

## RESEARCH ARTICLE

# Environmental history impacts gene expression during diapause development in the alfalfa leafcutting bee, *Megachile rotundata*

George D. Yocum<sup>1,\*</sup>, Anna K. Childers<sup>1,2</sup>, Joseph P. Rinehart<sup>1</sup>, Arun Rajamohan<sup>1</sup>, Theresa L. Pitts-Singer<sup>3</sup>, Kendra J. Greenlee<sup>4</sup> and Julia H. Bowsher<sup>4</sup>

## ABSTRACT

Our understanding of the mechanisms controlling insect diapause has increased dramatically with the introduction of global gene expression techniques, such as RNA sequencing (RNA-seq). However, little attention has been given to how ecologically relevant field conditions may affect gene expression during diapause development because previous studies have focused on laboratory-reared and -maintained insects. To determine whether gene expression differs between laboratory and field conditions, prepupae of the alfalfa leafcutting bee, *Megachile rotundata*, entering diapause early or late in the growing season were collected. These two groups were further subdivided in early autumn into laboratory- and field-maintained groups, resulting in four experimental treatments of diapausing prepupae: early and late field, and early and late laboratory. RNA-seq and differential expression analyses were performed on bees from the four treatment groups in November, January, March and May. The number of treatment-specific differentially expressed genes (97 to 1249) outnumbered the number of differentially regulated genes common to all four treatments (14 to 229), indicating that exposure to laboratory or field conditions had a major impact on gene expression during diapause development. Principle component analysis and hierarchical cluster analysis yielded similar grouping of treatments, confirming that the treatments form distinct clusters. Our results support the conclusion that gene expression during the course of diapause development is not a simple ordered sequence, but rather a highly plastic response determined primarily by the environmental history of the individual insect.

**KEY WORDS:** RNA-seq, Dormancy, Omnigenic, Toolkit

## INTRODUCTION

The ability to survive seasonal environmental changes (e.g. temperature, humidity, nutrient availability) is critical for organisms to complete their life cycles, thereby influencing their ability to occupy certain geographic ranges. Representatives of all animal and plant taxa have evolved some form of dormancy in their life history in response to seasonal changes. Although the terminology differs by taxa (e.g. hibernation, aestivation, diapause), the outcome is the same: organisms alter their physiology to survive harsh conditions.

Within Insecta, the physiological form of dormancy is known as diapause. Most insects that undergo diapause in the temperate zone have a narrow window of a few months from early spring to late summer in which to complete development. Across species, diapause occurs in a diverse range of genetically determined developmental stages, ranging from early blastoderm embryo to adult. However, among the various stages, the diapause phenotype is strikingly similar. Arrested development, cell-cycle arrest, decreased metabolism, increased metabolic reserves and higher stress tolerance all characterize insect diapause (Tauber et al., 1986; Danks, 1987; Denlinger, 2002).

Prior to the initiation of diapause, there is a distinct physiological stage known as the prediapause phase. During the prediapause phase, the insect receives the environmental cues forewarning of future adverse environmental conditions and makes preparations for entering diapause. These preparations may include accumulation of metabolic reserves, migration, or selection of an environmentally buffered site to spend the duration of the adverse season. After the prediapause period, the insect enters into diapause. Diapause is further subdivided into three phases, initiation, maintenance and termination. Transition through these phases is known as diapause development (Košťál, 2006). During the initiation phase, direct development stops and metabolism is suppressed. The maintenance phase is characterized by endogenous blocking of direct development even under conditions favorable for development. During the termination phase, control of direct development is switched from being endogenously to exogenously controlled, i.e. from hormonally to thermally controlled. Diapause commonly terminates before the period of adverse environmental conditions have ended, whereupon the insect enters into a period referred to as post-diapause quiescence. During this period, an insect retains many aspects of diapause such as decreased metabolism and increased tolerance to environmental stress, but will slowly regain the ability for normal development as environmental factors return to more favorable conditions. Although these divisions of diapause development are useful for referring to a specific suite of physiological functions, they are also somewhat misleading, because diapause development does not abruptly shift from one phase to another but instead progresses along a continuum (Sawyer et al., 1993; Košťál, 2006; Košťál et al., 2017). Another factor impeding the development of a comprehensive understanding of diapause physiology is that currently the majority of our understanding of diapause is derived from laboratory investigations. But insects diapausing under field conditions are subjected to a wide range of environmental factors that vary within and between seasons in both duration and severity. This raises the question as to the degree to which environmental factors shape diapause physiology (as reflected in the transcriptome response) over the course of overwintering.

Global gene expression techniques have greatly expanded our ability to understand the molecular regulation of diapause (Emerson

<sup>1</sup>USDA-ARS Insect Genetics and Biochemistry Research Unit, Fargo, ND 58102, USA. <sup>2</sup>USDA-ARS Bee Research Lab, Beltsville, MD 20705, USA. <sup>3</sup>USDA-ARS Pollinating Insect-biology, Management, Systematics Research Unit, Logan, UT 84322, USA. <sup>4</sup>Biological Sciences, North Dakota State University, Fargo, ND 58105, USA.

\*Author for correspondence (george.yocum@ars.usda.gov)

© G.D.Y., 0000-0001-6285-0623

et al., 2010; Ragland et al., 2010, 2011; Poelchau et al., 2011, 2013a,b; Gong et al., 2013; Dong et al., 2014; Huang et al., 2015; Qi et al., 2015; Yocum et al., 2015; Hao et al., 2016; Kang et al., 2016; Meyers et al., 2016; Košťál et al., 2017). Observed similarities in specific gene expression patterns between species has led a number of research teams to propose the existence of a conserved 'diapause genetic toolkit' (Ragland et al., 2010; Poelchau et al., 2013a; Huestis and Lehmann, 2014; Amsalem et al., 2015; Yocum et al., 2015; Ragland and Keep, 2017). Others have argued that if a diapause toolkit exists, it is conserved at the pathway level and not in expression patterns of individual transcripts (Ragland et al., 2010; Košťál et al., 2017). A significant challenge in testing these hypotheses is that environmental signals modify the physiological and chronological progression of diapause (Košťál, 2006). If the physiological processes underlying the phases of diapause development are impacted by the interaction of environmental factors, gene expression patterns should also reflect this complexity, thereby compounding the technical challenge of identifying any possible diapause toolkit.

The alfalfa leafcutting bee, *Megachile rotundata* (Fabricius 1787) (Hymenoptera: Megachilidae), is an intensely managed solitary bee used as the primary pollinator of alfalfa for seed production in North America (Pitts-Singer and Cane, 2011). Under agricultural management, *M. rotundata* adults emerge from leaf-wrapped cocoons in late June or early July, after spending the winter as prepupae. Each female constructs a series of brood cells in artificial nest blocks consisting of linear cavities. Once a cell is provisioned with pollen and nectar, the female will lay a single egg, seal the cell and then start to construct the next cell. The egg hatches, and the larva eats its provision mass, completes development within the cell, and either enters diapause as a prepupa or continues development to the adult stage for summer emergence (Kronic, 1972; Hobbs and Richards, 1979; Kemp and Bosch, 2001). The progeny of summer-emerging adults (second generation) are generated in the late summer and develop to the diapausing prepupal stage. These two brood cohorts (those entering diapause in June/July and those entering diapause in August/September) have very different thermal histories. The prepupae that enter diapause early in the field season are subjected to summer temperatures, whereas those that diapause late in the season experience only autumn temperatures before the onset of winter.

The objective of this investigation was to determine the degree to which environmental history impacts gene expression during the course of diapause development in *M. rotundata*. To achieve our objective, early and late diapausing prepupae were collected, and in the late autumn, the early and late diapausing groups were further separated into two groups: one that remained outdoors in ambient conditions and one that was maintained under constantly cool laboratory conditions. RNA sequencing (RNA-seq) and differential expression analyses were performed on all four treatment groups sampled in the autumn, winter and the following spring (March and May).

