Universiteit Leiden
The Netherlands

Targeting environmental and genetic aspects affecting life history traits

Baldal, E.A.

Citation
Baldal, E. A. (2006, November 23). *Targeting environmental and genetic aspects affecting life history traits*. Retrieved from https://hdl.handle.net/1887/4987

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4987

Note: To cite this publication please use the final published version (if applicable).

Chapter 5

Gene expression patterns of starvation resistant *D. melanogaster* under fed and starved conditions.

Gene expression patterns of starvation resistant *D. melanogaster* under fed and starved conditions.

E.A. Baldal[1, 3], D. Gunn[2, 3], P.M. Brakefield[1, 3] and B.J. Zwaan[1, 3]

[1] Institute of Biology, Leiden University, P.O. Box 9516, 2300 RA Leiden, The Netherlands
[2] Unilever R&D Colworth, United Kingdom
[3] On behalf of the "Lang Leven" consortium. The "Lang Leven" consortium consists of A. Ayrinhac, E.A. Baldal, M. Beekman, G.J. Blauw, D.I. Boomsma, P.M. Brakefield, B.W. Brandt, R. Bijlsma, D. van Heemst, B.T. Heijmans, J. van Houwelingen, D.L. Knook, I. Meulenbelt, P.H.E.M. de Meijer, S.P. Mooijaart, J. Pijpe, P.E. Slagboom, R.G.J. Westendorp, L.P.W.G.M. van de Zande, and B.J. Zwaan

Abstract

Earlier work on *Drosophila* lines selected for increased starvation resistance gave rise to the question of which genes underlie the response to selection. Therefore, in this pilot study whole genome gene expression is surveyed using micro arrays in new lines, one selected for increased starvation resistance and one control. These lines were predicted to show the largest differences in gene expression relative to a control line under selected, i.e. starved, conditions. They were assayed for their gene expression at three days from adult eclosion on normal rearing medium, and on starvation medium. Expression patterns were analysed by PCA and full factorial ANOVA. This revealed gene-by-environment interactions in the expression of several genes. We elaborate on the importance of the gene-by-environment interactions and examine genes that are known to be involved in specific life history related processes. To our knowledge, this is the first study that implies the involvement of candidate mechanisms in underpinning standing genetic variation for starvation resistance and longevity in natural populations. The outcome is discussed in the light of the literature.

Keywords

micro-arrays, *D. melanogaster*, starvation resistance, longevity, food condition, insulin, disposable soma

Introduction

This is a preliminary study and involves the first steps into micro-array analysis for this research group. The focus of this project has been to determine the technical issues and validity of this approach, as well as a test of how to interpret the biological data.

As the world has entered the genomic era so have the evolutionary and ecological sciences. That the linking of genomic data to evolutionary and especially ecological parameters with a clear phenotype will be informative is not under debate. However, there is some scepticism about the universal applicability of gene expression data (Feder and Walser 2005). This kind of scepticism is needed to remain objective because it is all too easy to be influenced by the potential promise of a technique. On the other hand, critically considering a technique can also be overdone. Thus far, genomics has not induced a revolution in the ecological field (but see van Straalen and Roelofs, 2006). Micro-arrays can be used to identify candidate genes that need to be more thoroughly investigated requiring a high degree of specialisation. To identify candidates, one needs to have a clear phenotype to assay. Also, the phenotype must have been assayed in an environment that is relevant to the trait, the one in which the phenotype usually occurs in and has been under selection in. Only in such circumstances the investigator can be sure that at least part of the altered gene expression is relevant to the traits under study. Here, all these requirements are met.

Instead of replacing “old-fashioned” scientific labour, the application of large scale, high throughput techniques has only increased its importance. The large amount of data and the issue of their significance offer new challenges for scientists in general.

Longevity and starvation resistance are correlated, but in part distinct, complex traits that are multifactorial in their determination. Next to complex environmental dependence, these traits also seem to be influenced by a large number of genes (Geiger-Thornsberry and Mackay 2004; Harbison et al. 2004; Harbison et al. 2005). Moreover, the environment and the genotypes interact (Leips and Mackay 2000; Vieira et al. 2000), resulting in different phenotypes, depending on environment and genotype. Large scale gene expression studies, such as micro-arrays, may help to strengthen and expand hypotheses about the complex of mechanisms underlying these traits.

Here, we present work on a limited set of micro-arrays. Of interest in this study are the genetic, environmental and gene-by-environment effects that affect gene expression and life span of *D. melanogaster*. Whole genome expression patterns of flies from a control and a starvation resistant line with distinct and known phenotypes were examined. The flies had been subject as adults to either normal rearing conditions, i.e. fed, or starved conditions.

We expected that under fed conditions the gene expression patterns would differ little between the control and starvation resistant lines, whereas differences would be much more apparent under starved conditions. Under fed conditions, gene expression patterns relevant to longevity may become apparent. Though the starvation resistant line (SR2) is significantly longer lived than its control line (C1), we

expect few genes to be reminiscent of this difference at an adult age of 3 days. In the case of starvation, three days is the point where the large difference between SR2 and C1 in starvation resistance is apparent. Also, it was thought that the impact of the environmental difference between starved and fed conditions would affect a large part of the genome as a response. A difference in response to the environmental change between the lines is called a genotype-by-environment interaction. Because of the selection regime we expect to find a considerable number of genes that display genotype-by-environment interactions. Further, we expect to find genes involved in insulin signalling, lipid, sugar and protein catabolism, stress perception and neural signalling to be differentially expressed and we will thus examine these classes of genes.

In this way, it was thought that genes of interest for further, detailed research could be identified. This was done using bioinformatic tools currently available and by personal observation and interpretation. Because this is a pilot study, the data were approached with a critical view for bioinformatic or personal interpretation biases. The genes of interest that were found are considered in the light of existing literature.

Materials and methods

Flies

Previously, we selected 4 lines of *D. melanogaster* for increased starvation resistance and maintained 2 control lines (Baldal et al. 2006). The starvation resistant line examined here, SR2, is highly starvation resistant and long-lived. The control line, C1, is not starvation resistant and does not display as long a life span as line SR2 under fed conditions.

