Original Article

TITLE

Gene regulatory mechanisms underlying the evolutionary loss of a polygenic trait

RUNNING TITLE

Transcriptional profile of loss of lipogenesis

AUTHORS

Mark Lammers[1], Ken Kraaijeveld[1], Janine Mariën[1], Jacintha Ellers[1]

[1] Vrije Universiteit Amsterdam, Department of Ecological Sciences, Section Animal Ecology, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

Corresponding author: Mark Lammers, marklammers6@gmail.com

Abstract

Trait loss is a pervasive phenomenon in evolution, yet the underlying molecular causes have been identified in only a handful of cases. Most of these involve loss-of-function mutations in one or more trait-specific genes. Parasitoid insects are fatty acid auxotrophs: they lost the ability to convert dietary sugars into fatty acids, a trait that is ubiquitous among non-parasitoid animals and is enabled by a highly conserved set of genes. Earlier research suggests that lack of lipogenesis is caused by changes in gene regulation, rather than gene decay. We compared transcriptome-wide responses to sugar-feeding in the non-lipogenic parasitoid species *Nasonia vitripennis* and the lipogenic *Drosophila melanogaster*. Both species adjusted their metabolism within four hours after feeding, but there was little overlap at the gene level between the responses of the two species. Even at the pathway-level, there were sharp differences between the expression profiles of the two species, especially in carbohydrate and lipid metabolic pathways. Several genes coding for key enzymes in acetyl-CoA metabolism, such as *malonyl-CoA decarboxylase* (MCD) and *HMG-CoA synthase* differed in expression between the two species. Their combined action likely blocks lipogenesis in the parasitoid species. Network analysis indicates most genes involved in these pathways to be highly connected, which suggest they have pleiotropic effects and could explain the absence of gene degradation. Our results indicate that modification of expression levels of only a few non-connected genes, such as MCD, is sufficient to enable complete loss of lipogenesis in *N. vitripennis*.

Key words

Fatty acid auxotrophy, trait loss, lipogenesis, parasitoid lifestyle, transcriptomics, pleiotropy, RNAseq

Introduction

In recent years, there has been a growing awareness of the adaptive role of trait loss in evolution (Fong et al. 1995; Lahti et al. 2009; Ellers et al. 2012; Royer et al. 2016). Trait loss has been shown to affect resource use efficiency (D'Souza et al. 2014), speciation (Stoks et al. 2003), host-parasitoid co-evolution (Pascoal et al. 2016) and the evolutionary potential of lineages (Bejder and Hall 2002; McLean et al. 2011).

The molecular changes underlying trait loss have been resolved in only a small number of cases. In a number of these, one or several key genes of the underlying pathway have degraded. For example, loss of vitamin C production in primates is caused by a single frameshift mutation in the gene *Gulo* (Chatterjee 1973; Ohta and Nishikimi 1999). Loss of four opsin genes, combined with reduced expression of nine important transcriptional factors, underlies eye vestigialization in cave fish (Yang et al. 2016). Accumulation of frameshift mutations in enamel-specific genes was found in a number of toothless and enamelless mammal lineages (Meredith et al. 2011; 2013). These studies reveal the molecular changes underlying trait loss for particular traits with relatively simple genetic bases. However, the genomic basis of trait loss in more complex, polygenic traits is still unresolved.

The expression of a complex trait is generally regulated by a network with many interconnected pathways, containing intermediate steps encoded by a large number of genes. Trait loss can be the result of a disruption in any of the intermediate steps, but gene network theory predicts that degradation of highly connected genes (so-called hub genes) is usually detrimental or even lethal (Jeong et al. 2001; Lee et al. 2008; Cho et al. 2012; Winterbach et al. 2013) due to pleiotropic functions of these genes. Other studies show that gene pleiotropy also correlates to evolutionary constraints on gene expression (Papakostas et al. 2014; Morandin et al. 2017; Yang and Wittkopp 2017). Therefore, these genes are unlikely targets for molecular changes underlying decay of complex traits. In contrast, genes with a more peripheral position in the gene network have lower connectivity and fewer pleiotropic functions (He and Zhang 2006; Zhu et al. 2007). When such genes decay or their expression is suppressed, relatively few

pleiotropic effects would result on other functions (reviewed in Winterbach et al. 2013). Regulatory changes might be a common mechanism underlying trait loss, but the role of gene expression and pleiotropic function in relation to trait loss is poorly understood (Gompel and Prud'homme 2009). *De novo* synthesis of fatty acids is a highly conserved complex process involving many deeply conserved and highly pleiotropic genes (Towle et al. 1997; Letunic et al. 2008; Arrese and Soulages 2010). It is an integral part of the life-history of most animals, enabling them to convert dietary carbohydrates to lipids, which allows storing energy for leaner times and for resource allocation to reproduction. Despite these essential functions of fatty acid synthesis, lack of lipogenesis has repeatedly evolved across the eukaryotic tree of life. For example, several lineages of fungi are fatty acid auxotrophs, caused by loss or degradation of the fatty acid synthase gene (Xu et al. 2007; Luginbuehl et al. 2017). Also, multiple insect lineages lack lipogenesis as shown by labelling studies (Giron and Casas 2003; Visser et al 2012; Visser et al 2017) or a lack of increase in adult fat reserves, despite feeding on sugar *ad libitum* (Ellers 1996; Visser and Ellers 2008). In insects, this recurrent loss of lipogenesis is phylogenetically linked to the parasitoid lifestyle: parasitoid clades of flies, beetles and wasps have lost lipogenesis independently (Visser et al. 2010). Phylogenetic analysis also shows that the ability to synthesize fatty acids has re-evolved independently in a number of parasitoid wasp species (Visser et al. 2010). The repeated regain of lipogenesis, as well as the central role of the lipogenic pathway in the carbon metabolism, suggests that the loss of lipogenesis is due to modification of gene expression rather than genetic changes in protein coding regions of the genome. In this study, we aim to unravel the changes in gene expression underlying the loss of lipogenesis in the parasitoid wasp *Nasonia vitripennis*. In *N. vitripennis*, there is no apparent gene loss or pseudogenization in pathways related to carbohydrate and lipid: a screening of the genome did not reveal gene losses in pathways relevant to lipogenesis (Werren et al. 2010). Visser et al. (2012) showed that in *N. vitripennis* sugar feeding does not cause a transcriptional response in several key genes in fatty acid biosynthesis, whereas these are upregulated under the same conditions in *Drosophila melanogaster*. These findings

suggest that lipogenesis is decoupled from other metabolic processes, sharply contrasting with regulation of these processes in other organisms (Towle et al. 1997; Zinke et al. 2002; Sassu et al. 2012).

We characterized transcriptomic changes in response to sugar-feeding in the non-lipogenic parasitoid *N. vitripennis*. No lipogenic strains of *N. vitripennis* are known, therefore we compared these changes to the response of a well-characterized representative lipogenic species, *D. melanogaster*. Both species have annotated genomes, well-studied life-histories and readily feed on sugar solutions. Although these species are in different insect orders, the metabolic pathways underlying lipid synthesis are highly conserved. We first compared the transcriptomic response to sugar-feeding for both species separately. Next, we compared these responses between the two species. Our results show that both species respond to sugar-feeding by adjusting transcription of key genes involved in multiple metabolic pathways, but network-based analyses indicate that both organisms evolved contrasting strategies in metabolizing dietary sugar.

Material and Methods

Study species

We used *Drosophila melanogaster* (Diptera: Drosophilidae) Bloomington stock 2057 (RRID:BDSC_2057, Adams 2000) and *Nasonia vitripennis* (Hymenoptera: Pteromalidae) strain AsymCX (Werren et al. 2010) in our experiments since the reference genomes originated from these strains. All strains were kept at 25°C, 75% relative humidity and 16:8h L:D cycle prior to and throughout the experiment.

Experimental setup

All newly emerged insects were kept without access to food for 24-36h (only water). This pre-treatment period ensured that all were hungry and eager to feed. Next, mated females were randomly assigned to either of two treatments: starved (St) and sucrose-fed (Sc). Starved insects were kept without food for another 4h; sucrose-fed insects were given *ad libitum* access to a sucrose solution (20% w/v). All insects

of the latter treatment were observed to feed within a few minutes. After exactly 4h all insects where killed by freezing in liquid nitrogen.

