunit organization	17	39	5.12E-05
GO: biological process	modification-dependent protein catabolic process	42	145	7.95E-05
GO: biological process	macromolecule catabolic process	54	210	2.44E-04
GO: biological process	nucleosome assembly	8	13	2.92E-04
GO: biological process	DNA endoreduplication	8	13	2.92E-04
GO: biological process	mitotic spindle assembly checkpoint	6	8	3.88E-04
GO: biological process	spindle checkpoint	6	8	3.88E-04
GO: biological process	pyrimidine-containing compound metabolic process	10	20	4.92E-04
GO: biological process	DNA packaging	19	54	5.57E-04
GO: biological process	regulation of imaginal disc-derived wing size	8	14	5.85E-04
GO: biological process	regulation of cell cycle process	39	146	8.57E-04
GO: biological process	regulation of tube size	13	32	8.64E-04
GO: biological process	canonical Wnt signaling pathway	8	21	*1.35E-02**
GO: biological process	regulation of Wnt signaling pathway	15	54	*2.29E-02**
GO: molecular function	glucose transmembrane transporter activity	8	11	5.78E-05
GO: molecular function	copper ion binding	10	17	9.61E-05
GO: molecular function	monosaccharide transmembrane transporter activity	9	17	6.31E-04
GO: molecular function	calcium ion binding	37	135	9.43E-04
KEGG Pathway	Proteasome	23	33	6.69E-12
KEGG Pathway	Ascorbate and aldarate metabolism	9	13	2.82E-05
KEGG Pathway	Retinol metabolism	8	13	2.76E-04
KEGG Pathway	DNA replication	12	26	3.13E-04
KEGG Pathway	Non-homologous end-joining	4	4	6.94E-04
Day7				
GO: biological process	amino sugar metabolic process	18	55	2.27E-07
GO: biological process	chitin-based cuticle development	22	79	2.62E-07
GO: biological process	chitin metabolic process	15	46	2.48E-06
GO: biological process	cuticle chitin catabolic process	3	3	5.82E-04
GO: biological process	regulation of hormone levels	10	43	2.34E-03
GO: biological process	regulation of tube size	10	44	2.81E-03
GO: biological process	regulation of tube architecture, open tracheal system	13	67	3.17E-03
GO: biological process	homophilic cell adhesion	7	25	3.48E-03
GO: biological process	cuticle pattern formation	7	25	3.48E-03
GO: biological process	aminoglycan catabolic process	5	14	4.25E-03
GO: biological process	pyrimidine nucleotide biosynthetic process	4	9	4.33E-03
GO: biological process	glucosamine-containing compound catabolic process	4	9	4.33E-03

GO: biological process	pyrimidine nucleoside monophosphate biosynthetic process	3	5	5.12E-03
GO: biological process	intercellular transport	3	5	5.12E-03
GO: biological process	septate junction assembly	7	28	6.88E-03
GO: biological process	negative regulation of hormone metabolic process	2	2	6.98E-03
GO: biological process	negative regulation of sequestering of calcium ion	2	2	6.98E-03
GO: biological process	detoxification of arsenic-containing substance	2	2	6.98E-03
GO: biological process	organonitrogen compound metabolic process	64	572	7.93E-03
GO: biological process	cellular hormone metabolic process	7	29	8.43E-03
GO: biological process	system process	35	279	9.17E-03
GO: biological process	activation of immune response	3	6	9.60E-03
GO: biological process	'de novo' pyrimidine nucleobase biosynthetic process	3	6	9.60E-03
GO: biological process	pyrimidine nucleoside biosynthetic process	3	6	9.60E-03
GO: molecular function	structural constituent of chitin-based cuticle	6	7	2.32E-06
GO: molecular function	chitin binding	13	39	1.22E-05
GO: molecular function	extracellular matrix structural constituent	7	13	3.64E-05
GO: molecular function	structural constituent of chitin-based larval cuticle	11	32	4.16E-05
GO: molecular function	serine-type endopeptidase activity	17	69	5.16E-05
GO: molecular function	serine hydrolase activity	20	91	6.74E-05
GO: molecular function	iron ion binding	16	67	1.28E-04
GO: molecular function	monosaccharide transmembrane transporter activity	7	17	3.04E-04
GO: molecular function	molybdenum ion binding	3	3	6.34E-04
KEGG Pathway	Drug metabolism - other enzymes	8	29	4.35E-04

