Mutational pleiotropy and the strength of stabilizing selection within and between functional modules of gene expression

Julie M. Collet*, Katrina McGuigan, Scott L. Allen, Stephen F. Chenoweth, Mark W. Blows

School of Biological Sciences, The University of Queensland, 4072, Australia

1 current affiliation: CBGP, INRA, CIRAD, IRD, Montpellier SupAgro, Univ. Montpellier,

Montpellier, France

Data Archival Location: NCBI Gene Expression Omnibus (N lines: GSE49815; S lines: GSE54777).
Variational and functional modularity

**Keywords:** modularity, pleiotropy, genotype-phenotype map, *Drosophila serrata*, gene expression

* corresponding author: j.collet@uq.edu.au

**ABSTRACT**

Variational modules, sets of pleiotropically covarying traits, affect phenotypic evolution, and therefore are predicted to reflect functional modules, such that traits within a variational module also share a common function. Such an alignment of function and pleiotropy is expected to facilitate adaptation by reducing the deleterious effects of mutations, and allowing coordinated evolution of functionally related sets of traits. Here, we adopt a high-dimensional quantitative genetic approach using a large number of gene expression traits in *Drosophila serrata* to test whether functional grouping, defined by GO terms, predicts variational modules. Mutational or standing genetic covariance was significantly greater than among randomly grouped sets of genes for 38% of our functional groups, indicating that GO terms can predict variational modularity to some extent. We estimated stabilizing selection acting on mutational covariance to test the prediction that functional pleiotropy would result in reduced deleterious effects of mutations within functional modules. Stabilizing selection within functional modules was weaker than that acting on randomly grouped sets of genes in only 23% of functional groups indicating that functional alignment can reduce deleterious effects of pleiotropic mutation, but typically does not. Our analyses also revealed the presence of variational modules that spanned multiple functions.
Pleiotropy has the potential to generate substantial evolutionary costs that scale with the number of traits affected by each mutation. Assuming a mutation has the same magnitude effect on each trait, the probability that a mutation will be favourable decreases as the number of traits (n) influenced by a mutation increases (Fisher 1930). Under the assumption of universal pleiotropy, the rate of adaptation may also decline by a factor of $n^{-1}$ (Orr 2000).

Modularity has been proposed as a mechanism to reduce such potential costs of organismal complexity (Wagner 1996). Variational modules occur when phenotypic traits share genetic variance through pleiotropy, while displaying lower covariation with traits belonging to different variational modules (Wagner 1996; Schlosser and Wagner 2004; Wagner et al. 2007). Independently, functional modularity describes an architecture where traits within a functional module share a common function (Wagner et al. 2007), which implies that the effects on fitness of a trait depend on the other traits contained in its functional module.

Functional integration is predicted to select for variational modularity because: (i) variational modularity reduces the range of effects of deleterious mutations, as mutations would only affect the traits belonging to the targeted functional module rather than the entire organism, (ii) all traits of a module can respond to natural selection as a unit, and (iii) it preserves the module’s function during evolutionary change (Olson and Miller 1958; Wagner and Altenberg 1996; Wagner and Mezey 2004). Several different approaches to theoretical modelling have determined that coincidence of functional and variational modules can indeed promote adaptation under some conditions (Welch and Waxman 2003; Griswold 2006; Pavlicev and Hansen 2011; Melo and Marroig 2015).
Despite the intuitive, and theoretically supported, potential evolutionary benefits of variational modules coinciding with functional modules, it is difficult to establish empirically whether functionally related sets of traits correspond to variational modules. Macro-evolutionary approaches can identify evolutionary modules (sets of traits that can potentially evolve independently of other such sets) as groups of genes with conserved physical proximity, co-occurrence in the genome, or fused genes. If variational modules help preserve functional modules, selection will favour the conservation of those modules and thus evolutionary modules are also expected to coincide with functional modules (Cheverud 1984). Comparative genomic analyses of functional modules have provided some evidence that evolutionary modules are, overall, more stable between genes that interact functionally than between unrelated genes, although there are large discrepancies between modules, and some functions are more conserved than others (Spirin et al. 2006; Peregrin-Alvarez et al. 2009; Snell and Huyten 2004; Moreno-Hagelsieb and Jokic 2012). Evolutionary preservation of functional modules may also cause differential rates of gene sequence evolution between functional modules. Consistent with this, Chen and Dokholyan (2006) reported that, in yeast, protein sequences and expression levels within functional modules evolve at more similar rates than between modules.