## MATERIALS AND METHODS

### Insects

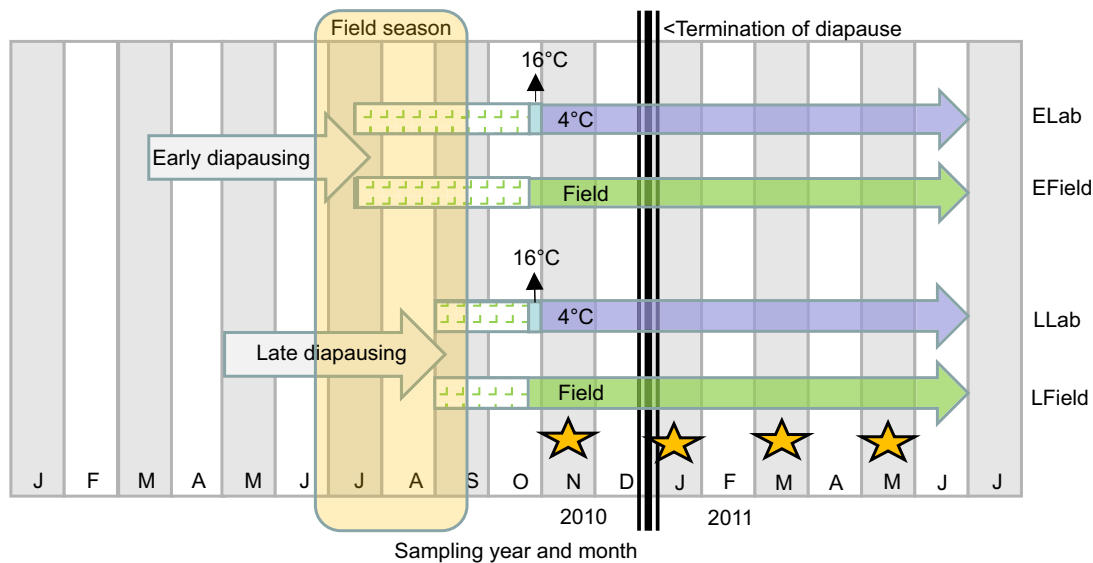
The *M. rotundata* that nested in Utah in the summer of 2009 originated from brood reared in commercial alfalfa fields in Wyoming, USA; the bees were then purchased by a Utah alfalfa seed grower for pollination in summer. Bees were incubated in an on-farm facility and released as adults on 26 June 2010 into a Utah alfalfa field (41°47'37.04"N; 112°8'18.35"W). Polystyrene nesting blocks (Beaver Plastics, Ltd, Alberta, Canada) had been placed in

field domiciles for bee nesting (Pitts-Singer and Cane, 2011), and in several of the blocks, paper straws were inserted into holes as cavity liners. Twice weekly, straws that had been filled with nest cells were removed from the holes and empty straws were inserted into vacated holes. Straws were brought into the USDA-ARS Pollinating Insects Research Unit (Logan, UT, USA), and the date when straws were recovered was recorded. Straws were aligned and adhered to plastic boards to secure the straws for placement into a digital X-ray machine (Faxitron 43804N, Faxitron Bioptics, Tucson, AZ, USA) and to align digital images with the actual nest cells. Nests on boards were then stored outdoors in an open shelter (protected from direct sunlight and rain), where the temperature was recorded with a HOBO Datalogger (Fig. S1; Onset Computer Corp., Bourne, MA, USA). Nests were X-radiographed weekly throughout the summer to determine the developmental stages of the bees still in their nest cells. Development to the pupal stage indicated that those individual bees had averted diapause and were destined to metamorphose to adulthood without spending the winter in the prepupal (diapausing) stage.

For this experiment, bees that entered diapause early and late in the nesting season were needed (Fig. 1). Early season nests were designated as those from straws pulled between 30 June and 19 July 2010. Late season nests were designated as those pulled on 1 September 2010. Based on the female bee's lifespan of 7 to 8 weeks, the early season diapausing progeny were generated by the adults of the 2009 brood, but the late season nests most likely were generated by non-diapausing bees that emerged during early summer 2010 (Pitts-Singer and Cane, 2011). For early season nests, only those in which all cells contained diapausing prepupae were used. Bee cells were dissected from nests in September 2010, transferred to individual gelatin capsules, and tracked using unique identifiers. Early and late season groups were divided into two overwintering management groups: (1) laboratory bees, overwintered in the laboratory according to standard commercial storage practices; and (2) field bees, overwintered outside at ambient temperatures while shielded from direct sunlight or rain. Therefore, four treatment groups were designated: (1) early season, field managed (EField), (2) early season, laboratory managed (ELab), (3) late season, field managed (LField) and (4) late season, laboratory managed (LLab). To assure that bees assigned to treatments were genetically diverse, prepupae from the same nest were distributed between field and laboratory treatments for both the early and late diapausing groups. All individual bee cells were maintained outdoors at ambient temperature until 22 October 2010. On that date, early and late season bee cells that were designated for laboratory-managed groups were moved to an incubator held at a constant 16°C, and on 1 November 2010 were placed into an environmental chamber maintained at 4–5°C and in darkness. Temperature continued to be recorded outdoors for the duration of the study (Fig. S1). Monthly from October 2010 to June 2011, we removed prepupae for future RNA extraction from each of the four treatment groups. Twenty whole prepupae per treatment were excised from their cocoons, placed individually into Eppendorf tubes, flash-frozen in liquid nitrogen and stored at –80°C.

### RNA

Samples from four time points (November, January, March and May) were chosen to span the diapause maintenance (November), termination (January) and post-diapause quiescent stages (March and May) of development (Johansen and Eves, 1973; Yocum et al., 2006). RNA was extracted from three prepupae haphazardly chosen



**Fig. 1. Schematics of experimental design.** *Megachile rotundata* diapausing prepupae were collected early and late in the nesting season. All diapausing prepupae were maintained under field conditions until October when the prepupae from each group [early (E) and late (L)] were further subdivided into laboratory or field management types resulting in four treatment groups: early laboratory, late laboratory, early field and late field. RNA-seq libraries were constructed from samples taken from each group in November, January, March and May (stars) for paired-end Illumina sequencing.

from samples for each of the four treatments and months (48 total samples). Bees were ground in liquid nitrogen, combined with 1 ml of TRIzol (Invitrogen, Life Technologies, Grand Island, NY, USA), and extracted according to the manufacturer's instructions. The isolated RNA pellets were stored under absolute ethyl alcohol at  $-80^{\circ}\text{C}$  until needed.

### Illumina sequencing

Illumina sequencing was carried out by the Georgia Genomic Facility of the University of Georgia. Stranded Illumina libraries were constructed using the TruSeq RNA kit (Illumina Inc., San Diego, CA, USA). Two lanes of 100 bp paired-end sequencing were performed on each RNA sample using an Illumina HiSeq 2000 sequencer.

### RNA-seq quality trimming, assembly and differential gene expression analysis

Raw 100 bp Illumina reads were quality trimmed using a Phred quality score cut-off of 19 with DynamicTrim (SolexaQA\_v.2.2), and only those reads with a minimum length of 50 bp were retained using LengthSort (SolexaQA\_v.2.2) (Cox et al., 2010).

Transcript assembly and differential gene expression analyses were carried out using the Tuxedo Suite (Trapnell et al., 2012). Based on an Agilent 2100 Bioanalyzer report (Agilent Technologies, Santa Clara, CA, USA) on the sequenced libraries, we calculated the  $-\text{mate-inner-dist}$  and the  $-\text{mate-std-dev}$  parameters for TopHat2 (v2.0.10) (Kim et al., 2013) transcript alignment to the *Mrot\_1.0* genome assembly (Kapheim et al., 2015). The inner distance between the mate pairs was calculated to be 0 bp by subtracting the adapter length ( $\sim 125$  bp) and the length of the paired-end reads (200 bp) from the average size of the Illumina libraries (324 bp). Because the size distribution of the library had a right skew, we used a conservative standard deviation of 24 bp, one-third the calculated standard deviation of the library. In addition, because the libraries were stranded, we used the library-type fr-first strand parameter in TopHat2. Data from both Illumina lanes for each sample were pooled. Next, Cufflinks (v2.1.1;

<http://cole-trapnell-lab.github.io/cufflinks/>) was used to generate transcripts from the TopHat2 output for each sample. Then all transcripts were combined using Cuffmerge (v1.0.0) to generate a single reference transcript assembly. This reference transcript assembly and each sample's individual output from TopHat2 was input into Cuffdiff (v2.1.1), which performed pairwise comparisons between all 48 samples for differential gene expression analyses. The completeness of the transcript assembly was estimated using BUSCO software (Benchmarking Universal Single-Copy Orthologs; v1.1b1) (Simão et al., 2015) with the arthropod reference gene set (released December 2014, contained 2675 genes).

### Assignment of gene descriptions

In order to minimize uninformative annotations (such as 'hypothetical protein'), we conducted three rounds of BLAST. In the first round, NCBI's BLASTX (version 2.2.30+) (Camacho et al., 2009) was used to align all transcripts assembled by Cufflinks to RefSeq proteins (downloaded 29 January 2015 with 21,778 entries) from *Apis mellifera*, the most closely related well-annotated species, keeping only the best hit with an E-value cut-off of  $1 \times 10^{-6}$ . Any transcripts with a poor alignment (hypothetical, uncharacterized or unnamed proteins) or without an alignment were aligned to NCBI's nr database (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>; downloaded 29 January 2015) using an E-value cut-off of  $1 \times 10^{-6}$  and the `negative_glist` option to exclude hypothetical, uncharacterized or unnamed proteins (GI numbers downloaded 13 February 2015). Finally, for those remaining unidentified transcripts, a third round of BLAST was performed against the full nr database using an E-value cut-off of  $1 \times 10^6$ , which allowed less informative descriptions (hypothetical, uncharacterized or unnamed proteins) in order to preserve knowledge of known homologs.