* Asterix indicates *P*-value <0.05

Table S2. Gene enrichment analysis table for the seven significant STEM profiles (Bonferroni corrected *P*-value cutoff ≤ 0.0001)

Strain	Profile	GO Category Name	#Genes Category	#Genes Assigned	#Genes Expected	#Genes Enriched	p-value	Corrected p-value	Fold
E	1	single-organism developmental process	1967	83	30.8	52.2	2.70E-15	1.10E-11	2.7
E	1	macromolecule metabolic process	1960	82	30.6	51.4	6.30E-15	2.70E-11	2.7
E	1	regulation of biological process	2220	84	34.7	49.3	6.10E-13	2.60E-09	2.4
E	1	regulation of cellular process	2066	79	32.3	46.7	1.70E-12	7.30E-09	2.4
E	1	organic substance metabolic process	2606	91	40.7	50.3	5.40E-12	2.30E-08	2.2
E	1	endopeptidase activity	168	19	2.6	16.4	2.90E-11	1.20E-07	7.2
E	1	proteolysis	330	26	5.2	20.8	3.10E-11	1.30E-07	5
E	1	anatomical structure development	1064	49	16.6	32.4	6.00E-11	2.50E-07	2.9
E	1	homophilic cell adhesion via plasma membrane adhesion molecules	29	9	0.5	8.5	4.20E-10	1.80E-06	19.8
E	1	developmental process involved in reproduction	520	31	8.1	22.9	4.70E-10	2.00E-06	3.8
E	1	anatomical structure morphogenesis	900	42	14.1	27.9	9.00E-10	3.80E-06	3
E	1	proteasome regulatory particle	22	8	0.3	7.7	9.40E-10	4.00E-06	23.3
E	1	plasma membrane	417	27	6.5	20.5	1.00E-09	4.30E-06	4.1
E	1	cellular macromolecule metabolic process	1726	63	27	36	1.60E-09	6.90E-06	2.3
E	1	single organism reproductive process	548	31	8.6	22.4	1.70E-09	7.00E-06	3.6
E	1	macromolecule catabolic process	214	19	3.3	15.7	1.80E-09	7.70E-06	5.7
E	1	negative regulation of cellular process	756	37	11.8	25.2	2.50E-09	1.00E-05	3.1
E	1	peptidase activity, acting on L-amino acid peptides	268	21	4.2	16.8	2.50E-09	1.10E-05	5
E	1	primary metabolic process	2469	80	38.6	41.4	2.70E-09	1.10E-05	2.1
E	1	protein binding	1272	51	19.9	31.1	2.80E-09	1.20E-05	2.6
E	1	ubiquitin-dependent protein catabolic process	131	15	2	13	2.80E-09	1.20E-05	7.3
E	1	peptidase activity	274	21	4.3	16.7	3.80E-09	1.60E-05	4.9
E	1	modification-dependent macromolecule catabolic process	136	15	2.1	12.9	4.80E-09	2.00E-05	7.1
E	1	modification-dependent protein catabolic process	136	15	2.1	12.9	4.80E-09	2.00E-05	7.1
E	1	proteolysis involved in cellular protein catabolic process	136	15	2.1	12.9	4.80E-09	2.00E-05	7.1
E	1	cellular component organization	1534	57	24	33	5.10E-09	2.10E-05	2.4
E	1	protein catabolic process	99	13	1.5	11.5	5.90E-09	2.50E-05	8.4
E	1	proteasome-mediated ubiquitin-dependent protein catabolic process	83	12	1.3	10.7	7.40E-09	3.10E-05	9.2
E	1	proteasomal protein catabolic process	84	12	1.3	10.7	8.50E-09	3.60E-05	9.1
E	1	cellular developmental process	1249	49	19.5	29.5	1.10E-08	4.60E-05	2.5
E	1	intracellular part	3372	98	52.7	45.3	1.20E-08	5.00E-05	1.9
E	1	negative regulation of biological process	844	38	13.2	24.8	1.40E-08	5.90E-05	2.9
E	1	calcium-dependent cell-cell adhesion via plasma membrane cell adhesion molecules	21	7	0.3	6.7	2.20E-08	9.30E-05	21.3
E	2	nucleus	1056	48	17.1	30.9	4.80E-10	2.00E-06	2.8
E	2	protein catabolic process	99	14	1.6	12.4	9.20E-10	3.90E-06	8.7
E	2	single-organism developmental process	1967	71	31.9	39.1	1.20E-09	5.10E-06	2.2
E	2	modification-dependent macromolecule catabolic process	136	15	2.2	12.8	7.80E-09	3.30E-05	6.8
E	2	modification-dependent protein catabolic process	136	15	2.2	12.8	7.80E-09	3.30E-05	6.8
E	2	proteolysis involved in cellular protein catabolic process	136	15	2.2	12.8	7.80E-09	3.30E-05	6.8
E	2	proteasome-mediated ubiquitin-dependent protein catabolic process	83	12	1.3	10.7	1.10E-08	4.70E-05	8.9
E	2	proteasomal protein catabolic process	84	12	1.4	10.6	1.30E-08	5.40E-05	8.8
E	2	anatomical structure development	1064	45	17.3	27.7	1.40E-08	5.90E-05	2.6
E	2	cellular response to DNA damage stimulus	234	19	3.8	15.2	1.40E-08	6.10E-05	5
E	2	macromolecule catabolic process	214	18	3.5	14.5	2.00E-08	8.40E-05	5.2