For the experiments, eggs were reared at a density of 100 eggs per vial. From both lines, female virgins were collected within 8 hours post-eclosion and subject to fed conditions on standard medium or starved conditions on agar medium. Standard medium consists of 20 gr. agar, 9 gr. kalmus [kalmus consists of 10 parts (weight) acidum tartaricum, 4 parts ammonium sulphate, 1 part magnesium sulphate and 3 parts potassium phosphate], 10 ml. nipagin [100 grams of 4-methyl hydroxy benzoate per liter ethanol], 50 gr. saccharose and 35 gr. of granulated yeast per liter water. Agar medium consists of 20 gr. agar, 9 gr. kalmus, and 5 cc. nipagin per liter water. Adult flies were maintained at a density of 5 flies per vial. After 3 days the flies were collected and flash-frozen in liquid nitrogen, before storage at
-80ºC.

RNA extraction and micro-array handling

Total RNA was extracted using Macherey Nagel® Nucleospin II columns. Samples were checked for quantity and degradation using a Nanodrop® ND-1000 Spectrophotometer. Subsequently, RNA quality was analysed using Lab-on-a-Chip. Only samples were used that did not show signs of degradation (comparison of 18S and 28S rRNA peaks) and contained at least 2µg of total RNA. The samples were

then amplified and biotin-labelled using the Ambion kit and Ambion standard protocols by ServiceXS (www.servicexs.com). Samples of at least 12.5 µg were analysed using Affymetrix *Drosophila* 2.0 chips. The *Drosophila* 2.0 chip contains 18952 probe sets, analysing 18500 different transcripts. Hybridisation and readout were performed using standard protocols by the LGTC (www.lgtc.nl). In total 14 arrays were run, 3 for each line under feeding conditions and 4 for each line under starved conditions.

Bioinformatics and statistics

Quality control, normalization and ANOVA
The data in the .cel files were converted to excel sheets using the DCHIP programme (www.dchip.org). Tab-delimited files were imported into Genespring 7.2, which can link gene information to the Unigene database (http://www.ncbi.nlm.nih.gov). Measurements with an intensity of lower than 0.01 were changed to 0.01, and effectively removed from the analysis. Of the total 18952 probe sets we analysed, 14930 were labelled Marginal or Present by GeneSpring. These samples were normalized (MAS 5.0) per chip by the 50th percentile (median). They were then normalized (MAS 5.0) per gene to the median of all 14 arrays. Sample 2F1 (line 2, fed, first sample) showed signs of RNA degradation, when examined in the R-based Bioconductor (www.bioconductor.org). However, the chip was retained in the analysis to maintain a balanced design.

Samples were analysed in Genespring 7.2 by using Principal component Analysis (PCA) and full factorial ANOVA. A strict Benjamini and Hochberg False Discovery Rate (FDR) correction of 0.01 was also used to reduce the number of genes that were found but are actually not differentially expressed. Finally, a further 5247 probe sets were excluded due to lack of data for ANOVA analysis. The remaining 9683 probe sets were used for the PCA and ANOVA analysis. Differing FDRs were used in lower level analyses for arbitrary reasons and are listed in the results section. Thus, the number of genes found to be differentially expressed in the overall analysis does not match the number found in the lower level analyses.

FDR
The number of genes found to be differentially expressed is determined by the FDR. In our overall ANOVA analysis we use one FDR (0.01) and, therefore, the gene numbers between the different factors (genetic, environmental and genotype-by-environment interactions) can be compared. In our lower level analyses that focus on genotypic or environmental effects rather than an assay of the whole model, we always tried several FDRs and examined whether the trends followed expectations. Using a very low FDR will yield very few differentially expressed genes and is thus likely to miss many such genes (type I error). Conversely, a high value will display many that are actually not differentially expressed (i.e. falsely discovered, type II error). Because the FDR does not display a linear behaviour as a function of the proportion of total genes identified, a slight rise in FDR may yield a large number of extra genes. During this analysis, the point where an increase in FDR does not involve an equally large increase in the number of genes identified was used as a guide. This resulted in the difference in FDR among the lower level analyses.

Interpreting the data
Differentially expressed genes were grouped according to their Gene Ontology (GO) Biological Function level 6 category by the programme FatiGO (http://www.FatiGO.org). In Biological Function level 6 genes, are categorised in very specific processes. Approximately 10% of all *Drosophila* genes have GO annotation. Though there can be a bias in these gene lists towards processes of former scientific interest, the GO annotated samples were used to keep an overview of the processes. Only GO categories with over 10% of all the differentially expressed genes found or with a minimum of 3 genes of the annotated genes represented were used in the analysis.

From the literature, relevant categories and pathways were deduced. These include: insulin signalling, lipid, sugar and protein catabolism, stress perception and neural signalling.

Public availability

The full dataset will be made publicly available after future planned research in combination with additional data.

Results

Principal component analysis

The first principal component (PC1) reveals a large effect of nutritional status (25.7%, figure 1). PC2 indicates a genotype effect (22.3%, figure 2) with a clear separation between SR2 and C1. Combined, PC1 and PC2 show a genotype-by-environment effect (figure 2). There is a cluster of the lines SR2 and C1 on standard medium and a separate cluster for each line under starvation. Though the environmental component has a large effect, the clear distinction between the genotypes in one of the two conditions prompted a focus on the effects of genetics and genotype-by-environment interactions rather than on the environmental component. PC3 (13.5%) picked out the handling of the arrays in two batches (data not shown). The fourth PC (7.7%) explained the variation due to the sample 2F1 (data not shown). Both of these latter components indicated some technical variations in the data obtained here, and this was taken into account during the analysis.

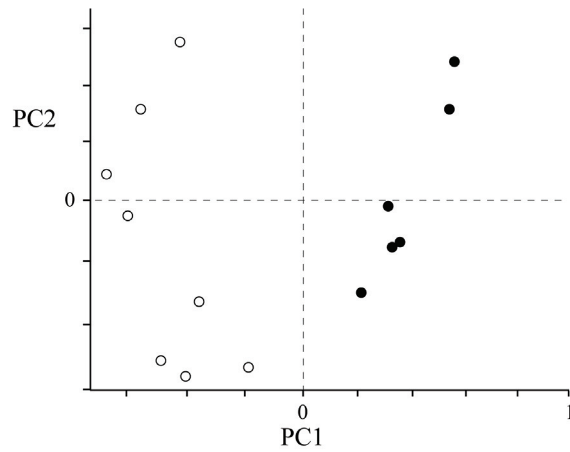

Figure 1. The first principal component (PC1, horizontal axis) on whole genome expression data, contrasting fed (filled symbols) and starved (open symbols) samples of two *D. melanogaster* lines, one control and one selected for increased starvation resistance.