### GO term and KEGG pathway enrichment analysis

Gene ontology (GO) terms were assigned using Blast2GO Pro (Conesa et al., 2005). KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways (December 2015 release; Kanehisa and Goto, 2000) were assigned using the DAVID Knowledgebase (DAVID



6.8; Huang et al., 2009) based on the protein GI identifiers obtained from protein alignments using BLASTX (see above) with all species used for annotations. Because some genes had multiple transcripts, identifiers from proteins aligned to all transcripts for a given gene were included in the analysis. Genes expressed over the average threshold of all genes in the hierarchical cluster analysis (see below) that determined the formation of the primary clusters were used in gene category over-representation analyses using the GOseq package (version 1.18.0) (Young et al., 2010) in R (version 3.1.3, <http://www.R-project.org/>). Enrichment analyses were run on the gene level to remove bias owing to variable numbers of transcripts for each gene. Each gene's length was taken as the median of the lengths of all transcripts. All GO terms and KEGG pathways associated with any of the gene's transcripts were mapped to the gene feature. A 0.05 false discovery rate (FDR) cut-off (Benjamini and Hochberg, 1995) was applied for multiple testing correction.

### Statistical analysis and Venn diagrams

Principle component analysis (PCA) and hierarchical cluster analysis were carried out using JMP (v. 12.0, SAS Institute, 2015, Cary, NC, USA). The data, consisting of 13,564 genes and 16 treatment groups (described above), were log-transformed and first subjected to PCA. Following the PCA, hierarchical cluster analysis by both Ward- and average-linkage methods was performed on normal standardized and robustly standardized data (Eisen et al., 1998). Satisfactory cluster separation by treatment variables was noted with the Ward-linkage method and normal standardization. The effects of overwintering site (laboratory or field), month (month the samples were collected; November, January, March or May) and timing of diapause initiation (early or late) noted in the hierarchical cluster analysis were further confirmed using linear modeling in JMP Genomics 8.2 (SAS Institute). Standardized and log-transformed cluster data derived from the cluster analysis were subject to ANOVA in JMP Genomics. Venn diagrams were generated using Venny 2.0 (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>).

## RESULTS

### Transcriptome assembly and annotation

Sequencing resulted in a total of 409,970,200 raw reads for all 48 samples, with an average of 8,541,045 raw reads per sample. On average, 14% of the raw reads for each sample were discarded in the quality trimming stage. Quality-trimmed RNA-seq reads had an average 87.8% overall and 84.3% concordant pair alignment rate to the *M. rotundata* genome assembly (Table S1). Using the data from all samples, we assembled a transcriptome representing 13,566 gene loci with 50,081 total transcripts (isoforms). Based on the BUSCO transcriptome completeness assessment, more than 92% of *M. rotundata* protein-coding genes were transcribed during the developmental periods evaluated (86.8% of expected genes were complete, another 5.9% were fragmented and 7.3% were missing).

The staged BLAST approach resulted in 85.7% of the transcripts being assigned a descriptive annotation (the description was not derived from hypothetical, uncharacterized and unnamed proteins) and another 1.8% of the transcripts had similarity to hypothetical proteins.

Blast2GO associated a total of 81,545 GO terms with 29,387 of the transcripts (58.7%), such that at least one of the transcripts from 5847 of the genes (43.1%) had a GO annotation. DAVID associated a total of 3372 KEGG pathways with 7165 transcripts (14.3%), such that at least one of the transcripts from 1623 of the genes (12%) was associated with a KEGG pathway. Within the full

data set, 2066 unique GO terms and 131 unique KEGG pathways are represented.

### Differential gene expression analysis

With the current experimental design, two types of comparisons could be carried out. We first compared all biologically relevant combinations within each sampling month (EField versus LField, ELab versus LLab, ELab versus EField, and LLab versus LField) (Table S2). The number of differentially regulated genes varied from as low as 104 upregulated genes in the November ELab versus LLab comparison to 1273 downregulated genes in the May ELab versus EField comparison. The percentage of differentially regulated genes varied from 0.77 to 9.39%, with the largest differences seen in the May comparisons.

The second type of comparison conducted was a sequential, month-to-month analysis to characterize gene expression during diapause development by treatment. Within each treatment, the November, January and March samples were used as the reference for determining differentially regulated genes in the succeeding months. This resulted in three comparisons with November as the reference group (January, March and May), two for January (March and May) and one for March (May). In these six between-month comparisons, the total number of differentially regulated genes for the combined four treatments varied from 722 (January to March comparison) to 2515 (November to May comparison) (Table S2). Venn diagrams of these between-month differentially regulated genes revealed that only 0.9% (March to May) to 14.5% (January to March) of the differentially regulated genes were shared between all four treatments (Figs 2 and 3). Considering all of the comparisons, a total of 566 genes (406 non-redundant) were shared between all four treatments, and only 130 of these genes were common to more than one comparison (Table S3). The total number of treatment-specific genes (upregulated plus downregulated) ranged from 97 in the November to January EField comparison to 1249 in the March to May EField comparison.

### Principal component analysis

Principal component analysis was carried out to determine the relationship between the 16 experimental groups (four treatments at each of four time points) (Fig. 4A). The first two principle components (PCs) explained 89.74% of the total variance. PC1 accounted for 62.53% of the total variance, was explained by the months November, January and March, and encompassed the experimental groups including ENovField, EJanLab, LJanLab, LJanField, EMarLab, EMarField and LMayLab. PC2 accounted for 27.21% of the total variance, was explained by the experimental groups ENovLab, LNovLab, LNovField and EMayField. These four treatments form two distinct clusters: the November laboratory cluster (ENovLab and LNovLab) and the May field cluster (EMayField and LMayField). Both of these clusters were negatively correlated with the primary cluster on the PC1 axis. The May field cluster was negatively correlated with each of the other clusters on the PC2 axis.

### ANOVA of differentially regulated genes

To determine which factor(s) had a statistically significant impact on gene expression, ANOVA was performed (Table 1). Over the course of diapause development, the overwintering site (field or laboratory) and month the samples were collected had a significant impact on gene expression ( $F_{1,1760}=233.71$ ,  $P=0.00$  and  $F_{3,1760}=44.32$ ,  $P=0.00$ , respectively). But, timing of diapause initiation (early or late) was not significant ( $F_{1,1760}=0.03$ ,  $P=0.87$ ). The interaction term for the