E	3	oxidoreductase activity	380	42	6.1	35.9	3.00E-22	1.30E-18	6.9
E	3	oxidation-reduction process	325	33	5.2	27.8	1.00E-16	4.30E-13	6.3
E	3	single-organism metabolic process	1304	61	20.9	40.1	4.60E-13	1.90E-09	2.9
E	3	integral component of membrane	613	36	9.8	26.2	6.00E-11	2.50E-07	3.7
E	3	organic acid metabolic process	264	23	4.2	18.8	9.10E-11	3.80E-07	5.4
E	3	oxoacid metabolic process	264	23	4.2	18.8	9.10E-11	3.80E-07	5.4
E	3	carboxylic acid metabolic process	249	22	4	18	1.80E-10	7.40E-07	5.5
E	3	small molecule metabolic process	498	30	8	22	1.20E-09	5.20E-06	3.8
E	3	carbohydrate metabolic process	207	18	3.3	14.7	9.80E-09	4.10E-05	5.4
E	3	organonitrogen compound metabolic process	431	26	6.9	19.1	1.50E-08	6.20E-05	3.8
E	4	oxidoreductase activity	380	22	3.8	18.2	7.60E-11	3.20E-07	5.8
E	4	oxidation-reduction process	325	20	3.2	16.8	1.80E-10	7.80E-07	6.2
E	4	single-organism catabolic process	190	14	1.9	12.1	9.80E-09	4.10E-05	7.4

Table S3. Significance (FDR) values for differentially expressed genes between E and Z strain moths in long-day conditions

Gene Symbol	Contigs	FDR Day 1	FDR Day 7
Ecdysone			
Usp	comp37290_c0_seq1*; comp10851_c0_seq1; comp108932_c0_seq1	8.26E-03	5.41E-03
Spo	comp30567_c0_seq1	4.63E-07	9.21E-10
Phm	comp24584_c0_seq1	3.76E-06	2.69E-05
Eip75B	comp18210_c0_seq1; comp11010_c0_seq2*	6.48E-03	-
Eip78C	comp122877_c0_seq1; comp151979_c0_seq1	-	-
EcR	comp34369_c0_seq1	1.01E-03	-
Nvd	comp8057_c0_seq2	8.88E-03	2.62E-03
Circadian			
Clk	comp120765_c0_seq1	-	-
Cpo	comp89983_c0_seq1	-	-
Cyc	comp120765_c0_seq1	-	-
Pdp1	comp36140_c0_seq1*; comp15799_c0_seq1*	9.33E-03; 8.35E-04	-
Tim	comp27153_c0_seq1	-	-
Vri	comp23405_c0_seq1	-	-
Per	comp7426_c0_seq1; comp8184_c0_seq1; comp97089_c0_seq1	-	-
Cell Cycle			
Cdk4	comp127995_c0_seq1*; comp233925_c0_seq1*	2.61E-03; 1.22E-04	-
CycB	comp21605_c0_seq1	1.32E-07	5.73E-04
Pcna	comp9566_c0_seq1	2.38E-06	-
Polo	comp21903_c0_seq1*; comp10877_c0_seq1*	2.22E-03; 1.56E-03	-
CycA	comp29198_c0_seq1	-	-
Trbl	comp20015_c0_seq1; comp25406_c0_seq1	-	-
Heat Shock			
Hsp83	comp22537_c0_seq1	5.03E-04	-