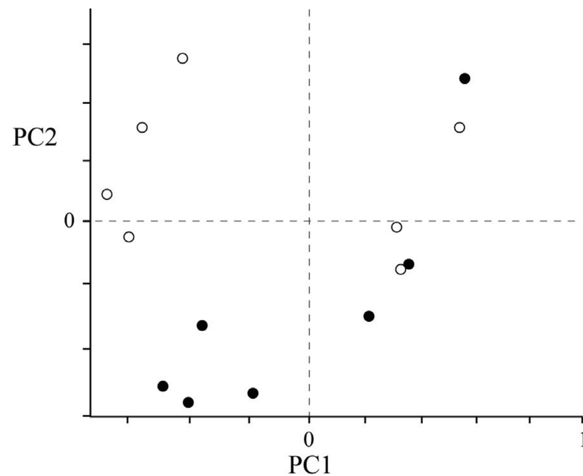

Figure 2. The second principal component (PC2, vertical axis) plotted against PC1 (horizontal axis) of whole genome expression data displaying the genotype-by-environment effect. On the right, in the fed samples as identified in the first PC (see figure 1), we observe that the starvation resistant line (SR2) and its control (C1) samples cluster together. On the left, in the starved samples, we see that the SR2 samples (open symbols) cluster separately from the C1 samples (filled symbols).

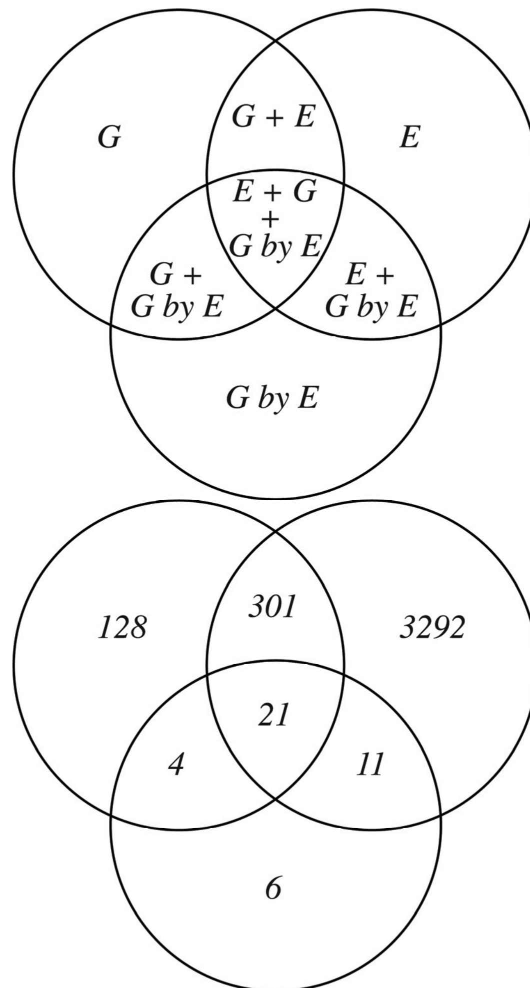

Figure 3. Outcome of the overall ANOVA. The upper Venn-diagram shows which effects the genes in a particular part of the circle are involved in. They can be involved in genetic (G), environmental (E) and gene-by-environment interactions (GbyE). Also there is overlap, where genes show, for example, genetic and environmental effects, but no gene-by-environment interaction. The lower Venn-diagram corresponds to the upper diagram. The numbers of genes that are involved in G, E effects and/or display GbyE interactions are listed in that part of the circle that represents that/those effects.

ANOVA

Figure 3 shows the numbers of genes that were detected to be differentially expressed in relation to the genotype and the environment. The full factorial ANOVA over all data showed that most genes were affected by the change in the

environment (3625 differentially expressed probe sets). This was followed by the difference between line SR2 and line C1 (454 differentially expressed probe sets). The genotype-by-environment interactions formed a minority (42 differentially expressed probe sets). A considerable number of probe sets were affected by more than one of the three factors from the full factorial design. These are shown in the overlapping regions between the circles. The low number of probe sets (138 of a total of 3763) unaffected in a consistent manner by the change in environment, but that are apparently only influenced by a genetic or genotype-by-environment cause, was striking. Of course in a genotype-by-environment interaction the environment plays an important role, but it is the genotype that responds to the environment that determines such a pattern. In this case no environment specific response could be identified from the gene expression patterns, hence the word consistent. The low number of genotype-by-environment (6 probe sets) was validated by its non-interacting counterpart, where G and E overlap (301 probe sets). In this last category, genes showed a difference in expression between the genotypes and between the environments. However, the response to the change in environment was similar in the lines.

Bioinformatics

Gene ontology analysis per food condition

The overall analysis of the genotypes (G effect) does not distinguish between the different environments, which are of considerable interest. Therefore, the G effects were subject to lower level analyses per feeding condition. Appendix 1 lists the affected biological processes from GO analysis per feeding condition. The difference between the lines when fed (G effect, FDR 0.5) comprised 62 probe sets of which 12 had a gene ontology classification. When starved, the effect of the genotype increased with 911 differentially expressed gene products (FDR 0.01), 79 of which had a GO category for biological processes at level 6.

Under fed conditions we found differential expression of GO categories involved in transcriptional regulation, nucleic acid related metabolism and neuronal signalling. Under starved conditions, GO categories were found to be involved mainly in transcription, development and catabolism.

Analysis of candidate genes

Full factorial ANOVA analysis

Genotype-by-environment interactions

In the list of genes that displayed genotype-by-environment interactions and had known functions, several functional categories were identified (Appendix 2). One set of four genes was down regulated in the control line in response to starvation, whereas there was no significant expression change in the starvation resistant line. One of these genes is the Juvenile hormone epoxide hydrolase 3 which is involved in

the catabolism of juvenile hormone. Juvenile hormone metabolism has been associated with longevity (Tatar et al. 2001).