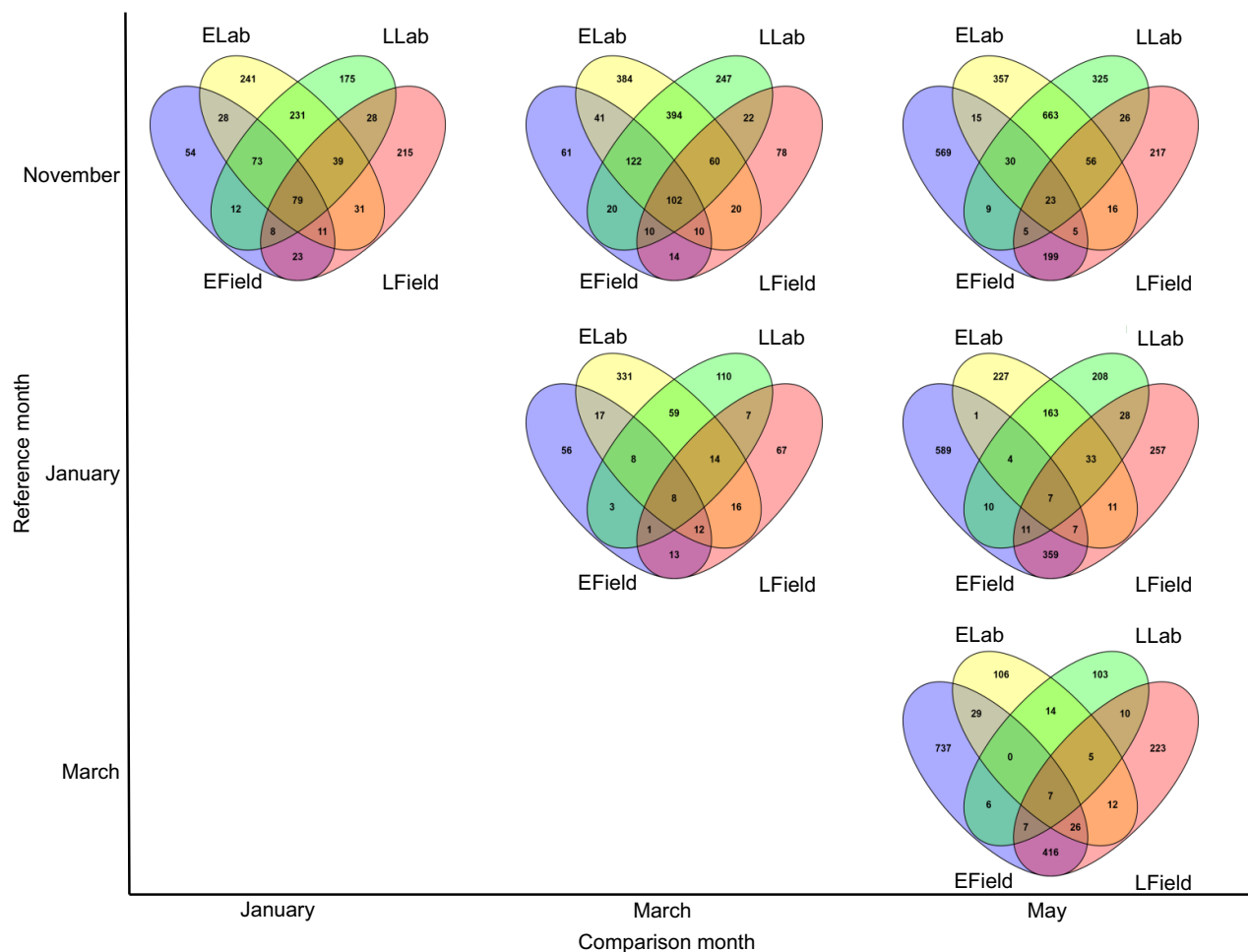


overwintering site by month was significant ( $F_{3,1760}=45.30$ ,  $P=0.00$ ). No interactions that incorporated the timing of diapause initiation were significant (timing  $\times$  month,  $F_{3,1760}=0.04$ ,  $P=0.99$ ; timing  $\times$  overwintering site,  $F_{1,1760}=0.04$ ,  $P=0.85$ ; or timing  $\times$  month  $\times$  overwintering site,  $F_{3,1760}=0.00$ ,  $P=1.0$ ). The adjusted  $R^2$  value for the analysis was 0.2432.

The grouping of ENovLab, LNovLab, EMayField and LMayField into supercluster 1 was an unexpected result from the hierarchical cluster analysis. To determine the physiological basis for this grouping, enrichment analysis was carried out. Genes contributing to the formation of supercluster 1 were enriched for 17 over-represented GO terms and three KEGG pathways. The majority of the enriched GO terms were associated with ribosomal function, translation and ATP synthesis (Table S4), whereas the KEGG pathways included the oxidative phosphorylation, proteasome and ribosome pathways (Table S4). There were no under-represented GO terms or KEGG pathways associated with supercluster 1.

Hierarchical cluster analysis of the expression level of all the genes yielded two superclusters and nine clades. Supercluster 1 consisted of clades comprising the treatments separated by PC2, ENovLab, LNovLab (1.1) and EMayField, LMayField (1.2) (Fig. 4B). An alternate method for visualizing results from hierarchical cluster analysis that includes distance comparisons is the constellation plot (Fig. 4C). The constellation plot of the present cluster analysis yielded two main observations. The first is a high degree of similarity between the May field (early and late) samples with the November laboratory (early and late) samples. The second observation is that the most dissimilar within-month comparison is the November laboratory (early and late) versus the November field (early and late) samples. These two treatment groupings are situated at opposite ends of the constellation plot, indicating a low level of similarity.

Because the November field and laboratory samples were the least similar based on the results of the constellation plot, we compared gene expression levels between these groups to determine the physiological basis for this result. Four comparisons were carried out: LNovField to LNovLab, LNovField to ENovLab, ENovField to LNovLab, and ENovField to ENovLab. Only those genes that were differentially regulated in the same direction in all comparisons were kept for further consideration. Between the November field samples



**Fig. 3. Overlap of downregulated genes.** Within each row and treatment group, gene expression status is relative to the reference month (November, January or March). Venn diagrams are labeled according to the timing that the prepupae entered diapause, early (E) or late (L) in the nesting season and the overwintering site, field (Field) or laboratory (Lab).

and the November laboratory samples, 107 genes were differentially regulated (Table S2). Within the differentially regulated genes, 74 were upregulated in the field samples. Notably, five of these differentially regulated genes were regulatory genes: cyclin-dependent kinase 1 (*CDK1*), anaphase-promoting complex subunit 15 (*ANAPC15*), Cactus (*Ikb*), Bicaudal C (*BicC*) and *Samui*.

## DISCUSSION

The goal of this study was to determine whether environmental history affects gene expression over the course of diapause development. We concentrated on two environmental factors, the timing of diapause initiation (early and late) and overwintering conditions (constant or fluctuating temperatures), to address this question. Both of these factors had a major impact on the gene expression patterns. The differences in gene expression patterns between the four treatment groups in both the within- and between-months comparisons strongly suggest that these bees were physiologically different. The impact of environmental history on insect physiology during the course of diapause development presents a significant challenge for identifying those genes that make up a possible 'genetic toolkit'.

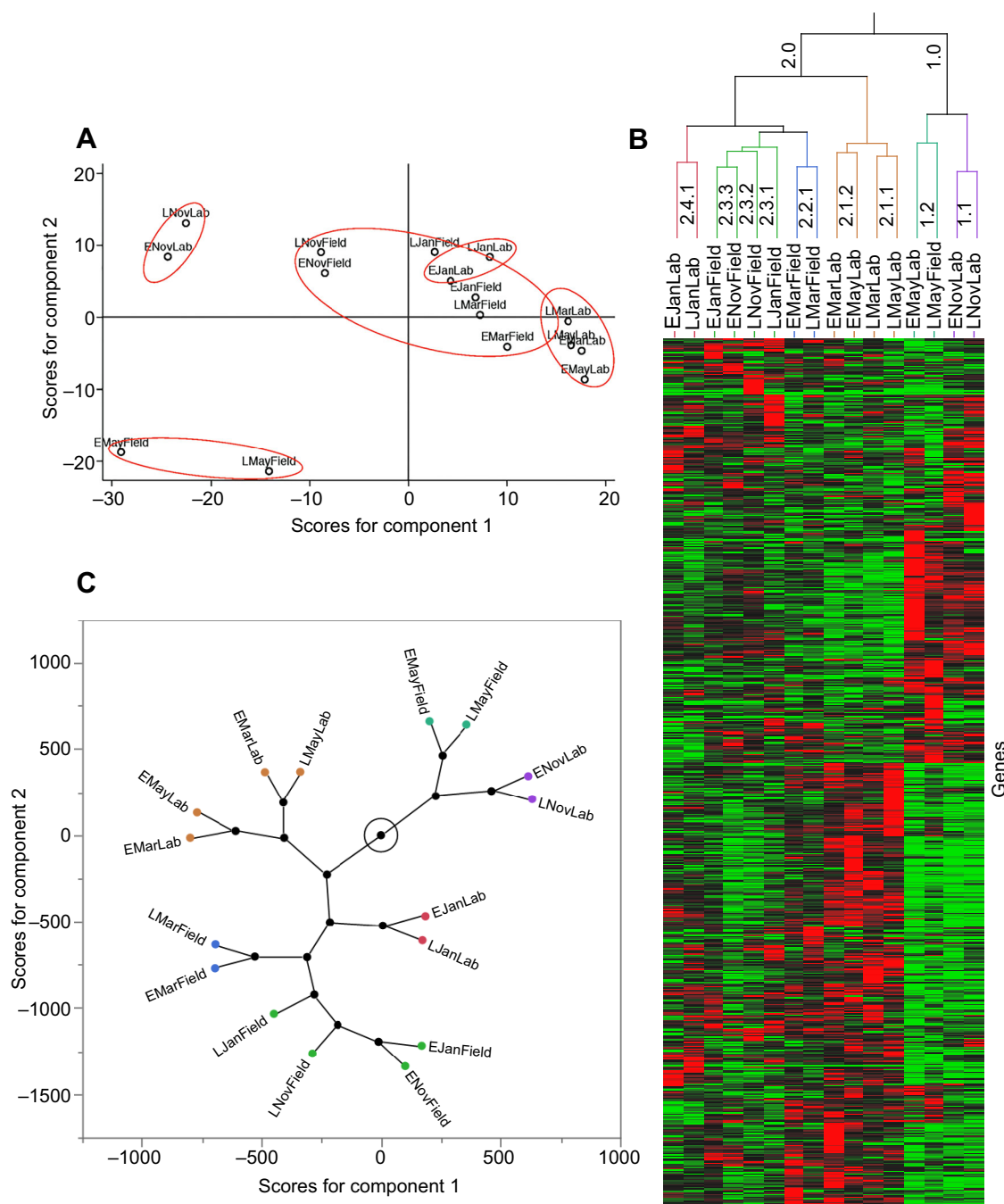
### Influence of early versus late diapause initiation

The analysis of the effect of timing of entering diapause (early or late) on gene expression gave conflicting results. Time was an

insignificant factor in our ANOVA model and the PCA and hierarchical cluster analyses failed to clearly separate the early and late treatments. But, timing of entering diapause clearly impacted gene expression patterns over the course of diapause development (Figs 2 and 3). The shift in gene expression patterns between the early and late samples had occurred by the time of the November sampling. We would argue that although statistically insignificant, timing is not biologically irrelevant. However, based on the ANOVA, PCA and hierarchical cluster analyses, the overwintering conditions (field versus laboratory) and month of sampling are the primary drivers for the physiological separation between the various treatment groups. Yet, these two factors only explained 24% of total variance in the ANOVA model, indicating that there are other unaccounted-for factor(s) impacting gene expression during diapause.