Hsp60	comp19138_c0_seq1	8.82E-03	-
Hsp70Bc	comp418731_c0_seq1; comp427912_c0_seq1	-	-
Hsp70Ab	comp9675_c0_seq1	-	-
Canonical Wnt Signaling			
Fz3	comp15297_c0_seq1	3.93E-03	-
Arm	comp17057_c0_seq1	9.38E-03	-
Smo	comp14234_c0_seq1	7.71E-03	-
Wg	comp13981_c0_seq1	1.91E-03	-
Wnt6	comp25466_c0_seq1	-	-
Hs6st	comp78453_c0_seq1	-	-
Gro	comp8218_c0_seq1	-	-
Pan	-	7.04E-03	-
Insulin			
Akt1	comp18543_c0_seq1; comp38567_c0_seq1; comp53008_c0_seq1; comp10465_c0_seq1	-	-
Foxo	comp19046_c0_seq1	-	-
InR	comp25612_c0_seq1; comp14287_c0_seq1; comp16806_c0_seq1; comp22132_c0_seq1; comp223642_c0_seq1	-	
Pi3k	comp190582_c0_seq1; comp28365_c0_seq1; comp31462_c0_seq1	-	-
Pten	comp6369_c0_seq1; comp7648_c0_seq1	-	-
SNF4y	comp26437_c0_seq1	-	3.11E-03
Tor	comp16536_c0_seq1; comp17899_c0_seq1; comp19143_c0_seq1; comp26489_c0_seq1	-	-

**Some genes had multiple transcripts, asterix denotes which transcript shows differential expression*

Table S4. Short read archive accession numbers

Strain	Days After Long-day Exposure	Sample Name	SRA Accession
E	0	Shortday_E_biologicalrep1	SRX1206058
E	0	Shortday_E_biologicalrep2	SRX1206872
E	0	Shortday_E_biologicalrep3	SRX1211504
E	0	Shortday_E_biologicalrep4	SRX1211546
E	0	Shortday_E_biologicalrep5	SRX1211572
Z	0	Shortday_Z_biologicalrep1	SRX1211612
Z	0	Shortday_Z_biologicalrep2	SRX1211648
Z	0	Shortday_Z_biologicalrep3	SRX1211686
Z	0	Shortday_Z_biologicalrep4	SRX1211701
Z	0	Shortday_Z_biologicalrep5	SRX1211703
E	1	Longday1_E_biologicalrep1	SRX1211704
E	1	Longday1_E_biologicalrep2	SRX1211706
E	1	Longday1_E_biologicalrep3	SRX1211707
E	1	Longday1_E_biologicalrep4	SRX1211711
E	1	Longday1_E_biologicalrep5	SRX1211709
Z	1	Longday1_Z_biologicalrep1	SRX1211714
Z	1	Longday1_Z_biologicalrep2	SRX1211718
Z	1	Longday1_Z_biologicalrep3	SRX1211716
Z	1	Longday1_Z_biologicalrep4	SRX1211720
Z	1	Longday1_Z_biologicalrep5	SRX1211723
E	7	Longday7_E_biologicalrep1	SRX1211725
E	7	Longday7_E_biologicalrep2	SRX1211741
E	7	Longday7_E_biologicalrep3	SRX1211744
E	7	Longday7_E_biologicalrep4	SRX1211746
E	7	Longday7_E_biologicalrep5	SRX1211748
Z	7	Longday7_Z_biologicalrep1	SRX1211750
Z	7	Longday7_Z_biologicalrep2	SRX1211752
Z	7	Longday7_Z_biologicalrep3	SRX1211782
Z	7	Longday7_Z_biologicalrep4	SRX1211784
Z	7	Longday7_Z_biologicalrep5	SRX1211786