Tissue collection and RNA extraction

Each experimental condition consisted of 10 individual females and the experiment was repeated three times, yielding three independent biological replicates. Frozen females were dissected on a clean liquid nitrogen-cooled steel block: only abdomens were retained for total RNA extraction. For *N. vitripennis*, eight females were pooled per RNA extraction. For *D. melanogaster*, each sample of RNA was extracted from two pools of four females which were subsequently combined. The remaining two individuals per replicate were stored frozen as backup. RNA was extracted using the Promega SV Total RNA Isolation System kit (Promega Corporation, USA) following manufacturer's instructions except that RNA extracts were eluted in 30µL water. RNA concentrations were measured on a NanoDrop 2000 (Thermo Scientific) and the RNA Integrity Number was measured on a BioAnalyzer 2100 (Agilent). All samples had sufficient quantities of the required quality of RNA (supplementary table S1).

Illumina sequencing

All samples were sent to the Beijing Genomics Institute where library preparation and sequencing were performed. This strand-specific TruSeq library preparation included poly-A RNA purification, mRNA fragmentation, cDNA synthesis from size-selected fragments using random primers and adapter ligation for sample identification. Resulting short-insert libraries were pooled and sequenced (90bp paired-end) on one lane of Illumina HiSeq2000. This yielded 13-20M reads per sample (detailed in supplementary table S1). Raw sequence data have been deposited in the NCBI Short Read Archive under the study accession number SRP127311.

Transcriptomes and reference genomes

Sequence reads were checked for quality using FastQC version 0.10.1 [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/]: only high quality clean reads were delivered. The reference genome annotations used for *N. vitripennis* and *D. melanogaster* were GCF000002325.3_Nvit_2.1 (downloaded from NCBI RefSeq 27 October 2014) and dmel_r6.11 (downloaded from FlyBase on 9 June 9 2016), respectively. Complete gene identifier conversion tables where created from the genome annotation files using a custom R script.

Indices to the reference genomes were built with Bowtie2-build version 2.2.4 [http://bowtie-bio.sf.net/bowtie2]. Reads were mapped to the respective reference genomes using TopHat2 version 2.0.13 (Kim et al. 2013) with the following settings: -r 0 -p 4 --library-type fr-firststrand. Statistics of mapping success are reported in supplementary table S1. Alignment files were converted to SAM-format using samtools version 0.1.19 (Li et al. 2009). GFF-files were converted to comply to HTSeq format requirements with a python script. Expression levels were counted per gene using HTSeq version 0.6.1 (Anders et al. 2015) with default settings. Only genes for which five or more libraries (for each species separately) had non-zero counts were retained because zero counts could cause spurious similarity between samples (Storey and Tibshirani 2003).

Differential gene expression analyses were performed in EdgeR version 3.10.2 (Robinson et al. 2010) as recommended by Guo et al. (2013). Gene expression was compared between the two treatments (starved vs. sucrose-fed) for each species separately. Normalization factors were calculated with the function calcNormFactors with default settings. A negative generalized log-linear model was fitted to the data and a Likelihood ratio test was used to obtain P-values (McCarthy et al. 2012). P-values were corrected for multiple testing using the R function p.adjust following the method of Benjamini-Hochberg (1995). A gene was considered significantly differentially expressed when the p-value associated with this comparison was below 0.05 after FDR correction.

Gene Ontology Enrichment analyses

For each species, the set of differentially expressed genes was checked for enriched GO-terms with the R-package 'topGO' version 2.20.0 (Alexa et al. 2006). We used the weighed P-values that TopGO calculates using an algorithm that weighs statistical significances of higher GO categories by the significance of its lower categories: Higher categories are only recovered as significant when more genes are found at that category than expected by chance. GO-indexes for the *D. melanogaster* gene annotation were downloaded from FlyBase (Attrill et al. 2016). For *N. vitripennis* we used the Blast2GO-based GO-index for the Nvit2.1 assembly, available via DOI:10.13140/RG.2.1.2194.1929. Only GO-terms in the category 'biological process' were considered.

Orthologous gene-based comparisons

Orthologs of *D. melanogaster* and *N. vitripennis* were obtained from OrthoDB7 (Waterhouse et al. 2013). This list of orthologous groups is based on previous versions of the *D. melanogaster* and *N. vitripennis* genome annotations and gene predictions that are not supported in the current annotations were omitted from the analysis. All single-copy orthologs were extracted from the database. Next, we calculated the Pearson's correlation of the expression levels of the non-differentially expressed orthologs of *N. vitripennis* and *D. melanogaster*.

Species-specific transcriptional responses

For each species, we manually checked the lists of differentially expressed genes for involvement in lipogenesis, regulation of lipogenesis, or carbohydrate metabolism pathways. Sets of differentially expressed genes between treatments were screened for regulatory components: (1) transcription factors, as listed in REGULATOR (Wang and Nishida 2015) and (2) non-coding RNAs, as annotated in the respective reference genomes.

Gene pleiotropy was estimated by the number of protein-protein interactions as collected in the database STRINGdb (Franceschini et al. 2013). The absolute change in expression level (logFC) in our treatments was tested for a correlation with the number of protein-protein interactions after log2 transformation.

Visualization of KEGG-pathways

Each gene for which we obtained sufficient expression data was queried against the KEGG Pathway database (Kanehisa and Goto 2000; Kanehisa et al. 2008) in order to match it to a specific enzymatic reaction number. This analysis was repeated for *D. melanogaster* and *N. vitripennis* independently. These reactions were coupled to the corresponding gene expression data in our transcriptomes and imported into iPath2 (Letunic et al. 2008, http://pathways.embl.de/iPath2.cgi) to visualize the metabolic maps of both species. Our functional interpretation of biochemical pathways is based on Berg et al. (2006).

Pathway-based comparison between transcriptomes of D. melanogaster and N. vitripennis

The number of DE genes per KEGG Pathway were summed for both species and divided by the total number of genes that each species had for that pathway. This fraction of induced genes per pathway is presented as a heatmap using the R function heatmap.2 from package gplots.

Results

Global gene expression patterns

We obtained 13.7-20.4 million high-quality 90bp paired-end reads for each of the 12 libraries (2 species, 2 treatments, 3 biological replicates). 17.7-24.2% of reads were unequivocally mapped to a single locus on the respective reference genomes and kept for subsequent analyses (supplementary table S1).

We first established that our treatment of four hours *ad libitum* access to sugar had a measurable effect on the insect's abdominal transcriptome. Heatmaps of these transcriptomes are shown in figure 1. We found significant differences in expression level for 195 genes of *D. melanogaster* and 188 genes of *N.*

vitripennis (table 1, supplementary tables S2 and S3). The number of enriched gene ontology terms (GO-terms) in the sets of DE genes are presented in table 1 and supplementary tables S4 and S5. 59 GO-terms were enriched in *D. melanogaster* and 19 in *N. vitripennis* (Fisher's exact tests, weighted p-values <0.05). These significantly enriched GO-terms contain a broad spectrum of processes altered upon sugar feeding, including terms related to amino acid metabolism, reproduction, and carbohydrate and fat metabolism.

To validate further comparison between the transcriptomic responses of these species, we compared the expression level of the 1822 non-differentially expressed single-copy orthologs of *D. melanogaster* against their expression level in *N. vitripennis* (figure 2). The overall gene expression patterns are positively correlated (Pearson's r=0.58, t=30.155, df=1820, p<0.0001).

Transcriptional response to sugar-feeding in D. melanogaster

D. melanogaster upregulated *fatty acid synthase 1* (FBgn0283427) and *lipid storage droplet 1* (FBgn0039114) upon ingestion of sucrose. Genes related to catabolism of amino acids like g*lutamate oxaloacetate transaminase 1* (FBgn0001124) were downregulated upon feeding as well.

1181 genes (11.0% of all genes for which we had expression data) were successfully matched with known reactions in KEGG Pathway. The resulting overview of the active and induced metabolic pathways is presented in supplementary figure S1. If we assume unaltered enzyme efficiencies and protein turnover rates overall, this figure provides a bird's eye view on the organism's abdominal metabolism. It shows that many processes other than our focal pathways were altered upon sugar-feeding, e.g. enzymes involved in nucleotide metabolism and amino acid metabolism.