### Impact of field versus laboratory overwintering conditions

Transcriptomes of diapausing *M. rotundata* diverged quickly when transferred from a thermally fluctuating field environment to the constant environment of the laboratory. By the time the November samples were collected, the laboratory samples had been transferred from the field to 16°C for 9 days. During this same period, the field bees were exposed to a range of temperatures (3.7–14.8°C) with a mean of 7.7°C (Fig. S1). The most striking result of the PCA and



**Fig. 4. Multivariate analysis of the impact of environmental history on diapause development in *Megachile rotundata*.** The 16 different treatments are labeled according to the timing that the prepupae entered diapause, early (E) or late (L) in the nesting season, and the overwintering site, field (Field) or laboratory (Lab), as well as the month sampled. Illumina sequencing determined the expression levels for a total of 13,566 genes. The normalized gene expression levels for each gene were employed for the multivariate analysis: principle component analysis score plot (A), geometric dendrogram (B) and constellation plot (C). Constellation plots use the groups' constellation coordinates and distance calculations to give a relative measurement of similarity between the groups. The greater the distance, the less similar the groups. The node circled in the constellation plot is the root node.

hierarchical cluster analyses was the supercluster 1 grouping of the November early and late laboratory treatments with the May early and late field treatments (Fig. 4B). This supercluster was characterized by overrepresentation of genes associated with GO terms and KEGG pathways for increased protein synthesis (ribosome, translation and ATP synthesis). Because bees in the May field treatments experienced temperature pulses above the developmental threshold for *M. rotundata* (~16–18°C) (O'Neill

et al., 2011), they would be expected to have resumed direct development. Therefore, the increase in protein synthesis is reasonable. The indication of a higher level of protein synthesis in the November laboratory treatments is somewhat more perplexing. Like the May field samples, the higher level of protein synthesis in the November laboratory samples may be due to the elevated temperature (16°C) relative to the temperatures experienced by the field prepupae during this same time period. Nevertheless, it



**Table 1. ANOVA results of *Megachile rotundata* diapause regulated genes**

Source	Partial SS	d.f.	MS	F	Prob>F
Model	1249.28	15	83.29	39.03	0.00
Timing <sup>a</sup>	0.05	1	0.05	0.03	0.87
Month <sup>b</sup>	283.74	3	94.58	44.32	0.00
Location <sup>c</sup>	498.73	1	498.73	233.71	0.00
Timing×Month	0.24	3	0.08	0.04	0.99
Timing×Location	0.08	1	0.08	0.04	0.85
Month×Location	290.03	3	96.68	45.30	0.00
Timing×Month×Location	0.11	3	0.00	0.00	1.00
Residual	3755.82	1760	2.13		
Total	5005.10	1775	2.82		

<sup>a</sup>Timing of diapause initiation (early or late).<sup>b</sup>Month samples were collected (November, January, March or May).<sup>c</sup>Location of the overwintering site (laboratory or field).

indicates that the environmental history of the November laboratory treatments made this group physiologically distinct.

Another striking result was the degree of dissimilarity between the November laboratory (early and late) and November field (early and late) treatments, best seen in the constellation plot (Fig. 4C). Because the November treatments were at the start of the experiment, we expected that these groups would have a high degree of similarity. Yet, the November laboratory and field treatments are situated on the opposite ends of the plot, indicating a low level of similarity relative to the other treatment groups. Examination of the differentially regulated genes between the field and laboratory November samples revealed a shift in key regulatory genes associated with diapause.

A universal hallmark of insect diapause is cell cycle arrest (Nakagaki et al., 1991; Tammariello and Denlinger, 1998; Košťál et al., 2009; Poelchau et al., 2011, 2013a; Yocum et al., 2015; Ragland and Keep, 2017). Comparison between the November field and laboratory samples found evidence for differences in this key trait, with the cell cycle regulatory genes *CDK1*, *ANAPC15* and *ING3* being upregulated in the laboratory samples. *CDK1* targets approximately 75 genes that control various aspects of the cell cycle, and therefore is an essential gene for the progression of the cell cycle (Malumbres and Barbacid, 2009; Enserink and Kolodner, 2010). The anaphase-promoting complex (*ANAPC15*) is essential for the initiation of DNA replication and sister chromatid separation during cell proliferation (Peters, 2006). The ING family of tumor suppressor proteins are highly conserved across taxa and are involved in the regulation of cell cycle arrest, DNA repair, regulation of gene expression, apoptosis and chromatin remodeling (Campos et al., 2004; Doyon et al., 2006; Cole and Jones, 2008; Aguiusa-Touré et al., 2011). *ING3* in particular is known primarily to function in cell cycle arrest, apoptosis and chromatin remodeling. However, in prostate cancer, *ING3* has been shown to promote cell proliferation, which demonstrates that it is environmentally dependent in cell cycle progression (Nabbi et al., 2017). Regardless of the physiological role of *CDK1*, *ANAPC15* and *ING3* in diapause, their differential expression lends more evidence to the physiological distinctions between the November field and laboratory samples.

Further evidence for differences in cell cycle regulation between the November field and laboratory samples is indicated by the differential regulation in key regulatory pathways. *Ikb*, an inhibitor of the NF- $\kappa$ B pathway (Belvin et al., 1995; Ray et al., 1995), and *BicC*, an inhibitor of the wnt pathway (Maisonneuve et al., 2009; Park et al., 2016), are both upregulated in the field samples relative to laboratory, indicating that both the NF- $\kappa$ B and wnt pathways are

being downregulated in the field bees. Inhibiting either the NF- $\kappa$ B or the wnt pathway would decrease progression through the cell cycle (Joyce et al., 2001; Niehrs and Acebron, 2012). Taken together, these results indicate that the cell cycle is being differentially regulated between the laboratory and field samples.

Of the differentially regulated genes we identified, the most thoroughly studied diapause-related gene is *Samui*, which was upregulated in the November field samples. *Samui* is a cold-induced gene that has been isolated in *Bombyx mori* (Moribe et al., 2001). *Samui* is a member of the BAG family of proteins (Moribe et al., 2001), which are highly conserved and have been demonstrated to interact with signaling pathways and molecular chaperones (Doong et al., 2002; Kabbage and Dickman, 2008). Based on the correlation between the temperature-dependent profile of *Samui* expression and termination of diapause, as well as the timing of *Samui* expression early during low temperature exposure, it was proposed that *Samui* serves as a trigger for terminating diapause in *B. mori* (Moribe et al., 2001, 2002). The upregulation of *Samui* in the November field samples may indicate that the process of diapause termination has started in these samples.

### Thermal history is a confounding factor in diapause transcriptome investigations

The present study considered only two factors: (1) the timing of entering diapause and (2) the overwintering conditions. We found a wide variation in gene expression patterns between the four treatments at each of the four time points examined. The simplest hypothesis for this observation is that the treatments desynchronized the bees' diapause development, resulting in the observation of different expression patterns for each group. As such, this hypothesis presupposes that diapause development is a fixed genetic pathway following a single order of gene expression patterns. A counter argument to desynchronization is that chilling of diapausing *M. rotundata* prepupae typically synchronizes adult emergence. Previous work has shown that chilling prepupae for 7 months (the same duration as the current May samples) decreased the emergence interval to 12 days from approximately 11 months when no chilling occurred (Johansen and Eves, 1973). Therefore, because all the bees experienced several months of chilling in this study, it can be reasonably argued that if *M. rotundata* diapause was driven by a single fixed pathway, the bee physiology should have been more similar as diapause progressed compared with the dissimilarity we found. Given these data, we propose an alternate hypothesis: that there are multiple genetic profiles through which diapause development may progress even though phenotypes may appear similar.



A key aspect of insect diapause is its anticipatory nature. Using various environmental cues, insects enter diapause prior to the start of unfavorable conditions. But many aspects of the environmental challenges diapausing insects face are totally unpredictable, including the severity, duration and frequency of individual stressors, as well as the temporal relationship between different environmental stressors. To survive the unpredictable complexity of environmental stressors, diapausing insects likely do not use a one-size-fits-all approach to overwintering. This plasticity has been demonstrated at the molecular level in a number of insect species. For example, in some diapausing insects, expression patterns of the heat shock proteins following a cold shock were determined by the insects' thermal histories [gypsy moth, *Lymantria dispar* (Lepidoptera: Erebidiae), Yocum et al., 1991; Colorado potato beetle, *Leptinotarsa decemlineata*, Yocum, 2001]. Similarly, in the present study, we found differences in expression patterns between the four treatment groups at each time point surveyed. Together, these results suggest that there is a complex interaction between an insect's environmental history and current thermal conditions that regulate diapause development and which may vary between individuals even within the same geographic location.