A set of 10 genes showed down regulation in the control line and up regulation in SR2. The most striking example of this category is Turandot A, a gene involved in stress response. Gene-by-environment interactions also occur when the response of one of the lines is more extreme than that of the other line. In this category we have identified 10 genes, of which only 3 had a more extreme response in the starvation resistant line than the control line. Two of these latter genes are associated with sugar catabolism. Only one gene showed up regulation in the control line and no response in SR2, namely CG5933, an RNA methylation gene. We observe that of the genes that display genotype-by-environment interactions, 10 out of 25 have a response in the same direction only more or less extreme. Of the remaining 15 genes, 5 are not differentially expressed between environments in line SR2, which may indicate resource depletion or severe stress in line C1. The other 10 genes show a gene-by-environment effect with opposite reaction norms.

An analysis of the types of genotype-by-environment interactions that the genes displayed was performed (figure 4). In the 42 genes, 7 types of interactions were found, of a total of 10 possible (C1 up-SR2 up, C1 stronger response; C1 up-SR2 up. SR2 stronger response; C1 down- SR2 down, C1 stronger response; C1 down- SR2 down, SR2 stronger response; C1 up- SR2 down; C1 down- SR2 up; C1 stable- SR2 down; C1 stable- SR2 up; C1 up- SR2 stable; C1 down- SR2 stable). Of these, both options where C1 is stable and C1 down-SR2 down are not present. X^2 analysis showed a value of 17.4. The critical value for a two tailed test with 9 degrees of freedom was 19 at the 5% level. Thus there is no significant deviation from the expected distribution and no evidence for any forbidden categories.

Analyses per feeding condition

Fed flies

Closer examination of the genes that were differentially expressed on the basis of the GO molecular function revealed that sugar catabolising genes and chorion protein, a reproduction-related gene, were expressed at a lower level in the SR2 line. In contrast, alcohol dehydrogenase, constitutive proteins and translation initiation factors were expressed at a higher level. This suggests that SR2 flies, which have been found to have a metabolism similar to that of C1 (Baldal et al. 2006), have a different allocation pattern which involves lower expression of sugar metabolism genes and genes involved in reproduction. Instead, SR2 females could be increasing protein synthesis and detoxifying cells to help survival. The breakdown of alcohol is in line with the theory of McElwee et al. (2004) and Gems and McElwee (2005) that detoxification is a major factor involved in longevity determination. The indication of decreasing sugar metabolism matches the earlier finding that glycogen reserves increase considerably over lifetime in lines selected for increased starvation resistance, but are not elevated from eclosion (see Baldal et al. 2006, Appendix 3). High carbohydrate levels have, via desiccation resistance, often been associated with both longevity and starvation resistance (Graves et al. 1992; Force et al. 1995; Djawdan et al. 1996; Harshman et al. 1999b). Also, the decreased breakdown of

sugars may imply high hemolymph sugar concentrations, which matches the data on high glucose levels in flies that were long-lived because their insulin producing cells had been ablated (Broughton et al. 2005).

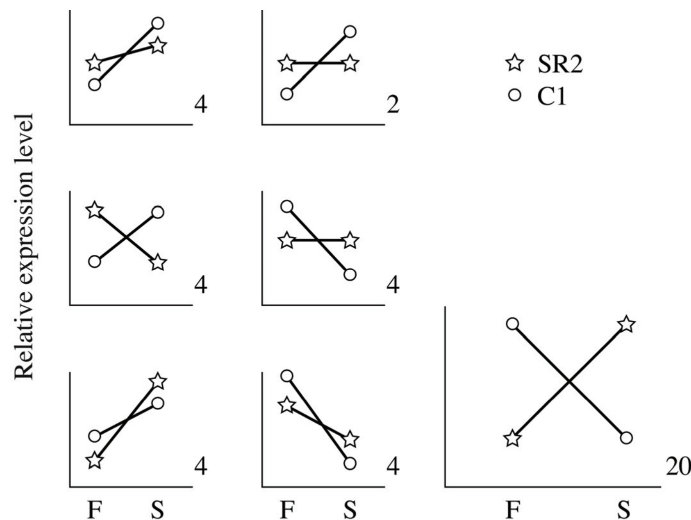

Figure 4. Gene-by-environment interactions. Graphs representing the relative gene expression level of the SR2 and C1 flies under fed (F) and starved (S) conditions. The numbers on the right side of the graph represent the number of genes out of a total of 42 we identified to be represented in this group. Depicted are only those types of genotype-by-environment interactions that were found in the data.

Starved flies: “The selection environment”

Since the gene lists are too large to give here, we have rather picked out those genes thought to be of interest. Sometimes genes occur in more than one group because they have more than one function or are involved in more than one process. We have listed the results of all genes we know to be involved in these processes.

The usual suspects

Genes that are often associated with starvation and ageing related processes are found in the gene lists. For example, *chico* is expressed at a lower level in the starvation resistant line which fits the observation that decreased *chico* expression in mutant flies shows increased life span (Clancy et al. 2001; Clancy et al. 2002; Tu et al. 2002). Juvenile hormone catabolising enzyme genes (2) are expressed at a higher level. This relates well to previous data (Tatar et al. 2001; Flatt and Kawecki 2004; but see Rauschenbach et al. 2004). The low level of heat shock protein expression in SR2 is somewhat surprising because high HSP expression levels have previously

been associated with longevity (Sorensen and Loeschcke 2001; Walker and Lithgow 2003), but may be indicative of lower levels of perceived stress in the starvation resistant line. Furthermore, the high expression level of triacyl glycerol lipase (3 genes), involved in triglyceride breakdown follows the earlier finding that the SR lines possess more lipids and have therefore more resources to burn off.

MAPKKK cascade

To survive, an organism has to recognize and deal with environmental stress as quickly as possible. In this experimental design under starvation conditions, C1 is likely to be more stressed than SR2 and, thus, should induce a higher stress response. The Jnk and MAPKKK have been found to be activated by environmental stress (Matsukawa et al. 2004; Zhuang et al. 2006). Nine genes were found of the Jnk and MAPKKK cascades that were expressed at a lower level in SR2 than in C1 (see Appendix 4a). The fact that the controls have relatively high expression levels is consistent with their higher experience of environmental stress during starvation than SR2.

Insulin signalling and sugar metabolism

Insulin signalling has often been associated with longevity (Braeckman et al. 2001; Clancy et al. 2001; Tatar et al. 2001; Dillin et al. 2002; Tu et al. 2002; Bluher et al. 2003; Holzenberger et al. 2003; McCulloch and Gems 2003; Walker and Lithgow 2003). Here we find that in SR2 flies some of the genes from the insulin signalling pathway are expressed at a lower level than in C1 (Appendix 4b).