Gene expression is regulated by a range of mechanisms, including non-coding RNAs and cis-regulatory components like transcription factors. Of the differently expressed genes, nine were annotated as non-coding RNAs (table 2). These were all down-regulated upon sugar-feeding. Only one of the differentially expressed genes was a transcription factor as listed in the transcription factor database REGULATOR: *sugarbabe*. It was strongly upregulated in our experiment.

Transcriptional response to sugar feeding in N. vitripennis

In *N. vitripennis* none of the three copies of the enzyme *fatty acid synthase* (LOC100121447, LOC100122099, LOC100122083) was upregulated in the sugar-fed treatment. Several other genes related to lipid metabolism were upregulated: *acetyl-CoA carboxylase* (LOC100123347), *glucose 6-phosphate dehydrogenase* (LOC100120232) and *ATP-citrate lyase* (LOC100119651) and a *citrate transporter* (LOC100118210). A number of genes in other pathways linked to carbohydrate metabolism were differentially regulated upon sugar-feeding: *HMG-CoA synthase 1* (LOC100116401) and *phosphoenol pyruvate carboxykinase* (LOC10015526).

925 genes (8.5% of all genes for which we had expression data) had known reactions in KEGG Pathway. The resulting overview of the active and induced metabolic pathways is presented in supplementary figure S2, and the main changes of interest are summarized in figure 3.

Of the differently expressed genes in *N. vitripennis*, six genes were annotated as ncRNAs. Four of these were upregulated, two were downregulated (table 2). Of the transcription factors listed in REGULATOR, four were upregulated and two were downregulated expressed in our experiment (table 2).

Comparison between N. vitripennis and D. melanogaster

While the majority of single-copy orthologs showed a strong correlation in expression level between *N. vitripennis* en *D. melanogaster* (see above), there were a number of outliers to this correlation: several genes were highly expressed in *N. vitripennis* at all times, but had a much lower expression in *D. melanogaster* (upper-left in figure 2). The genes with the most extreme interspecific difference in expression were *radish*, *Phosphodiesterase 1c*, *CG9117*, and *olf413*. *Radish* is a Rap-Like GTPase Activating Protein involved in memory dynamics (Folkers et al. 2006). *CG9117* is a metallo-beta-lactamase domain-containing protein without further annotation. *olf413* is a copper type II ascorbate-dependent monooxygenase. The opposite expression pattern (highly expressed in *D. melanogaste*r, but

low in *N. vitripennis*) was found for *Gasp* and *walrus*. *Gasp* is a chitin-binding protein associated with embryonic development. *walrus* is an electron transfer flavoprotein, probably capable of accepting electrons of several dehydrogenases. The full table of expression levels of non-plastic single-copy orthologs is available in supplementary table S6.

KEGG-pathways associated with the DE genes were visualised as a heatmap in figure 3 for both species. Full tables are available in supplementary table S7. In the carbohydrate metabolic pathways, *D. melanogaster* showed multiple DE genes in fructose (two genes), galactose (nine genes) and sucrose (eight genes) metabolism, while *N. vitripennis* had no genes differentially expressed in these pathways. By contrast, differentially expressed genes in *N. vitripennis* were involved in the tricarboxylic acid (TCA) cycle (four genes), propanoate and butyrate metabolism (two genes each) and pyruvate metabolism (three genes). There were also divergent responses in the amino acid metabolisms, lipid metabolic pathways and in the pathways of the metabolism of co-factors and vitamins.

Differential expression correlates to gene pleiotropy

Figure 4 shows the correlation between absolute fold change of gene expression levels and the number of protein-protein interactions (PPI) as a measure for the level of pleiotropy. There was a significant negative correlation between fold change and the number of PPI: Pearson's $r = -0.195$ for *N. vitripennis* ($t=-20.772$, $df=10915$, $p<0.0001$) and $r = -0.216$ for *D. melanogaster* ($t=-25.192$, $df=12975$, $p<0.0001$). Many genes related to fatty acid metabolism are positioned at the high end of the pleiotropy spectrum: the median number of PPI for genes from KEGG Pathway map 'Fatty acid metabolism' (map01212) is 16, while this number is only 11.5 for all metabolic pathways combined (map01100).

Discussion

Global gene expression patterns

The set of differentially expressed genes upon sugar feeding in both species were enriched for a broad spectrum of GO-terms, showing that both species adjust their gene regulation within four hours after feeding on a sugar solution. The strong positive correlation of the gene expression levels of conserved (single-copy) genes shows that these patterns are maintained over long evolutionary timeframes (Diptera and Hymenoptera diverged about 340 million years ago (Misof et al. 2014)) and validates our subsequent comparisons of conserved metabolic pathways between *D. melanogaster* and *N. vitripennis*.

Transcriptional response to sugar-feeding in D. melanogaster

D. melanogaster accelerates its fatty acid synthesis after sugar-feeding, as indicated by the upregulation of the key enzyme *fatty acid synthase 1*. This was expected, as *D. melanogaster* is known to start lipogenesis shortly after sugar-feeding (Zinke et al. 2002). *Lipid storage droplet 1*, involved in storage of lipids in the fat body, was upregulated concomitantly, as would be expected when lipid production is increased. The downregulation of genes related to catabolism of amino acids upon feeding indicate that the flies may have used amino acids as fuel in the starvation treatment.

Nine non-coding RNAs were down-regulated upon sugar-feeding. The regulatory role of ncRNAs is a dynamic field of research (Rastetter et al. 2015; Su et al. 2016; Zhao et al. 2016), without generalized predictions of the function of specific ncRNAs. We show here that these nine ncRNAs are either co-regulated with the metabolic genes of *D. melanogaster* or regulating them. Note that our library preparation methods excluded all miRNAs. The transcription factor *sugarbabe* was strongly upregulated in our experiment. In an earlier microarray study this transcription factor was shown to be upregulated in *D. melanogaster* larvae shortly after sugar-feeding (Zinke et al. 2002).

Transcriptional response to sugar feeding in N. vitripennis

Fatty acid synthesis was not activated in response to sugar feeding in *N. vitripennis* as indicated by the lack of a response of the three *fatty acid synthases* in the sugar-fed treatment. However, several other

genes that play a direct or indirect role in fatty acid synthesis were upregulated. The gene *acetyl-CoA carboxylase* adds a carboxyl group to acetyl-CoA yielding malonyl-CoA which is required for fatty acid synthesis. High levels of malonyl-CoA also inhibit the activity of *carnitine acyltransferase I*, preventing fatty acid transport to the mitochondrion, thereby limiting β-oxidation. *Glucose-6-phosphate dehydrogenase* produces $NADPH^+$, required for anabolic processes like fatty acid synthesis. Activity of *ATP-citrate lyase* is indicative of acetyl-CoA transport from the mitochondrion to the cytoplasm, generally implicated in fatty acid synthesis. The upregulated *citrate transporter* potentially facilitates this transport. Citrate in the cytoplasm stimulates the activity of *acetyl-CoA carboxylase*.

These changes in gene expression in response to sugar-feeding indicate that most of the response of the lipid metabolism is intact, which contrasts with the phenotypic lack of lipogenesis. This paradox could be explained by evolutionary constraints on these enzymes: many enzymes have more than one function (pleiotropy), and the regulatory changes required to decouple all these processes might require many mutations.

A number of other pathways linked to carbohydrate metabolism were differentially regulated upon sugar-feeding. For example, gluconeogenesis was decelerated as indicated by a downregulation of *phosphoenol pyruvate carboxykinase* transcripts. A reduction of gluconeogenesis is expected to be concomitant with a reduction in ketogenesis. However, the upregulation of *hydroxymethylglutaryl-CoA synthase 1* indicates that ketogenesis was accelerated. Catabolism of acetyl-CoA through increased ketogenesis might be linked to loss of lipogenesis in *N. vitripennis*. An earlier study using qPCR to assess gene transcriptional responses to sugar feeding in *N. vitripennis*, found largely congruent results for key genes involved in carbohydrate, fatty acid, and glycerolipid metabolism, including *acetyl-CoA carboxylase*, *ATP citrate lyase*, *glucose-6-phosphate dehydrogenase* and *phosphoenolpyruvate carboxykinase* (Visser et al. 2012).