The majority of our understanding of insects' molecular mechanisms for overwintering and diapause development has come from laboratory studies. Because of the differences in environmental conditions between the laboratory and field, laboratory studies may not yield a complete picture of the dynamic range of diapause physiology. Indeed, under laboratory conditions, gene expression patterns of non-stress-associated genes in Colorado potato beetle entering diapause appear to be consistent between individual beetles (Yocum et al., 2009a,b), yet in the field, beetles collected from the same area exhibited eight different expression patterns (Yocum et al., 2011). This disparity between laboratory and field results supports the hypothesis that there is more than one order of gene expression leading to diapause initiation. This conclusion was further supported by work done by Lehmann et al. (2014) showing population differences in gene expression profiles as Colorado potato beetles entered diapause. One reason for these discrepancies is that studies of gene expression during diapause development under laboratory conditions are normally conducted under constant temperatures. However, diapausing insects under field conditions are exposed to fluctuating temperatures. Rearing post-diapause quiescent *M. rotundata* prepupae under a fluctuating thermal regime significantly impacted gene expression patterns of wide categories of genes (Torson et al., 2015). Together, these studies strongly indicate that environmental history has a major impact on gene expression patterns during diapause development. Additionally, these studies suggest that it may be difficult to compare results from different investigations owing to differences in the environmental histories of the individuals being compared.

### Identifying a diapause 'toolkit'

The hypothesis of a diapause 'genetic toolkit', which asserts that a conserved set of genes regulates diapause development across insect species, has been proposed by a number of research teams (Ragland et al., 2010; Poelchau et al., 2013a; Huestis and Lehmann, 2014; Amsalem et al., 2015; Yocum et al., 2015; Ragland and Keep, 2017). However, there is currently no agreement between the proposed lists of toolkit genes. The rationale for the existence of a toolkit is highly reasonable, given that there is only a limited number of regulatory pathways and checkpoints within those pathways for the regulation of diapause development. For example, it has been

well established that, depending on the developmental stage that enters diapause, either juvenile hormone or ecdysone is a part of the regulatory mechanism of diapause (Denlinger et al., 2005, 2011). The physiological effect of these hormones is exerted by regulating the rate of synthesis or degradation of the hormone, or by the differential regulation of the hormone receptor complex components. Therefore, at least with closely related insect species, the expectation for a common set of shared genes regulating diapause is based on solid physiological justification.

Nevertheless, a number of complexities compromise attempts to identify a diapause 'toolkit'. The first is that the insect's physiology is continuously changing as it proceeds through diapause development (Danks, 1987; Sawyer et al., 1993; Košťál, 2006; Košťál et al., 2017; Yocum et al., 2006). At the transcriptional level, each phase (initiation, maintenance and termination) of diapause in the fly *Chymomyza costata* (Diptera: Drosophilidae) can be further subdivided into subphases with specific gene expression patterns with very little overlap in the differentially regulated genes between these different subphases (Košťál et al., 2017). The dynamic nature of diapause development can also be seen in how rapidly gene expression patterns can change. Exposing a diapausing apple maggot, *Rhagoletis pomonella* (Diptera: Tephritidae), to a permissive temperature for the resumption of direct development induced 1394 differentially regulated genes over the period examined. Of the genes differentially regulated between the 0 and 24 h time points, 75% had reversed their direction of expression by the 48 h time point (Meyers et al., 2016). Therefore, to identify a robust and reliable diapause toolkit, a method must first be established to insure that physiologically equivalent stages of diapause are being compared.

Another significant challenge facing the identification of a diapause toolkit is that insects experiencing diapause development under field conditions may be physiologically different from insects diapausing under laboratory conditions. In addition to those studies discussed above, adult firebugs, *Pyrhocoris apterus* (Hemiptera: Pyrrhocoridae), whose diapause was terminated in the laboratory remained responsive to photoperiod, whereas those terminated under variable field conditions were nonresponsive to photoperiod (Hodek, 1968, 1983, in Košťál, 2006). Similarly, the median time for emergence from the overwintering substrate for postdiapausing Colorado potato beetle varied according to whether they were overwintered in the laboratory or the field (Tauber et al., 1988). Additionally, our study observed differences in gene expression between *M. rotundata* prepupae kept in the laboratory and the field. In light of all of these findings, a toolkit identified from experiments performed in controlled settings may not accurately reflect the physiological reality experienced by insects under natural environmental conditions. Development of an ecologically sound toolkit requires meeting the challenge of ensuring that physiologically equivalent stages are compared under relevant conditions. Therefore, developing a diapause toolkit of shared diapause-regulating genes may be technically challenging for all but a small group of closely related species with similar life histories.

The recently proposed hypothesis of omnigenics (Boyle et al., 2017) has major implications for the hypothesis of a genetic toolkit for complex traits such as diapause. The model of omnigenics proposes that complex traits (phenotypes) are directly impacted by a small number of genes that play an ascribed biological role in that trait as well as their direct regulators. These genes are referred to as core genes. But, owing to the interconnectivity of the regulatory pathways, any genes expressed are likely to affect the phenotype of a complex trait. A central tenet of omnigenics is that the core genes of

a complex trait are far outnumbered by these peripheral genes and that some complex traits may have no core genes at all. Classification of a particular gene as core or peripheral may be more a matter of degree than an either/or assignment. This ambiguity in classifying a gene as core or peripheral is yet another challenge in establishing a genetic toolkit for diapause.

The results presented here for *M. rotundata* suggest that the possible number of core genes (i.e. those genes shared between the four treatments in the between-month comparisons) may be as high as 406. A subset (130) of these potential core genes is shared between more than one comparison, increasing the likelihood that they are core genes (toolkit genes) (Table S3). Six of these genes are upregulated in four of the six possible between-month comparisons: November to March, November to May, January to March, and January to May. The physiological function of one of these six genes is currently unknown. The five remaining genes were compared with the list of differentially regulated genes isolated from *M. rotundata* diapausing and post-diapausing prepupae in Yocum et al. (2015). When directionality of expression is taken into consideration, only the gene *sex comb on midleg* (*Scm*) is shared between these two studies. The genes for both vitellogenin and phosphoglycolate phosphatase were downregulated whereas the gene for alkaline phosphate was not differentially regulated. One possible explanation for these discrepancies is that diapausing samples for Yocum et al. (2015) were collected in October rather than November as in this study.

The estimate of possible toolkit genes presented here is probably high, because the field samples used in this experiment may not be completely ecologically accurate for representing diapausing *M. rotundata*. One possible environmental factor not controlled for in this investigation was photoperiodism. The prepupae stored under laboratory conditions were kept in darkness and the field bees were exposed to the natural photoperiod. At the time of designing this experiment, *M. rotundata* was thought not to be sensitive to light (Tweedy and Stephen, 1970), a natural consequence of nesting in cavities. Under natural conditions, diapausing prepupae are cocooned and enclosed within a leaf-lined cell situated among a series of cells within a cavity. As such, light would be greatly restricted or totally blocked. We have recently discovered that developing *M. rotundata* are sensitive to light and that a limited amount of light can penetrate the leaf cell and cocoon (Bennett et al., 2018). If photoperiodism can account for a proportion of unexplained variance (75%) in the ANOVA model, this would strengthen our conclusion that the diapause transcriptome would vary between individuals based on their environmental history. In this hypothetical scenario, location and spatial orientation of the nest wherein some of the diapausing prepupae will be exposed to light and but not others will result in related gene expression differences. As more ecologically relevant experiments are developed, the number of possible core genes will likely decrease. Therefore, developing an accurate ecological understanding of diapause will require further understanding of the impact of peripheral genes upon a limited number of core genes.

## Conclusions

Environmental history influences gene expression patterns throughout the course of diapause development. The isolation of treatment-specific differentially regulated genes and the low level of overlap between these genes indicate that the prepupae were physiologically distinct at each time point examined. This study also underscores the need for future studies to establish that they are indeed comparing physiologically equivalent stages before

attempting to refine a genetic toolkit for diapause. Based on the results presented, we hypothesize that diapause is an omnigenic response and that the transcript expression profile of an individual insect is shaped more by its environmental history than by a single 'toolkit'. The conserved signaling pathways and molecular processes observed in various dormancies across taxa (Hand et al., 2016) suggest that our hypothesis may have wider application beyond Insecta.

## Acknowledgements

The authors thank Marnie Larson of the USDA-ARS, Fargo, ND, USA, for her technical assistance. We thank Marnie Larson and Lisa Yocum for editing the various drafts of the manuscript submitted for publication. We thank Ellen Klomps, USDA-ARS, Logan, UT, USA, and Utah State University undergraduates Martin Welker and Amy Spielmaker for their extremely valuable assistance in nest collections and dissections plus data entry and coordination. We appreciate the very helpful critiques of two anonymous reviewers.

## Competing interests

The authors declare no competing or financial interests.

## Author contributions

Conceptualization: G.D.Y., T.L.P.; Software: A.K.C.; Formal analysis: A.K.C., A.R.; Resources: T.L.P.; Data curation: A.K.C.; Writing - original draft: G.D.Y.; Writing - review & editing: G.D.Y., A.K.C., J.P.R., A.R., T.L.P., K.J.G., J.H.B.; Supervision: G.D.Y., J.P.R.; Project administration: G.D.Y., J.P.R.