Sugar catabolic enzymes are more highly expressed (10 genes) and the build up of glycogen is regulated at a lower level in the SR2 flies than in C1 (1 gene). We could not identify a pattern in sugar transporters (6 genes). Furthermore, glycolysis associated genes were expressed at a lower level in SR2 flies than in C1 flies (4 genes). This relates to data demonstrating that insulin signalling increases glycolysis (e.g. Beitner and Kalant 1971). Thus, there is a consistently low level of expression in the insulin pathway and its downstream targets such as glycolysis and gluconeogenesis in the SR2 flies.

Lipid metabolism

Lipid breakdown genes, such as lipases, generally have a higher expression level in the starvation selected line (9 genes, one has a lower expression level). Both cholesterol metabolism related genes found were expressed at a lower level in SR2 flies. Most genes involved in fatty acid biosynthesis (4 out of 6 genes) are also expressed at a higher level. This is consistent with higher availability of lipids of the starvation resistant flies and may be involved in fatty acid modification rather than de novo synthesis. As for the sugar metabolism related genes, the lipid transporters do not show a uniform response.

Protein metabolism

Proteolysis and peptidolysis associated genes (25 genes) were expressed at a higher level in the starvation resistant line than in the control line. Four genes with this proteolytic function were down regulated. Of these, *surfeit6* fits well with other gene expression changes because it is an ATP dependent proteolytic enzyme. Since glycolysis related genes show relatively low expression, so is the production of ATP and thus the ATP available for this enzyme to act. The remaining 10 genes, which have highly different functions, were all expressed at a lower level in line SR2. Four

of these were involved in protein modification and the other 6 were involved in diverse processes such as protein folding, transport and cell death.

Reproduction
Longevity and starvation resistance are known to be involved in a trade off with reproduction at a phenotypic (e.g. Chippindale et al. 1993) and genetic level (e.g. Zwaan et al. 1995b. We have already suggested that the insulin pathway genes were expressed at a lower level in the SR2 line compared to the control line. This pathway is thought to regulate longevity and reproduction (Dillin et al. 2002). Therefore, reproduction-associated genes can be expected to be down-regulated in starvation resistant lines relative to their control. Appendix 4c lists the 14 genes associated with the reproductive machinery, each of which was expressed at a lower level, supporting our hypothesis. Strikingly, of those genes we know to be involved in reproduction, none were up-regulated in line SR2. This consistency indicates that there is likely to be a strong effect on reproduction.

Neural tissue associated genes
A total of 26 genes associated with neural processes were found to be differentially expressed. Of these, 12 were known to be involved in development and neurogenesis, 4 of which were expressed at a higher level in SR2. The exact function of 2 genes remains unknown to us, not withstanding the fact that they are associated with neural tissue in GO annotation. A total of 12 genes were active in the transmission of neural signals. Of these, 3 were involved in neurotransmitter secretion (2 lower, 1 more expressed in SR2) and 9 were involved in nerve impulse (8 lower, 1 more expressed in SR2).Generally, we can state that neural signal transduction genes are expressed at a lower level.

Discussion

Technical issues

Many of the genes identified as differentially expressed have not been annotated yet or they do not have gene ontology categories. From this it follows that in this paper we are concentrating much more on previously investigated candidate mechanisms than on discovering new pathways and genes. This makes using current gene knowledge, bias prone. However, they supply us with valuable information about if, and to what extent candidate mechanisms are associated with standing genetic variation in natural populations for the traits they are supposed to underpin. Better annotation of the *Drosophila* genome is crucial to link the knowledge about this system to all the molecular processes responsible and not only to those that are already well-known (cf. Gems and McElwee 2005).

In the PCA analysis, 20% of all variation was explained by technical factors. On the other hand, in the ANOVA analysis, the FDR also generates errors. Micro-array data have thus not only noise because of inherent biological variation, but also because of technical and analysis variation. The question remains how worrisome this is. We have already acknowledged the necessity for caution and know that the micro-array data themselves are not enough to prove or disprove hypotheses. Therefore, we

should take the uncertainties into consideration and use them for the thing they have been designed for: revealing patterns in a large and complex dataset. In this way micro-arrays are used as a means to gain an indication of whether or not to pursue certain hypotheses.

The technical variation will not be as consistent as a biological signal. There may be a large number of genes involved in a pathway or a response mechanism. Single gene expression differences will typically be generated by technical variation. These will not yield as strong a basis for hypotheses as will pathway differences. Therefore, the technical variation need not pose an insurmountable problem upon careful consideration of the data.

General analysis

It is very important to acknowledge that micro-array experiments do not necessarily provide truth. Micro-array analysis usually does not comprise a large number of chips or high replication because of the costs. Therefore, one should be cautious in the interpretation of micro-array data and acknowledge the fact that hypotheses provided by micro-array analysis should be followed up using quantitative real time PCR. In addition, because we have 3 to 4 replicates for each treatment, the significant differences, especially in the ANOVA, are likely to be of interest. Here, we use micro-arrays within a hypothesis driven framework. The outcome of the gene expression study is not used as proof for a hypothesis, but rather to gain insight about whether testing certain aspects of the hypothesis is worthwhile. In doing so, lines were used with a distinct phenotype and known biology. We began this research with a clear hypothesis in mind; the gene expression differences of the selected line will show most differences when assayed in the environment of selection, these expression differences are likely to involve several well-known pathways, such as the insulin signalling pathway. This hypothesis drove this research, which was carried out under relevant circumstances, accordingly.

PCA analysis and ANOVA revealed that the large difference between the selected and control line is only present under starved conditions. This is not surprising since that is the condition for, and under which the starvation resistant line is explicitly selected. In our project we focus on the genes underlying starvation resistance and possibly longevity, we leave the environmentally differentially expressed genes for a more ecological genomics oriented path of research.

PCA 1 and 2 together reveal a genotype-by-environment effect; a distinct difference between the lines under starved conditions. These first two components show the power of micro-array analysis to identify processes and provide a good argument for continuing this line of research. The third and fourth components urge caution about the results. It is noteworthy that especially the third component may be reduced in strength by just normalising the expression data within batches.