Coordinated selection that operates on functional inter-relationships of traits can preserve functional modules among taxa. By comparing the topologic structure of metabolic networks among species and the environment where they live, it was found that metabolic network modularity varies with environmental conditions in bacteria (Parter et al. 2007; Kreimer et al. 2008), archae (Takemoto and Borjigin 2011) and across archae, bacteria and eukarya (Mazurie et al. 2010), consistent with the role of natural selection in shaping
variational relationships between traits belonging to functional modules (note that results
using metabolic network modularity should be taken with care as the same modularity
scores can be obtained with very different metabolic network structures ZHOU and NAKHLEH
2012). Preservation of interactions between traits belonging to functional modules might
also reflect a bias in the variation generated by mutation. However, potential biases in
mutation rates within modules, and how mutation can contribute to maintaining the
organisation of functional modules, remain little explored.

Micro-evolutionary studies within species offer the opportunity to directly determine the
distribution of phenotypic effects of new mutations, and selection on these mutations.
Identifying variational modules requires studying many traits, belonging to diverse
functional modules. Morphological studies, for example, have limited use for testing the key
hypothesis: statistical tests for lower covariance outside the module are relatively weak
because only relationships among similar types of traits (morphology) are considered, and
covariance between morphological and behavioural or physiological traits, which might
contribute to a common function, are not considered. Systems biology, which involves a
shift in focus from the function of individual genes in isolation to the interaction among
gene products to achieve a biological function (IDEKER et al. 2001; KITANO 2002; CIVELEK and
LUSIS 2014), provided the tools to study modularity across diverse functional modules. Work
in model organisms has empirically revealed the function, and functional interactions of
thousands of genes, valuable information that can be applied to non-model taxa through
tools such as database of The Gene Ontology (GO) Consortium (ASHBURNER et al. 2000) and
the Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway resources (KANEHISA et al.
2006).
Gene expression data allow questions about the distribution of variational and functional modules to be asked while considering a diverse range of biological functions. The development of high-throughput methods for simultaneous measurement of the expression of thousands of genes lead to statistical developments allowing identifying clusters of co-expressed genes (e.g. Eisen et al. 1998; D'Haesembleer et al. 2000; Tanay et al. 2002). When gene expression has been measured within a quantitative genetic breeding design, high levels of covariance among expression traits were found, which imply the existence of variational modules (Denver et al. 2005; Rifkin et al. 2005; Ayroles et al. 2009; Blows et al. 2015). Many studies have subsequently applied enrichment analyses to determine whether co-expressed genes were associated with the same GO terms more often than expected by chance (e.g. Ayroles et al. 2009; Blows et al. 2015; Storz et al. 2015; Rose et al. 2016).

Although it appears from these studies that variational modules can predict functional group membership to some extent, it is unknown what proportion of pleiotropic covariance resides within predefined functional groups, and whether pleiotropic covariance among functional groups is of a similar magnitude to that found within functional groups. Further, the role of selection in generating these variation modules remains relatively unexplored.

Here, we adopt a high-dimensional quantitative genetic approach to directly test whether trait functional associations predict the pleiotropic covariance of gene expression traits. Using 41 mutation accumulation lines of Drosophila serrata in which widespread mutational pleiotropy among small random sets of 3385 gene expression traits has been demonstrated (McGuigan et al. 2014b), we identified 13 groups of genes involved in a particular function (using GO terms and KEGG pathways), that were significantly enriched with genes that had mutational variance, and that showed significant mutational covariance; the opportunity for
genes to contribute to more than one of the 13 groups was explicitly limited. Using this set of Mutation Accumulation (M) lines and a matched set of set 30 inbred lines derived from an outbred population of *D. serrata*, we determined whether co-expression, which was due to pleiotropic effects of new mutations in the M lines or of alleles already present in the inbred lines, was greater within these functionally related sets of genes than it was among random sets of the same number of genes whose functions were not known. We also investigated whether we could detect pleiotropic mutations that spanned a larger array of functions. In addition to directly testing whether pleiotropic effects were stronger within functional modules, our experimental design (specifically the matched estimates of mutational and standing genetic variance) allowed us to directly estimate the strength of stabilizing selection acting against function-specific mutational pleiotropy and compare it to selection against mutational pleiotropy spanning multiple functions.