## Funding

This research was funded by the U.S. Department of Agriculture, Agricultural Research Service Project Plans 3060-21220-027-00D and 3060-21220-030-00D to G.D.Y. and J.P.R. Data were processed using cyberinfrastructure supported by the National Science Foundation under award numbers DBI-0735191 and DBI-1265383 (CyVerse, <http://www.cyverse.org/>).

## Data availability

Raw sequence data are deposited in NCBI's SRA database under SRA study SRP116229, BioProject PRJNA400265, as 16 BioSamples, one for each treatment group and collection date, with accessions SAMN07562395–SAMN07562410. Transcript assembly, along with its associated descriptions and GO term and KEGG pathway identifiers, is available through the i5k Workspace@NAL (<https://i5k.nal.usda.gov/megachile-rotundata>) and can be viewed in the genome browser (Poelchau et al., 2015).

## Supplementary information

Supplementary information available online at <http://jeb.biologists.org/lookup/doi/10.1242/jeb.173443.supplemental>

## References

- Aguiss-Touré, A.-H., Wong, R. P. and Li, G. (2011). The ING family tumor suppressors: from structure to function. *Cell. Mol. Life Sci.* **68**, 45–54.
- Amsalem, E., Galbraith, D. A., Cnaani, J., Teal, P. E. A. and Grozinger, C. M. (2015). Conservation and modification of genetic and physiological toolkit underpinning diapause in bumble bee queens. *Mol. Ecol.* **24**, 5596–5615.
- Belvin, M. P., Jin, Y. and Anderson, K. V. (1995). Cactus protein degradation mediates *Drosophila* dorsal-ventral signaling. *Gene Dev.* **9**, 783–793.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Soc. Series B Methodol.* **57**, 289–300.
- Bennett, M. M., Rinehart, J. P., Yocum, G. D., Doetkott, C. and Greenlee, K. J. (2018). Cues for cavity nesters: investigating relevant zeitgebers for emerging leafcutting bees, *Megachile rotundata*. *J. Exp. Biol.* **221**, jeb175406.
- Boyle, E. A., Li, Y. I. and Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.
- Campos, E. I., Chin, M. Y., Kuo, W. H. and Li, G. (2004). Biological functions of the ING family tumor suppressors. *Cell. Mol. Life Sci.* **61**, 2597–2613.
- Cole, A. H. and Jones, S. N. (2008). The ING gene family in the regulation of cell growth and tumorigenesis. *J. Cell Physiol.* **218**, 45–57.
- Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J. and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.

- Cox, M. P., Peterson, D. A. and Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485.
- Danks, H. V. (1987). *Insect Dormancy: An Ecological Perspective*. Ottawa, ON: Biological Survey of Canada.
- Denlinger, D. L. (2002). Regulation of diapause. *Annu. Rev. Entomol.* **47**, 93–122.
- Denlinger, D. L., Yocum, G. D. and Rinehart, J. P. (2005). Hormonal control of diapause. In *Comprehensive Molecular Insect Science*, vol. 3 (ed. L. I. Gilbert, K. Iatrou and S. S. Gill), pp. 615–650. Oxford, UK: Elsevier.
- Denlinger, D. L., Yocum, G. D. and Rinehart, J. P. (2011). Hormonal control of diapause. In *Insect Endocrinology* (ed. L. Gilbert), pp. 430–463. The Netherlands: Academic Press.
- Dong, Y., Desneux, N., Lei, C. and Niu, C. (2014). Transcriptome characterization analysis of *Bactrocera minax* and new insight into its pupal diapause development with gene expression analysis. *Int. J. Biol. Sci.* **10**, 1051–1063.
- Doong, H., Vrailas, A. and Kohn, E. C. (2002). What's in the 'BAG'?—a functional domain analysis of the BAG-family proteins. *Cancer Lett.* **188**, 25–32.
- Doyon, Y., Cayrou, C., Ullah, M., Landry, A.-J., Côté, V., Selleck, W., Lane, W. S., Tan, S., Yang, X.-J. and Côté, J. (2006). ING tumor suppressor proteins are critical regulators of chromatin acetylation required for genome expression and perpetuation. *Mol. Cell* **21**, 51–60.
- Eisen, M. B., Spellman, P. T., Brown, P. O. and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868.
- 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.
- Enserink, J. M. and Kolodner, R. D. (2010). An overview of CDK1-controlled targets and processes. *Cell Div.* **5**, 11.
- Gong, Z.-J., Wu, Y.-Q., Miao, J., Duan, Y., Jiang, Y.-L. and Li, T. (2013). Global transcriptome analysis of orange wheat blossom midge, *Sitotiplosis mosellana* (Gehin) (Diptera: Cecidomyiidae) to identify candidate transcripts regulating diapause. *PLoS ONE* **8**, e71564.
- Hand, S. C., Denlinger, D. L., Podrabsky, J. E. and Roy, R. (2016). Mechanisms of animal diapause: recent developments from nematodes, crustaceans, insects, and fish. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **310**, R1193–R1211.
- Hao, Y.-J., Zhang, Y.-J., Si, F.-L., Fu, D.-Y., He, Z.-B. and Chen, B. (2016). Insight into the possible mechanism of the summer diapause of *Delia antiqua* (Diptera: Anthomyiidae) through digital gene expression analysis. *Insect Sci.* **23**, 438–451.
- Hobbs, G. A. and Richards, K. W. (1976). Selection for a univoltine strain of *Megachile (Eutricharaea) pacifica* (Hymenoptera: Megachilidae). *Can. Entomol.* **108**, 165–167.
- Hodek, I. (1968). Diapause in females of *Pyrrhocoris apterus* L. (Heteroptera). *Acta Entomol. Bohemoslov.* **65**, 422–435.
- Hodek, I. (1983). Role of environmental factors and endogenous mechanisms in the seasonality of reproduction in insects diapausing as adults. In *Diapause and Life Cycle Strategies in Insects* (ed. V. K. Brown and I. Hodek), pp. 9–33. The Hague, The Netherlands: Dr. W. Junk Publishers.
- Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.
- Huang, X., Poelchau, M. F. and Armbruster, P. A. (2015). Global transcriptional dynamics of diapause induction in non-blood-fed and blood-fed *Aedes albopictus*. *PLoS Negl. Trop. Dis.* **9**, e0003724.
- Huestis, D. L. and Lehmann, T. (2014). Ecophysiology of *Anopheles gambiae* s. l.: persistence in the Sahel. *Infect. Genet. Evol.* **28**, 648–661.
- Johansen, C. A. and Eves, J. D. (1973). Effects of chilling, humidity and seasonal conditions on emergence of the alfalfa leafcutting bee. *Environ. Entomol.* **2**, 23–26.
- Joyce, D., Albanese, C., Steer, J., Fu, M., Bouzahzah, B. and Pestell, R. G. (2001). NF- $\kappa$ B and cell-cycle regulation: the cyclin connection. *Cytokine Growth Factor Rev.* **12**, 73–90.
- Kabbage, M. and Dickman, M. B. (2008). The BAG proteins: a ubiquitous family of chaperone regulators. *Cell. Mol. Life Sci.* **65**, 1390–1402.
- Kanehisa, M. and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30.
- Kang, D. S., Cotton, M. A., Denlinger, D. L. and Sim, C. (2016). Comparative transcriptomics reveals key gene expression differences between diapausing and non-diapausing adults of *Culex pipiens*. *PLoS ONE* **11**, e0154892.
- Kapheim, K. M., Pan, H., Li, C., Salzberg, S. L., Puiu, D., Magoc, T., Robertson, H. M., Hudson, M. E., Venkat, A., Fischman, B. J. et al. (2015). Social evolution. Genomic signatures of evolutionary transitions from solitary to group living. *Science* **348**, 1139–1143.
- Kemp, W. P. and Bosch, J. (2001). Postcocooning temperatures and diapause in the alfalfa pollinator *Megachile rotundata* (Hymenoptera: Megachilidae). *Ann. Entomol. Soc. Am.* **94**, 244–250.
- Kim, D., Pertea, G., Trapnell, C., Pimental, H., Kelley, R. and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36.
- Košťál, V. (2006). Eco-physiological phases of insect diapause. *J. Insect Physiol.* **52**, 113–127.
- Košťál, V., Šimůnková, P., Kobelková, A. and Shimada, K. (2009). Cell cycle arrest a hallmark of insect diapause: changes in gene transcription during diapause induction in the drosophilid fly (*Chymomyza costata*). *Insect Biochem. Mol. Biol.* **39**, 875–883.
- Košťál, V., Štětina, T., Poupardin, R., Korbellová, J. and Bruce, A. W. (2017). Conceptual framework of the eco-physiological phases of insect diapause development justified by transcriptomic profiling. *Proc. Natl. Acad. Sci. USA* **114**, 8532–8537.
- Kronic, M. D. (1972). Voltinism in *Megachile rotundata* (Hymenoptera: Megachilidae) in southern Canada. *Can. Entomol.* **104**, 185–188.
- Lehmann, P., Piironen, S., Kankare, M., Lyytinen, A., Paljakka, M. and Lindström, L. (2014). Photoperiodic effect on diapause-associated gene expression trajectories in European *Leptinotarsa decemlineata* populations. *Insect Mol. Biol.* **23**, 566–578.
- Maisonneuve, C., Guilleret, I., Vick, P., Weber, T., Andre, P., Beyer, T., Blum, M. and Constam, D. B. (2009). Bicaudal C, a novel regulator of Dvl signaling abutting RNA-processing bodies, controls cilia orientation and leftward flow. *Development* **136**, 3019–3030.
- Malumbres, M. and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nat. Rev. Cancer* **9**, 153–166.
- Meyers, P. J., Powell, T. H. Q., Walden, K. K. O., Schieferecke, A. J., Feder, J. L., Hahn, D. H., Robertson, H. M., Berlocher, S. H. and Ragland, G. J. (2016). Divergence of the diapause transcriptome in apple maggot flies: winter regulation and post-winter transcriptional repression. *J. Exper. Biol.* **219**, 2613–2622.
- Moribe, Y., Niimi, T., Yamashita, O. and Yaginuma, T. (2001). *Samui*, a novel cold-inducible gene, encoding a protein with a BAG domain similar to silencer of death domain (SODD/BAG-4), isolated from *Bombyx* diapause eggs. *Eur. J. Biochem.* **268**, 3432–3442.
- Moribe, Y., Shiota, T., Kamba, M., Niimi, T., Yamashita, O. and Yaginuma, T. (2002). Differential expression of two cold-inducible genes, *Samui* and *sorbitol dehydrogenase*, in *Bombyx* diapause eggs exposed to low temperatures. *J. Insect Biotechnol. Sericol.* **71**, 167–171.
- Nabbi, A., McClurg, U. L., Thalappolly, S., Almami, A., Mobahat, M., Bismar, T. A., Binda, O. and Riabowol, K. T. (2017). ING3 promotes prostate cancer growth by activating the androgen receptor. *BMC Med.* **15**, 103.
- Nakagaki, M., Takei, R., Nagashima, E. and Yaginuma, T. (1991). Cell cycles in embryos of the silkworm, *Bombyx mori*; G2-arrest at diapause stage. *Dev. Biol.* **200**, 223–229.
- Niehrs, C. and Acebron, S. P. (2012). Mitotic and mitogenic Wnt signaling. *EMBO J.* **31**, 2705–2713.
- O'Neill, K. M., O'Neill, R. P., Kemp, W. P. and Delphia, C. M. (2011). Effect of temperature on post-wintering development and total lipid content of alfalfa leafcutting bees. *Environ. Entomol.* **40**, 917–930.
- Park, S., Blaser, S., Marchal, M. A., Houston, D. W. and Sheets, M. D. (2016). A gradient of maternal Bicaudal-C controls vertebrate embryogenesis via translational repression of mRNAs encoding cell fate regulators. *Development* **143**, 864–871.
- Peters, J.-M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nat. Rev. Mol. Cell Biol.* **7**, 644–655.
- Pitts-Singer, T. L. and Cane, J. H. (2011). The alfalfa leafcutting bee, *Megachile rotundata*: the world's most intensively managed solitary bee. *Annu. Rev. Entomol.* **56**, 221–237.
- Poelchau, M. F., Reynolds, J. A., Denlinger, D. L., Elisk, C. G. and Armbruster, P. A. (2011). A denovo transcriptome of the Asian tiger mosquito, *Aedes albopictus*, to identify candidate transcripts for diapause preparation. *BMC Genomics* **12**, 619.
- Poelchau, M. F., Reynolds, J. A., Elisk, C. G., Denlinger, D. L. and Armbruster, P. A. (2013a). Deep sequencing reveals complex mechanisms of diapause preparation in the invasive mosquito, *Aedes albopictus*. *Proc. R. Soc. B.* **280**, 20130143.
- Poelchau, M. F., Reynolds, J. A., Elisk, C. G., Denlinger, D. L. and Armbruster, P. A. (2013b). RNA-Seq reveals early distinctions and late convergence of gene expression between diapause and quiescence in the Asian tiger mosquito, *Aedes albopictus*. *J. Exper. Biol.* **216**, 4082–4090.
- Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C.-Y., Lin, J.-W. and Hackett, K. (2015). The i5k Workspace@ NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Res.* **43**, D714–D719.
- Qi, X., Zhang, L., Han, Y., Ren, X., Huang, J. and Chen, H. (2015). De novo transcriptome sequencing and analysis of *Coccinella septempunctata* L. in non-diapause, diapause and diapause-terminated states to identify diapause-associated genes. *BMC Genomics* **16**, 1086.
- Ragland, G. J. and Keep, E. (2017). Comparative transcriptomics support evolutionary convergence of diapause responses across Insecta. *Physiol. Entomol.* **42**, 246–256.
- 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. *Proc. Natl. Acad. Sci. USA* **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. Exper. Biol.* **214**, 3948–3960.