Genotype-by-environment interactions

A priori we expected to find a large number of gene-by-environment interactions by comparing the transcriptome of lines C1 and SR2 under fed and starved circumstances. Line SR2 displays high starvation resistance and longevity, compared to the control (Baldal et al. in prep.; Baldal et al. 2006). By introducing a clear shift in the environment, feeding vs. starvation, we observe large differences in the gene expression patterns. The environmental change from feeding to starving induced 3625 gene products to change their expression pattern significantly. Selection on a quantitative trait, in this case starvation resistance, led to differential gene expression, but apparently in a smaller number of genes than in the case of the environmental change. Not many genes were found to show a genotype-by-environment interaction. Of the total number of 3763 differentially expressed gene products only about 1% showed a significant genotype-by-environment interaction. Genotype-by-environment interactions are established by non-parallel slopes of reaction norms; thus different genotypes show a different response to the environmental gradient (Stearns 1992).We examined one selected and one control line, SR2 and C1, respectively, for the performance of a single trait (life span) in two environments (fed and starved).

The ability to survive a period of extreme adversity such as starvation is an important feature in the life history of an organism. Selection pressure on such a trait can be very harsh in certain environments, as has been pointed out clearly for the human situation by Diamond (2003). Starvation resistance is called for by most, if not all, living organisms in adversity. The genes responsible for starvation resistance are thought to overlap with those determining resource allocation and somatic maintenance, hence the association with longevity. Life history traits such as longevity and starvation resistance are thought to rely on public mechanisms (Partridge and Gems 2002) that are shared throughout the animal kingdom and maybe more. This indicates that the mechanisms underlying these traits are evolutionarily conserved. One might argue that if such basal genes affect many important processes they will have to be more stringently regulated and are not likely to show radical changes in expression pattern. This provides a hypothesis for why in this specific example few gene-by-environment interactions are found. On the other hand, when the environment changes radically, a lot of processes may need to be changed, which is exactly the thing these key-pathways do. In the response to the environment we observe this. In a selection process, changing the expression of only a few genes, which will come on top of the normal large shift in gene expression as a result of the environment, may suffice. If the gene-by-environment interactions are reminiscent of the response to selection, then they should turn up in the loading of the factors involved in the PCA. Testing this hypothesis is one the foci of future research.

The strong genotype-by-environment effect we found in the principal components can be explained by the fact that indeed those genes that have been affected most by selection display the largest differences, add most to the variation and thus end up being reflected in the principal component. It should be noted again that PCA is a data reduction tool and should be used to gain insight from a large and difficult dataset. Close examination of the factors that attribute most to variation in combination with ANOVA analysis should reveal the extent to which those genes

contribute most to the variation are also those displaying genotype-by-environment interactions. Future work will focus on this.

Genes

The differentially expressed genes under starved conditions revealed that flies of the control line experience more stress than those of the starvation resistant line. Line SR2 shows a clearly changed pattern of gene expression and seems to burn off resources at a higher rate, down regulates insulin signalling and, consequently, glycolysis. Though burning off resources and down regulating glycolysis seems contradictory this is not necessarily the case. Glucose is only one of the resources available and its down-regulated catabolism may be a necessary consequence, which is compensated by the breakdown of other substances. Furthermore, reproduction related gene expression appears to be lowered in SR2 flies, as is neural transmission.

Resources and metabolic rate
The fact that starvation resistant lines possess more resources and can thus burn off more calories is clear from the high levels of lipid and carbohydrate down breaking gene expression levels. Earlier research in the related lines SR3 and SR4 showed that glycogen content was not higher in these lines than in control lines from eclosion onwards, but was increased considerably upon feeding. Here, we see that carbohydrate breakdown in line SR2 is higher which implies that carbohydrate levels will have been higher in this line than in C1 from eclosion onwards. Increased carbohydrate levels have been suggested to be associated with longevity through desiccation resistance (for a review, see Hoffmann and Harshman 1999). Therefore, we hypothesise that the long lived lines, SR1 and SR2, have a considerably higher desiccation resistance than lines, SR3 and SR4. Carbohydrate levels of lines SR1 and SR2 could provide additional mechanisms to increase longevity. Also, from the down-regulation of glycolysis and increased breakdown of carbohydrates it can be deduced that free glucose levels will be higher. This semi-diabetic phenotype matches that found in Broughton et al. (2005).

The burning off of excess resources explains why in Baldal et al. (2006) we found that SR2 females did not have a reduced metabolic rate after three days of starvation and had a higher metabolic rate than control lines. This suggests that in SR2 increased resources are required for increased starvation resistance. Also they explain the longer development time (unpublished result).

Proteolysis is also higher in SR2. Together with the breakdown of sugars other than glucose, it seems that SR2 flies clean themselves from excess resources to maintain homeostasis and prevent damage from secondary metabolites in the cell. Though this is speculative, it at least indicates that the flies are gaining energy from substances other than glucose. The fact that insulin signalling, glycogenesis and glycolysis related genes are expressed at a low level in SR2 indicates that the animals have down regulated these physiological processes associated with glucose metabolism. Here, we hypothesise that the high metabolic rate may be attributed by the breakdown of resources and surplus material.

Reproduction
Insulin signalling is known to regulate reproductive output in nematodes (Dillin et al. 2002) and insects (Wu and Brown 2006). SR2 is a line for which increased starvation resistance comes with an increased longevity. Here, reproduction related gene expression is suggested to be lower in the starvation resistant line. This is in line with the disposable soma theory of ageing (Kirkwood 1977; Kirkwood and Holliday 1979; Kirkwood and Rose 1991). Where reproduction is thought to trade off with somatic maintenance, therewith reducing longevity. Such a physiological trade off between starvation resistance and reproductive output has been shown to exist in *D. melanogaster* (Chippindale et al. 1993). To our knowledge this is the first study that links genome-wide expression data to the disposable soma theory of ageing.

Neural transmission
Starvation selected flies have a lower expression of genes involved in neural transmission. Earlier findings of Alcedo and Kenyon (2004) showed that ablation of specific gustatory nerve cells resulted in extended longevity in *C. elegans*. Broughton et al (2005) showed a similar increase in longevity in *D. melanogaster* flies whose insulin producing cells in the brain had been ablated. The signal that food is present (insulin like peptides) is not transmitted in the absence of these neurons in the flies. In the case of the starvation resistant lines, it may be so that neural transmission is reduced from eclosion on, as a result of selection for starvation resistance from eclosion onwards. This may result in overall lower insulin signalling, directing resource allocation to the soma instead of to reproduction and mitigating glucose metabolism, therewith extending longevity.