Four non-coding RNAs were upregulated and two were downregulated in *N. vitripennis*. This suggests that these non-coding RNAs in *N. vitripennis* could be involved in regulating its diverging metabolic response to sugar-feeding. Four transcription factors were upregulated and two were downregulated in our

experiment. As with the non-coding RNAs, we lack information on their targets and mechanism. It is possible that one of these, or their combined effects, regulate the decoupling of lipogenesis from sugar metabolism.

Comparison between N. vitripennis and D. melanogaster

It is unclear how the outliers in the comparison of expression levels of single-copy orthologs relate to differences in lipogenic abilities between the studied species. Similarly, it is unknown what the differences in expression patterns of non-coding RNAs and transcription factors mean. Nonetheless, our data indicate that *D. melanogaster* accelerates fatty acid synthesis upon sugar-feeding, while some key components of this pathway in *N. vitripennis* lack a response. In the carbohydrate metabolic pathways, *D. melanogaster* showed many differentially expressed genes in fructose, galactose and sucrose metabolism, while *N. vitripennis* had no genes differentially expressed in these pathways. By contrast, differentially expressed genes in *N. vitripennis* were involved in the tricarboxylic acid (TCA) cycle, propanoate and butyrate metabolism, and pyruvate metabolism. Moreover, there are contrasting responses in the amino acid metabolisms, lipid metabolic pathways and in the pathways of the metabolism of co-factors and vitamins. Although we currently cannot exclude temporal differences in the patterns of gene expression, these results suggest that these species use a different method of metabolizing dietary sugar and have a divergent response to sugar-feeding overall.

Differential expression correlates to gene pleiotropy

The negative correlation between fold change and the number of PPI indicates that highly pleiotropic genes are constrained in their extent of up- or downregulation, which is particularly relevant when the insect's metabolism is under selection. Many genes related to fatty acid metabolism are positioned at the high end of the pleiotropy spectrum, which means that most genes are rather constrained in their change in gene expression. We expected that the regulatory reorganization of fatty acid and acetyl-CoA

metabolism of *N. vitripennis* is likely to be directed by genes of low pleiotropy. A good candidate with zero known PPI would be *malonyl-CoA decarboxylase* (LOC100120093). This gene encodes the enzyme that converts one of the substrates for *fatty acid synthase*, malonyl-CoA, to acetyl-CoA. It has a high constitutive expression in *N. vitripennis* in both feeding treatments (logCPM=7.415). The high expression of this gene could therefore potentially deplete available malonyl-CoA, which would impede fatty acid synthesis. *Malonyl-CoA decarboxylase* has been lost in *D. melanogaster*. Other potential candidates having low PPI that likely underlie loss of lipogenesis are enzymes catabolizing acetyl-CoA, or otherwise disposing of it, such as genes involved in ketogenesis and the TCA cycle. Indeed, we show *HMG-CoA synthase 1*, an intermediate step in ketogenesis, to be upregulated in *N. vitripennis* upon sugar-feeding.

Conclusion

We characterized the gene expression patterns of the non-lipogenic *N. vitripennis* upon sugar-feeding in order to find clues to the molecular mechanism underlying the evolutionary loss of lipogenesis in this species. Animals feeding on sugar generally upregulate their lipogenic pathways (Towle et al. 1997; Kersten 2001), as we observed for the lipogenic *D. melanogaster*. *N. vitripennis* seems to have evolved a regulatory mechanism that decouples sugar ingestion and lipogenesis. Rather than storing dietary sugar in the form of fat, it uses the high expression of *malonyl-CoA decarboxylase* to counteract the activity of *acetyl-CoA carboxylase* and subsequently direct the acetyl-CoA to ketogenesis. The carbohydrates from sugar-feeding are used in somatic maintenance and as a source of energy for physical activity as parasitoid wasps feeding on sugar live longer and loose less fat reserves (Jervis et al. 2008 and references therein). Catabolizing excess carbohydrates via ketone bodies probably helps to avoid adverse effects of a high glucose diet and simultaneously enable retention of fat stores carried over from the larval stage. Our results raise the question why these lipogenesis-genes are maintained and expressed at detectable levels in *N. vitripennis* despite apparent lack of function. One explanation could be that the genes involved in lipogenesis have undetected subtle forms of gene degradation that impair enzyme function.

Alternatively, genes involved in lipogenesis could be under purifying selection through their pleiotropic effects. The regulatory mechanism hypothesized here could be a way of blocking lipogenesis in adult wasps, while maintaining the genes' other pleiotropic functions. This could be tested in a knockdown experiment, for example by using RNAi. Future studies on the regulatory network of lipogenesis would give mechanistic insights in these evolutionary constraints.

Supporting information

Supplementary tables

S1. Sample quality measurements, number of recovered reads per sample and mapping success

S2. Differentially expressed genes of *D. melanogaster*

S3. Differentially expressed genes of *N. vitripennis*

S4. Enriched GO-terms of the differentially expressed genes of *D. melanogaster*

S5. Enriched GO-terms of the differentially expressed genes of *N. vitripennis*

S6. Non-plastic genes sorted by residual differences in expression level

S7. KEGG-pathways associated with the differentially expressed genes per species

Supplementary figures

S1. Overview of the active metabolic pathways in abdomens of *D. melanogaster*

S2. Overview of the active metabolic pathways in abdomens of *N. vitripennis*

Acknowledgements

Lea van de Graaf helped setting up the feeding treatments. The VUmc provided access to their BioAnalyzer and Riet Vooijs helped with measurements. Peter Neleman and Tjalf de Boer advised on bioinformatics analyses. Bart Pannebakker kindly provided the GO-index for *Nasonia vitripennis*. This work was supported by a grant from The Netherlands Organization for Scientific Research [NWO, VICI grant number 865.12.003].

References

Adams MD. 2000. The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195. doi:10.1126/science.287.5461.2185.

Alexa A, Rahnenfuhrer J, Lengauer T. 2006. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics 22:1600–1607. doi:10.1093/bioinformatics/btl140.

Anders S, Pyl PT, Huber W. 2015. HTSeq - A python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169. doi:10.1093/bioinformatics/btu638.

Arrese EL, Soulages JL. 2010. Insect fat body: energy, metabolism, and regulation. Annu. Rev. Entomol. 55:207–225. doi:10.1146/annurev-ento-112408-085356.

Attrill H, Falls K, Goodman JL, Millburn GH, Antonazzo G, Rey AJ, Marygold SJ. 2016. FlyBase: establishing a gene group resource for *Drosophila melanogaster*. Nucleic Acids Res. 44:D786–D792. doi:10.1093/nar/gkv1046.

Bejder L, Hall BK. 2002. Limbs in whales and limblessness in other vertebrates: mechanisms of evolutionary and developmental transformation and loss. Evol. Dev. 4:445–458. doi:10.1046/j.1525-142X.2002.02033.x.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57:289–300. doi:10.2307/2346101.

Berg J, Tymoczko J, Stryer L. 2006. Biochemistry. New York: WH Freeman and Company.

Chatterjee IB. 1973. Evolution and the biosynthesis of ascorbic acid. Science 182:1271–1272. doi:10.1126/science.182.4118.1271.

Cho D-Y, Kim Y-A, Przytycka TM. 2012. Chapter 5: Network biology approach to complex diseases. PLoS Comput. Biol. 8:e1002820. doi:10.1371/journal.pcbi.1002820.

D'Souza G, Waschina S, Pande S, Bohl K, Kaleta C, Kost C. 2014. Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution 68:2559–2570. doi:10.1111/evo.12468.

Ellers J. 1996. Fat and eggs: an alternative method to measure the trade-off between survival and reproduction in insect parasitoids. Netherlands J. Zool. 46:227–235. doi:10.1163/156854295X00186.

Ellers J, Toby Kiers E, Currie CR, McDonald BR, Visser B. 2012. Ecological interactions drive evolutionary loss of traits. Ecol. Lett. 15:1071–1082. doi:10.1111/j.1461-0248.2012.01830.x.

Folkers E, Waddell S, Quinn WG. 2006. The *Drosophila* radish gene encodes a protein required for anesthesia-resistant memory. Proc. Natl. Acad. Sci. 103:17496–17500. doi:10.1073/pnas.0608377103.