- Ray, P., Zhang, D.-H., Elias, J. A. and Ray, A. (1995). Cloning of a differentially expressed I $\kappa$ B-related protein. *J. Biol. Chem.* **270**, 10680-10685.
- Sawyer, A. J., Tauber, M. J., Tauber, C. A. and Ruberson, J. R. (1993). Gypsy moth (Lepidoptera: Lymantriidae) egg development; a simulation analysis of laboratory and field data. *Ecol. Modelling* **66**, 121-155.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212.
- Tammariello, S. P. and Denlinger, D. L. (1998). G0/G1 cell arrest in the brain of *Sarcophaga crassipalpis* during pupal diapause and the expression of the cell cycle regulator, proliferating cell nuclear antigen. *Insect Biochem. Mol. Biol.* **28**, 83-89.
- Tauber, M. J., Tauber, C. A. and Masaki, S. (1986). *Seasonal Adaptations in Insects*. Oxford, UK: Oxford University Press.
- Tauber, M. J., Tauber, C. A., Obrycki, J. J., Gollands, B. and Wright, R. J. (1988). Geographical variation in responses to photoperiod and temperature by *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae) during and after diapause. *Ann. Entomol. Soc. Am.* **81**, 764-773.
- Torson, A., Yocum, G. D., Rinehart, J. P., Kemp, W. P. and Bowsher, J. H. (2015). Transcriptional responses to fluctuating thermal regimes underpinning differences in survival in the solitary bee *Megachile rotundata*. *J. Exper. Biol.* **218**, 1060-1068.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelly, D. R., Pimental, H., Saizber, S. L., Rinn, J. L. and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562-578.
- Tweedy, D. G. and Stephen, W. P. (1970). Light refractive emergence rhythm in the leafcutter bee, *Megachile rotundata* (F.) (Hymenoptera: Apoidea). *Experientia* **26**, 377-379.
- Yocum, G. D. (2001). Differential expression of two *HSP70* transcripts in response to cold shock, thermoperiod, and adult diapause in the Colorado potato beetle. *J. Insect Physiol.* **47**, 1139-1145.
- Yocum, G. D., Joplin, K. H. and Denlinger, D. L. (1991). Expression of heat shock proteins in response to high and low temperature extremes in diapausing pharate larvae of the gypsy moth, *Lymantria dispar*. *Arch. Insect Biochem. Physiol.* **18**, 239-249.
- Yocum, G. D., Kemp, W. P., Bosch, J. and Knoblett, J. N. (2006). Thermal history influences diapause development in the solitary bee *Megachile rotundata*. *J. Insect Physiol.* **52**, 1113-1120.
- Yocum, G. D., Rinehart, J. P. and Larson, M. L. (2009a). Down-regulation of gene expression between the diapause initiation and maintenance phases of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *Eur. J. Entomol.* **106**, 471-476.
- Yocum, G. D., Rinehart, J. P., Chirumamilla-Chapara, A. and Larson, M. L. (2009b). Characterization of gene expression patterns during the initiation and maintenance phases of diapause in the Colorado potato beetle, *Leptinotarsa decemlineata*. *J. Insect Physiol.* **22**, 32-39.
- Yocum, G. D., Rinehart, J. P. and Larson, M. L. (2011). Monitoring diapause development in the Colorado potato beetle, *Leptinotarsa decemlineata*, under field conditions using molecular biomarkers. *J. Insect Physiol.* **57**, 645-652.
- Yocum, G. D., Rinehart, J. P., Horvath, D. P., Kemp, W. P., Bosch, J., Alroobi, R. and Salem, S. (2015). Key molecular processes of the diapause to post-diapause quiescence transition in the alfalfa leafcutting bee *Megachile rotundata* identified by comparative transcriptome analysis. *Physio. Entomol.* **40**, 103-112.
- Young, M. D., Wakefield, M. J., Smyth, G. K. and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* **11**, R14.