The high glucose levels that were found in the study of Broughton et al. (2005) are also hinted at in this study. Carbohydrate catabolic genes are expressed at a high level, whereas glycolysis specific genes are expressed at a low level. Because of this high supply, low demand situation, it can be expected that glucose levels will rise. This resembles mammalian pathological diabetes, which would, in mammals, lead to a decrease in life span. Broughton et al (2005) have acknowledged this fact. They explain that mammalian life span reduction as a result of diabetes comes from pathologies that do not occur in fruit flies, such as vascular damage.

Hypothesis

On the basis of these findings we hypothesise that as a result of selection our starvation resistant line has a lower body-wide neural transmission in the absence of food. Insulin producing cells in the *corpora cardiaca* (Kim and Rulifson 2004) produce less insulin, which results in the down regulation of the insulin pathway. This in turn results in down regulation of glycolysis and consequently, the down regulation of reproduction. The resources are then allocated to the soma side of the disposable soma theory. Because of the increased amount of resources, breakdown can continue for a longer time. Carbohydrates are effectively broken down, but the down regulated glycolysis will be a rate limiting step, through which energy will be generated over a longer time. This would have to lead to an increase in free glucose and a diabetic hyper phenotype. Figure 5 describes the pattern of changes in gene expression in broad classes of genes and pathways. More detailed analysis on these genes is needed and should be done using quantitative realtime PCR, and include

time series of gene expression. Despite the generality of the model, this study indicates that the physiological processes that have been identified in mutant studies in a variety of model organisms, are also underlying the responses to selection in lines that contain only naturally occurring variation.

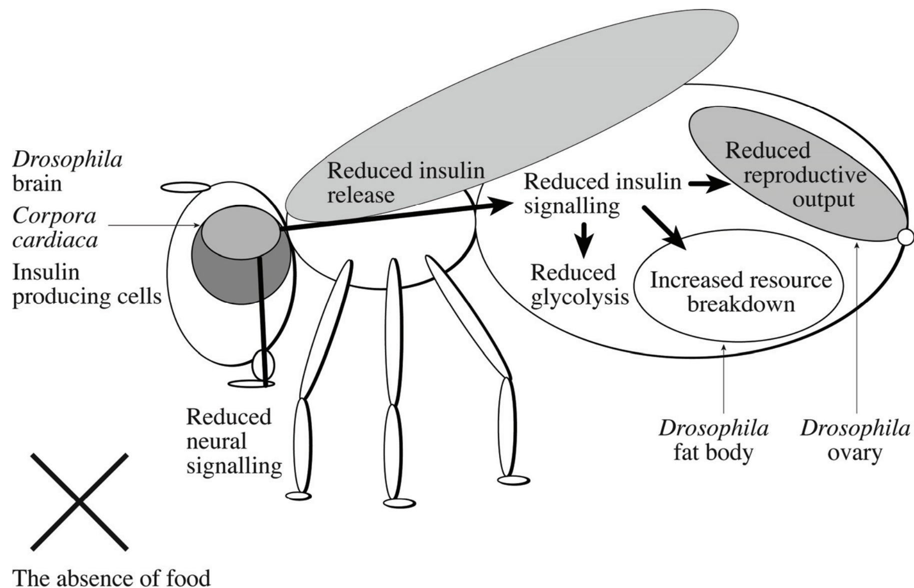

Figure 5. Schematic representation of what hypothetically happens in a female *Drosophila* in the absence of food.

Suggestions and a short outline of future research

Researchers should be aware of the technical variation that may occur in these studies. This noise can largely be overcome by synchronising experiments and controlled analysis. We, therefore, advise that in micro-array experiments all actions on samples that are to be compared take place at precisely the same time for all samples. Thus, the whole protocol for all samples should be completely synchronous. Batches of arrays that have been processed at different times should be normalized independently before analysis. In the case of time series and large sample sizes this may not be feasible. We then advise to use internal references in each separate batch.

Independent replication of the experiment (both biologically and technically) is necessary to be able to examine whether the same genes are found. Genes displaying a similar response can be regarded as being involved in the feature under study. A future project with Unilever will provide this replicate for the present pilot study.

Since pathway information is scarce for the *Drosophila* genome chips, in future analyses of this set of arrays and the replicate it would be interesting to use the human and murine orthologues and use those pathways as candidates for those involved in *Drosophila* longevity and starvation resistance. The orthologues can be found using the Ingenuity program. Also, this exercise may reveal genes of importance that have not been annotated yet.

Expression patterns of genes of interest should then be analysed in more depth by performing QRT-PCR on the remaining RNA samples. Furthermore, performing QRT-PCR on novel biological replicates will add statistical power and proof for the involvement of these genes in longevity and/or starvation resistance. Nevertheless, this pilot study illustrates the potential of applying micro-array technology to the analysis of selected lines and traits such as starvation resistance. More in depth analysis of the change in processes should be done in a time series.

Appendix 1. Gene ontology categories of genes that were differentially expressed between the genotypes under fed and starved conditions. The right hand column gives the percentage of genes involved in the GO category. Percentages add up to over 100 % because some genes are involved in several processes and thus add up to the percentage multiple times.

Fed	percentage (of 12)
transcription	67
regulation of transcription	58
regulation of nucleoside, nucleobase, nucleotide and nucleic	58
central nervous system development	42
neuroblast cell fate determination	25

Starved	percentage (of 79)
cellular protein metabolism	25
female gamete generation	23
transcription	20
imaginal disc metamorphosis	18
intracellular protein transport	18
eye morphogenesis	18
protein transport	18
cytoskeleton biogenesis and organisation	18
regulation of transcription	18
regulation of nucleoside, nucleobase, nucleotide and nucleic	18
eye-antennal disc development	15
peripheral nervous system development	13
phosphate metabolism	11
mesoderm development	11
sensory perception	11
ectoderm development	11
detection of abiotic stimulus	10
photoreceptor cell morphogenesis	10
eye photoreceptor cell differentiation	10
cell migration	10
wing morphogenesis	10
photoreceptor cell development	10
enzyme linked receptor protein signalling pathway	10
wing disc development	10
response to pest, pathogen or parasite	10
cellular macromolecule catabolism	10
M phase of mitotic cell cycle	10

Appendix 2. List of genes with known functions that showed gene-by-environment interactions. Listed are the unigene code, gene name, biological function or molecular function (indicated with an asterisk) and whether C1 and SR2 up or down regulated the expression of this gene under starved conditions relative to feeding.