Fong DW, Kane TC, Culver DC. 1995. Vestigialization and loss of nonfunctional characters. Annu. Rev. Ecol. Syst. 26:249–268. doi:10.1146/annurev.es.26.110195.001341.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P,

von Mering C, et al. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41:D808–D815. doi:10.1093/nar/gks1094.
Giron D, Casas J. 2003. Lipogenesis in an adult parasitic wasp. J. Insect Physiol. 49:141–147. doi:10.1016/S0022-1910(02)00258-5.
Gompel N, Prud'homme B. 2009. The causes of repeated genetic evolution. Dev. Biol. 332:36–47. doi:10.1016/j.ydbio.2009.04.040.
Guo Y, Li C-I, Ye F, Shyr Y. 2013. Evaluation of read count based RNAseq analysis methods. BMC Genomics 14:S2. doi:10.1186/1471-2164-14-S8-S2.
He X, Zhang J. 2006. Toward a molecular understanding of pleiotropy. Genetics 173:1885–1891. doi:10.1534/genetics.106.060269.
Jeong H, Mason SP, Barabási A-L, Oltvai ZN. 2001. Lethality and centrality in protein networks. Nature 411:41–42. doi:10.1038/35075138.
Jervis MA, Ellers J, Harvey JA. 2008. Resource acquisition, allocation, and utilization in parasitoid reproductive strategies. Annu. Rev. Entomol. 53:361–385. doi:10.1146/annurev.ento.53.103106.093433.
Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, et al. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36:D480–D484. doi:10.1093/nar/gkm882.
Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28:27–30. doi:10.1093/nar/28.1.27.
Kersten S. 2001. Mechanisms of nutritional and hormonal regulation of lipogenesis. EMBO Rep. 2:282–286. doi:10.1093/embo-reports/kve071.
Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 14:R36. doi:10.1186/gb-2013-14-4-r36.
Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, Foster SA. 2009. Relaxed selection in the wild. Trends Ecol. Evol. 24:487–496. doi:10.1016/j.tree.2009.03.010.
Lee D-S, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabasi A-L. 2008. The implications of human metabolic network topology for disease comorbidity. Proc. Natl. Acad. Sci. 105:9880–9885. doi:10.1073/pnas.0802208105.
Letunic I, Yamada T, Kanehisa M, Bork P. 2008. iPath: interactive exploration of biochemical pathways and networks. Trends Biochem. Sci. 33:101–103. doi:10.1016/j.tibs.2008.01.001.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352.
Luginbuehl LH, Menard GN, Kurup S, Erp H Van, Radhakrishnan G V, Breakspear A, Oldroyd GED, Eastmond PJ. 2017. Fatty acids in arbuscular mycorrhizal fungi are synthesized by the host plant. 1178:1175–1178. doi:10.1126/science.aan0081
McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40:4288–4297. doi:10.1093/nar/gks042.
McLean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT, et al. 2011. Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature 471:216–219. doi:10.1038/nature09774.
Meredith RW, Gatesy J, Cheng J, Springer MS. 2011. Pseudogenization of the tooth gene enamelysin (MMP20) in the common ancestor of extant baleen whales. Proc. R. Soc. B Biol. Sci. 278:993–1002. doi:10.1098/rspb.2010.1280.
Meredith RW, Gatesy J, Springer MS. 2013. Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. BMC Evol. Biol. 13:20. doi:10.1186/1471-2148-13-20.
Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG, et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science 346:763–767.

doi:10.1126/science.1257570.
Morandin C, Mikheyev AS, Pedersen JS, Helanterä H. 2017. Evolutionary constraints shape caste-specific gene expression across 15 ant species. Evolution 71:1273–1284. doi:10.1111/evo.13220.
Ohta Y, Nishikimi M. 1999. Random nucleotide substitutions in primate nonfunctional gene for l-gulono-γ-lactone oxidase, the missing enzyme in l-ascorbic acid biosynthesis. Biochim. Biophys. Acta - Gen. Subj. 1472:408–411. doi:10.1016/S0304-4165(99)00123-3.
Papakostas S, Vøllestad LA, Bruneaux M, Aykanat T, Vanoverbeke J, Ning M, Primmer CR, Leder EH. 2014. Gene pleiotropy constrains gene expression changes in fish adapted to different thermal conditions. Nat. Commun. 5:4071. doi:10.1038/ncomms5071.
Pascoal S, Liu X, Ly T, Fang Y, Rockliffe N, Paterson S, Shirran SL, Botting CH, Bailey NW. 2016. Rapid evolution and gene expression: a rapidly evolving Mendelian trait that silences field crickets has widespread effects on mRNA and protein expression. J. Evol. Biol. 29:1234–1246. doi:10.1111/jeb.12865.
Rastetter RH, Smith CA, Wilhelm D. 2015. The role of non-coding RNA in male sex determination and differentiation. Reproduction 150:R93–R107. doi:10.1530/REP-15-0106.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. doi:10.1093/bioinformatics/btp616.
Royer AM, Kremer C, George K, Pérez SG, Schemske DW, Conner JK. 2016. Incomplete loss of a conserved trait: function, latitudinal cline, and genetic constraints. Evolution 70:2853–2864. doi:10.1111/evo.13096.
Sassu ED, McDermott JE, Keys BJ, Esmaeili M, Keene AC, Birnbaum MJ, DiAngelo JR. 2012. Mio/dChREBP coordinately increases fat mass by regulating lipid synthesis and feeding behavior in *Drosophila*. Biochem. Biophys. Res. Commun. 426:43–48. doi:10.1016/j.bbrc.2012.08.028.
Stoks R, McPeek MA, Mitchell JL. 2003. Evolution of prey behavior in response to changes in predation regime: damselflies in fish and dragonfly lakes. Evolution 57:574–585. doi:10.1554/0014-3820(2003)057[0574:EOPBIR]2.0.CO;2.
Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. 100:9440–9445. doi:10.1073/pnas.1530509100.
Su Y, Wu H, Pavlosky A, Zou L-L, Deng X, Zhang Z-X, Jevnikar AM. 2016. Regulatory non-coding RNA: new instruments in the orchestration of cell death. Cell Death Dis. 7:e2333. doi:10.1038/cddis.2016.210.
Towle HC, Kaytor EN, Shih H-M. 1997. Regulation of the expression of lipogenic enzyme genes by carbohydrate. Annu. Rev. Nutr. 17:405–433. doi:10.1146/annurev.nutr.17.1.405.
Visser B, Ellers J. 2008. Lack of lipogenesis in parasitoids: A review of physiological mechanisms and evolutionary implications. J. Insect Physiol. 54:1315–1322. doi:10.1016/j.jinsphys.2008.07.014.
Visser B, Le Lann C, den Blanken FJ, Harvey JA, van Alphen JJM, Ellers J. 2010. Loss of lipid synthesis as an evolutionary consequence of a parasitic lifestyle. Proc. Natl. Acad. Sci. 107:8677–8682. doi:10.1073/pnas.1001744107.
Visser B, Roelofs D, Hahn DA, Teal PEA, Mariën J, Ellers J. 2012. Transcriptional changes associated with lack of lipid synthesis in parasitoids. Genome Biol. Evol. 4:864–874. doi:10.1093/gbe/evs065.
Visser B, Willett DS, Harvey JA, Alborn HT. 2017. Concurrence in the ability for lipid synthesis between life stages in insects. R. Soc. Open Sci. 4:160815. doi:10.1098/rsos.160815.Wang K, Nishida H. 2015. REGULATOR: a database of metazoan transcription factors and maternal factors for developmental studies. BMC Bioinformatics 16:114. doi:10.1186/s12859-015-0552-x.
Waterhouse RM, Tegenfeldt F, Li J, Zdobnov EM, Kriventseva E V. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. Nucleic Acids Res. 41:D358–D365. doi:10.1093/nar/gks1116.
Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJP, et al. 2010. Functional and evolutionary insights from the genomes of