Unigene	Gene name	Biological function	C1	SR2
NM_137810	CG3290	*alkaline phosphatase activity	down	equal
NM_137543	Juvenile hormone epoxide hydrolase 3	defense response; juvenile hormone catabolism	down	equal
NM_139861	CG8562	proteolysis and peptidolysis	down	equal
NM_079385	Multiple inositol polyphosphate phosphatase 1	*phosphoprotein phosphatase activity	down	equal
NM_142592	CG4462	extracellular transport	down	up
NM_080517	Turandot A	response to stress	down	up
NM_142648	CG5494	*structural constituent of cuticle	down	up
NM_143450	Odorant-binding protein 99a	*odorant binding	down	up
NM_079862	Alkaline phosphatase 4	*alkaline phosphatase activity	down	up
NM_142674	CG10827	*alkaline phosphatase activity	down	up
NM_136652	CG1809	*alkaline phosphatase activity	down	up
NM_134574	CG1304	proteolysis and peptidolysis	down	up
NM_140084	CG18180	proteolysis and peptidolysis	down	up
NM_136829	CG12350	proteolysis and peptidolysis	down	up
		stronger response in SR2		
NM_134831	CG3609	*oxidoreductase activity	up	up
NM_057277	Larval visceral protein D	glucose metabolism	up	up
NM_057280	Larval visceral protein L	glucose metabolism	up	up
		stronger response in C1		
NM_134818	CG17012	proteolysis and peptidolysis	down	down
NM_168021	Drosomycin-4	defense response	down	down
NM_141722	CG3940	*carbonate dehydratase activity	down	down
NM_079277	Astray	peripheral nervous system	up	up
NM_142777	CG7059, isoform A	glycolysis	up	up
NM_144443	CG18858	cholesterol metabolism; lipid metabolism	up	up
NM_135074	CG14031	*oxidoreductase activity	up	up
NM_142947	CG5933	RNA methylation	up	equal

Appendix 3. All the differentially expressed genes under fed conditions that have been described as genes, together with their Unigene code, gene name, molecular function or biological function (marked with an asterisk), and a column telling whether the gene was regulated at a higher or lower level in SR2 relative to C1.

Unigene code	Gene name	Molecular function	SR2 vs C1
NM_143563	CG1340	RNA binding; translation initiation factor	Higher
NM_080172	CG1856, isoform E	specific RNA polymerase II transcription factor activity; transcriptional repressor activity	Higher
NM_078990	Vismay	transcription factor activity	Higher
NM_057861	Type III alcohol dehydrogenase	alcohol dehydrogenase activity	Higher
NM_057950	Chitinase 2	chitinase activity; hydrolase activity,	Higher
NM_143456	Odorant-binding protein 99b	*Odorant binding	Higher
NM_132240	Companion of reaper	*Reaper associated	Higher
NM_057930	CG6955	structural constituent of larval cuticle	Higher
NM_057271	Larval cuticle protein	structural constituent of larval cuticle	Higher
NM_078533	Chorion protein 38	structural constituent of chorion	Lower
NM_140389	CG10154	structural constituent of peritrophic	Lower
NM_139416	CG13937, isoform A	HNK-1 sulfotransferase activity	Lower
NM_134831	CG3609	oxidoreductase activity	Lower
NM_137030	CG4812	serine-type peptidase activity; trypsin	Lower
NM_166152	CG8448, isoform D	ATPase activity, coupled	Lower
NM_057280	Larval visceral protein L	alpha-glucosidase activity; transporter activity	Lower
NM_135270	CG5177	trehalose-phosphatase activity	Lower

Appendix 4a. List of genes associated with the MAPKKK and JNK cascades. The biological function and whether the gene of interest had a higher or lower expression level in SR2 than in C1 are also listed.

Unigene code	Gene name	Biological function	SR2 vs. C1
NM_057238	Jun-related antigen	JNK cascade; MAPKKK cascade	Lower
NM_142288	CG14895, isoform A	MAPKKK cascade	Lower
NM_057532	Twins	MAPKKK cascade; response to stress	Lower
NM_140830	Mitogen-activated protein kinase phosphatase 3	MAPKKK cascade	Lower
NM_080005	Rab-related protein 4	MAPKKK cascade	Lower
NM_057750	CG8416, isoform B	JNK cascade	Lower
NM_057960	CG7693, isoform A	MAPKKK cascade	Lower
NM_079821	Protein C kinase 98E	MAPKKK cascade	Lower
NM_080050	CG12437, isoform B	regulation of JNK cascade	Lower

Appendix 4b. insulin signalling pathway associated genes.

Unigene code	Gene name	Biological function	SR2 vs. C1
NM_164899	Chico	insulin-like growth factor receptor binding	Lower
NM_080382	Protein kinase 61C	protein kinase activity	Lower
NM_057366	CG2621, isoform A	glycogen synthase kinase 3 activity	Lower

Appendix 4c. Reproduction associated genes that were differentially expressed between SR2 and C1.

Unigene code	Gene name	Biological function	SR2 vs. C1
NM_057670	Myocyte enhancing factor 2	ovarian follicle cell development	Lower
NM_167558	CG32562	oogenesis	Lower
NM_080322	New glue 1	oogenesis; reproduction	Lower
NM_080382	Protein kinase 61C	oogenesis	Lower
NM_057366	CG2621, isoform A	oogenesis	Lower
NM_057594	Cactus	oogenesis	Lower
NM_080364	Bunched	oogenesis; ovarian follicle cell development	Lower
NM_079659	Cheerio	female germ-line cyst encapsulation; ovarian ring canal formation	Lower
NM_080303	Wings apart-like	female meiosis chromosome	Lower
NM_139364	CTP	oogenesis	Lower
NM_057469	Nebbish	oogenesis	Lower
NM_078659	Bazooka	follicle cell adhesion; maintenance of oocyte identity; oocyte cell fate determination	Lower
NM_164899	Chico	egg chamber growth;	Lower
NM_080084	Spinster	courtship behavior; regulation of female receptivity	Lower