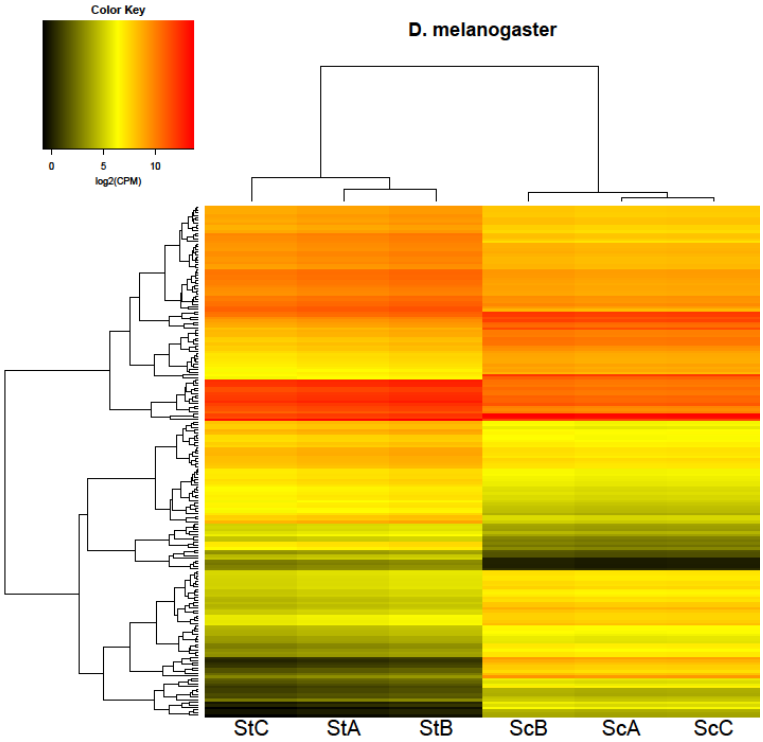
three parasitoid *Nasonia* species. Science 327:343–348. doi:10.1126/science.1178028.
Winterbach W, Mieghem P Van, Reinders M, Wang H, Ridder D de. 2013. Topology of molecular interaction networks. BMC Syst. Biol. 7:90. doi:10.1186/1752-0509-7-90.
Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, Kuramae EE, Kronstad JW, DeAngelis YM, Reeder NL, Johnstone KR, et al. 2007. Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. Proc. Natl. Acad. Sci. 104:18730–18735. doi:10.1073/pnas.0706756104.
Yang B, Wittkopp PJ. 2017. Structure of the transcriptional regulatory network correlates with regulatory divergence in *Drosophila*. Mol. Biol. Evol. 34:1352–1362. doi:10.1093/molbev/msx068.
Yang J, Chen X, Bai J, Fang D, Qiu Y, Jiang W, Yuan H, Bian C, Lu J, He S, et al. 2016. The *Sinocyclocheilus* cavefish genome provides insights into cave adaptation. BMC Biol. 14:1. doi:10.1186/s12915-015-0223-4.
Zhao J, He Q, Chen G, Wang L, Jin B. 2016. Regulation of non-coding RNAs in heat stress responses of plants. Front. Plant Sci. 7:1213. doi:10.3389/fpls.2016.01213.
Zhu X, Gerstein M, Snyder M. 2007. Getting connected: analysis and principles of biological networks. Genes Dev. 21:1010–1024. doi:10.1101/gad.1528707.
Zinke I, Schütz CS, Katzenberger JD, Bauer M, Pankratz MJ. 2002. Nutrient control of gene expression in *Drosophila*: microarray analysis of starvation and sugar-dependent response. EMBO J. 21:6162–6173. doi:10.1093/emboj/cdf600.

Data Accessibility

Raw sequence data have been deposited in the NCBI Short Read Archive under the study accession number SRP127311. Derived data (gene expression levels, GO-term enrichment, comparisons of orthologs, KEGG Pathway reaction-based results) are all included as supplementary material to this manuscript.

Author Contributions

JE conceived the project and performed the feeding experiments. JM performed most molecular wet-lab activities. ML ran the data analysis and interpreted the results. KK and JE advised on data analysis and interpretation. ML wrote the manuscript under supervision of JE and KK.



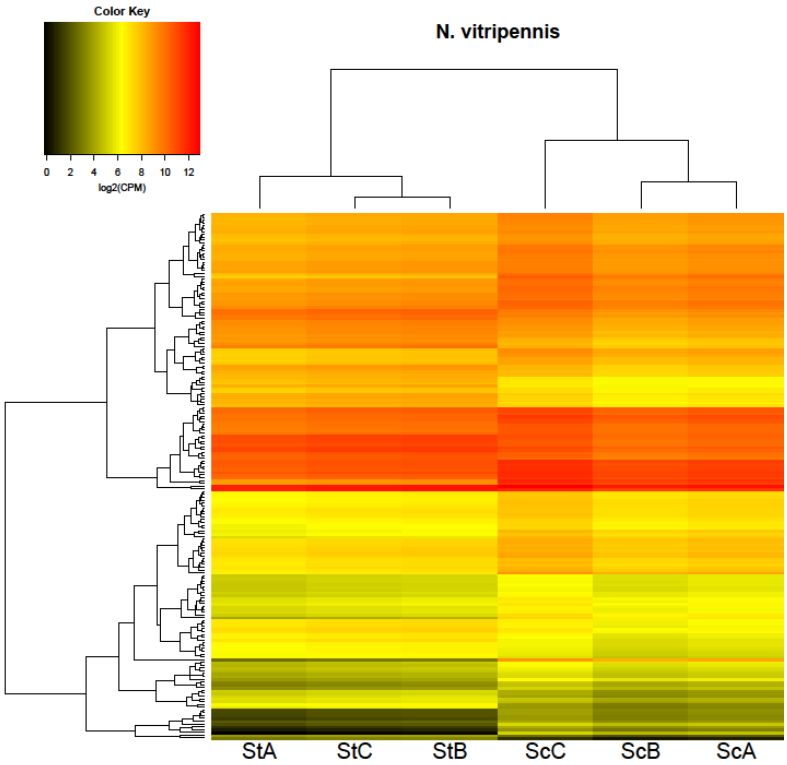


Figure 1. Heatmaps of the *D. melanogaster* and *N. vitripennis* transcriptomes. CPM = counts per million, St = starved, Sc = sucrose-fed, A/B/C = replicate.

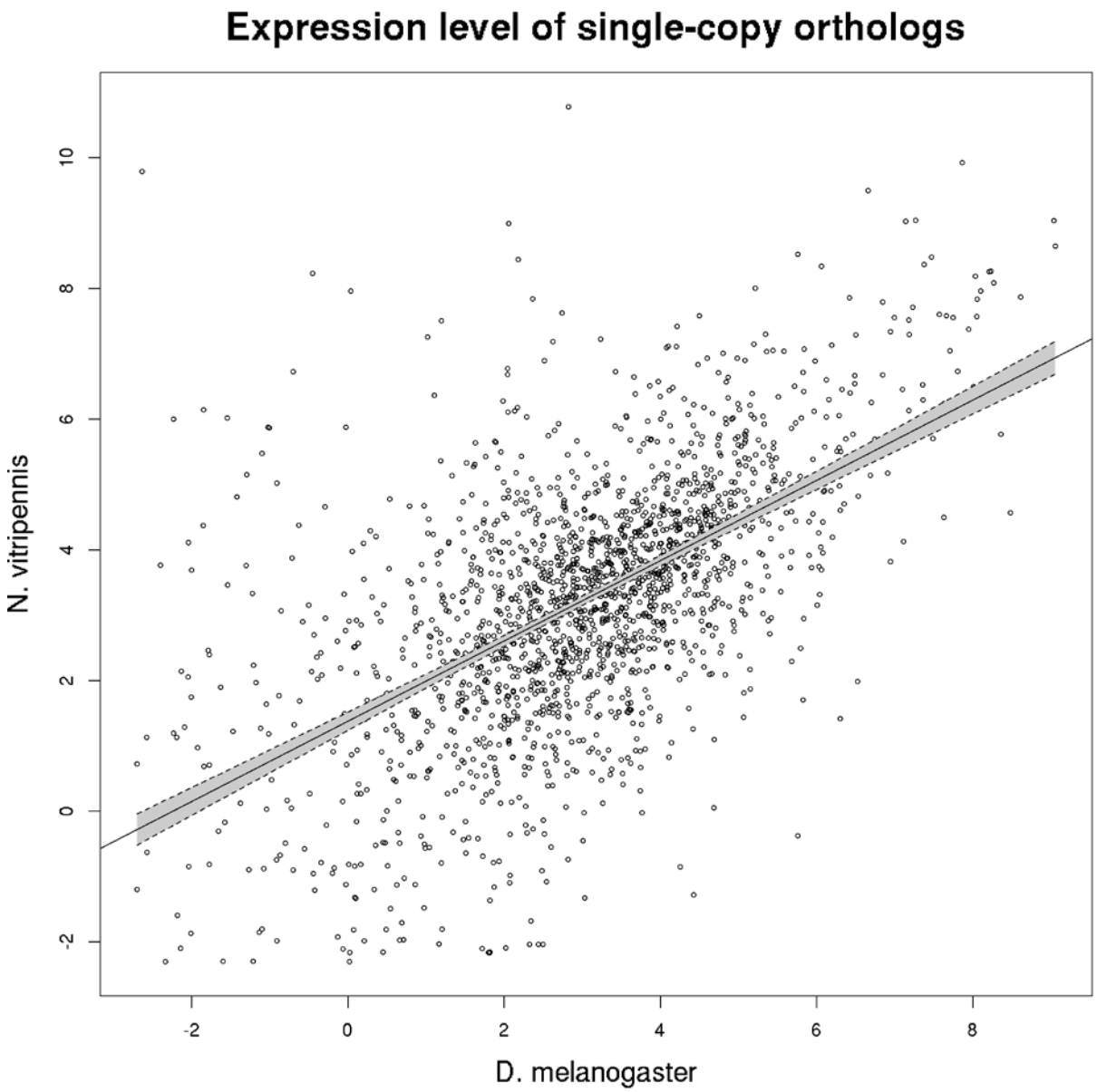


Figure 2. Interspecific correlation of expression level of non-differentially expressed single-copy orthologs between *D. melanogaster* and *N. vitripennis* when comparing sucrose feeding and brief starvation treatments.

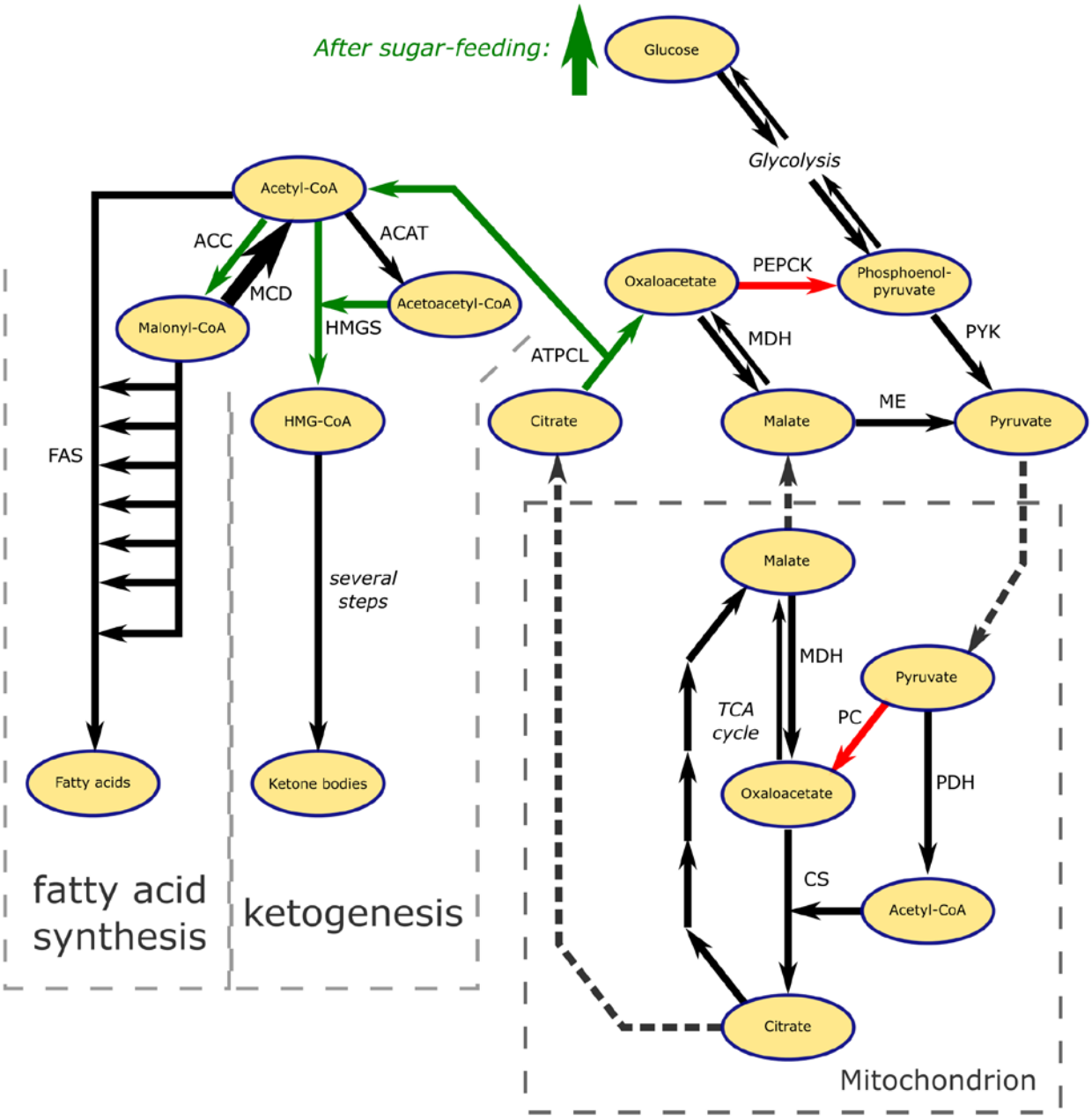


Figure 3. Differential gene expression in acetyl-CoA metabolism of *N. vitripennis* upon sugar feeding. Green and red arrows depict upregulated and downregulated genes, respectively. The thick arrow for MCD symbolize high constitutive expression. Dashed arrows represent transport of metabolites across the mitochondrial membrane.

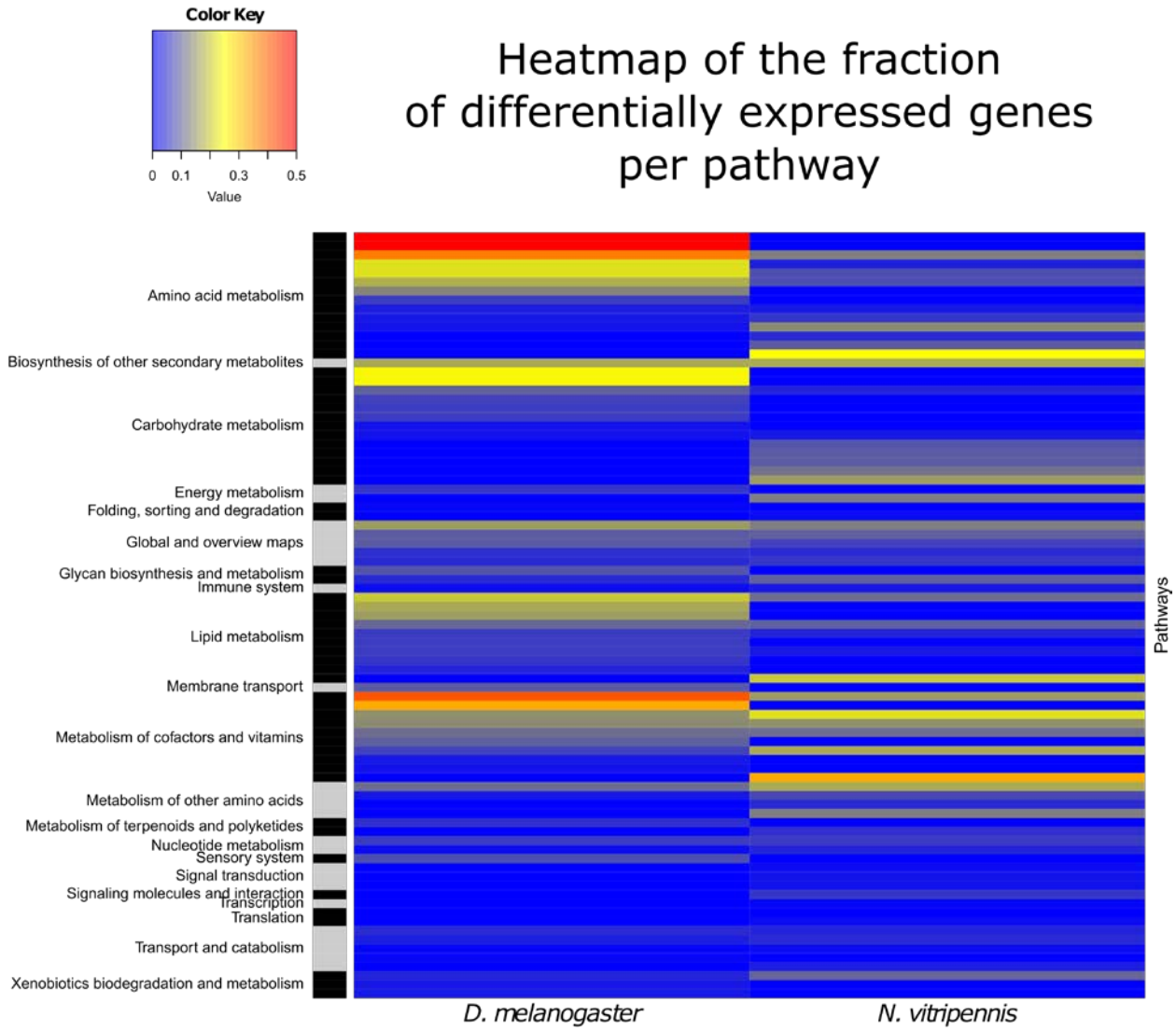
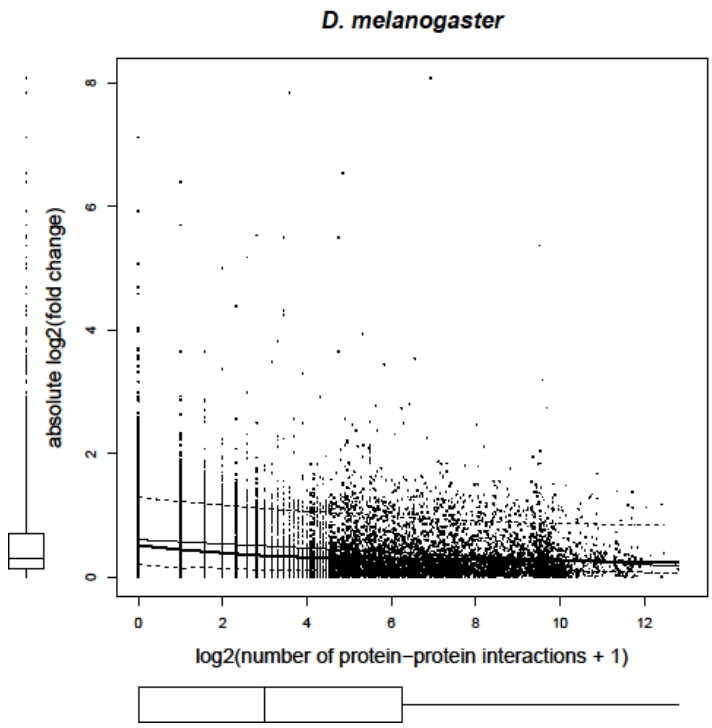


Figure 4. Heatmap of the fraction of genes per KEGG Pathway that was differentially expressed for *D. melanogaster* and *N. vitripennis* when comparing sucrose feeding and brief starvation treatments. The full table is provided in supplementary table S7.



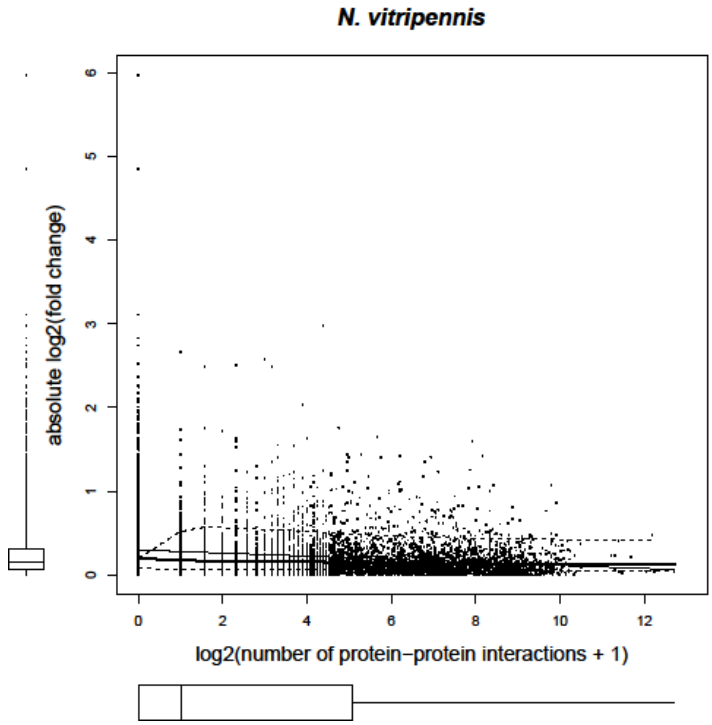


Figure 5. Scatterplots of the log2 fold change of gene expression upon sugar-feeding against the known log2-transformed number of protein-protein interactions for *D. melanogaster* and *N. vitripennis*. There is a significant negative correlation in both species.

606 **Tables**

607 Table 1. Summary of annotated genes for *D. melanogaster* and *N. vitripennis* and the number of genes up- and down-regulated upon sugar feeding
608 with the number of unique Gene Ontology (GO) terms associated to these differentially expressed genes.

Species	Number of genes in reference genome	Number of genes with expression data	Genes up-regulated	Genes down-regulated	GO terms
D. melanogaster	17,974	12,977 (72.2%)	84	111	59
N. vitripennis	14,321	10,914 (76.2%)	126	62	19

609

610 Table 2. Differentially expressed regulatory genes for *D. melanogaster* and *N. vitripennis* in response to sugar feeding compared to a brief
611 starvation treatment.

Species	Type	ID	logFC	logCPM	LR	PValue	Product
D. melanogaster	ncRNA†	FBgn0262904	-3.98328	1.510955	50.30487	1.42E-09	N/A
		FBgn0263448	-1.89139	1.986294	18.72951	0.002281	N/A
		FBgn0264939	-1.65076	5.328696	20.26412	0.001183	N/A
		FBgn0265150	-2.9169	-1.59583	15.07845	0.011061	N/A
		FBgn0266681	-2.36489	-0.00301	21.90987	0.000639	N/A
		FBgn0266686	-1.39943	3.940312	16.16845	0.007028	N/A
		FBgn0266702	-1.67772	3.741531	20.61357	0.001057	N/A
		FBgn0266705	-1.43091	3.972018	15.94068	0.00746	N/A
		FBgn0267617	-1.36943	4.795973	11.48211	0.047994	N/A
	TF‡	FBgn0033782	3.196841	2.536466	39.40898	2.35E-07	*sugarbabe*
N. vitripennis	ncRNA†	LOC103315329	1.22724	5.453971	60.4189	5.23E-12	N/A
		LOC103315513	0.79582	1.559863	13.07675	0.021931	N/A

	LOC103315648	0.884968	7.019906	29.15263	1.55E-05	N/A
	LOC103317395	0.882667	4.463351	17.37046	0.003292	N/A
	LOC103315928	-1.89299	3.81667	122.3051	5.40E-25	N/A
	LOC103315927	-1.53081	4.814787	60.44347	5.23E-12	N/A
TF‡	LOC100116547	-1.2319	1.170765	18.09523	0.002415	*muscle segmentation homeobox-like*
	LOC100124032	1.235618	0.241892	13.94669	0.015554	*GATA-binding factor A-like*
	LOC100115031	-0.45739	6.990395	19.5721	0.00129	*Krueppel-like factor 7*
	LOC100122625	0.510109	4.732156	12.6242	0.02631	*uncharacterized LOC100122625*
	LOC100678159	1.169649	3.616958	36.48003	5.91E-07	*Krueppel-like factor 10*
	LOC100115252	0.762056	3.469889	18.70013	0.001836	*forkhead box protein P1*

† ncRNA: non-coding RNA
‡ TF: transcription factor