**MATERIAL AND METHODS**

**Experimental Populations and Data Collection:**

We used two sets of highly inbred lines of *Drosophila serrata*, which were established, maintained and assayed in a similar way (McGuigan et al. 2014a), to measure the mutational and standing genetic covariance in sets of gene expression traits assigned to functional groups. Briefly, the first set of lines consisted of 45 mutation accumulation lines (hereafter called M lines) derived from a single inbred ancestral population. Each M line was maintained during 27 generations through full-sib inbreeding. The second set of 42 lines were derived from females collected from a wild outbred population and inbred for 15 generations of full-sib mating (G lines). As M lines were founded from a population depleted
of standing genetic variance, the differences among M lines originate from mutations, filtered by relaxed selection, and M lines captured mutational variance. On the other hand, the differences among G lines originated from the natural standing genetic variance captured in the original outbred population; hence the G lines captured standing genetic variance. We neglected the contribution of new mutations accumulated in the G lines because, as determined from the M lines, mutational variance is small relative to standing genetic variance (Table 1).

In generation 26 in the M lines and generation 14 in the G lines, we set up four replicate vials per M line and three replicate vials per G line in preparation for RNA extraction in the following generation. From these vials, 40 (M) and 60 (G) virgin male offspring were collected for each line. The use of replicate vials for each line ensured that micro-environmental variation was not confounded with the among line (mutational or genetic) variation. Males were held in groups of five until three (G lines) or four (M lines) days after emergence, when two RNA extractions on a subsample of 20 (M lines) or 30 (G lines) males were conducted. Given the scope of this study (ratio of variance between the two treatments compared between functions and within functions), we do not expect any of these small discrepancies in handling the two sets of lines to impact our results. Flies were snap-frozen using liquid nitrogen; total RNA was extracted using TRIzol (Invitrogen, Carlsbad, CA) and purified using RNeasy kits (QIAGEN, Valencia, CA), all according to manufacturer’s instructions.

As detailed in Allen et al. (2013) and McGuigan et al. (2014b), microarrays designed from a D. serrata EST library (Frentiu et al. 2009), were manufactured by NimbleGen (Roche); cDNA synthesis, labelling, hybridization and microarray scanning (NimbleScan) were performed by
the Centre for Genomics and Bioinformatics, Bloomington, Indiana. Microarrays contained 20K random probes plus 11604 features from ESTs, targeted by five 60-mer oligonucleotide probes; each probe appeared twice on each array. A single sample was hybridized to each array, with a single color (Cy-3). Twelve arrays appeared on each slide, and samples were randomly assigned to a slide. We performed quality control analyses with the oligo package in Bioconductor (GENTLEMAN et al. 2005) and removed data due to poor quality or high background signal. The expression data for the 41 M lines and 30 G lines remaining after this process are analysed below, and are available through the NCBI Gene Expression Omnibus (EDGAR et al. 2002; BARRETT et al. 2011) (M: GSE49815, G: GSE45801).

To determine which of the 11604 phenotypes on each microarray were associated with mutational or standing genetic variation, and therefore informative for analysis of variational and functional modules, we use linear mixed model analyses to partition the variance in each expression trait. First, a linear mixed model to characterize the mutational variance in the standardized (mean = 0, SD = 1) log_{10} mean expression of each probe (MCGUIGAN et al. 2014b) was implemented within a Restricted Maximum-Likelihood framework using the function lme in R 3.1.2 (library nlme, PINHEIRO and BATES 2000);

$$Y_{ijk} = \mu + SF + Line_i + Rep_{j(i)} + \epsilon_{ijk}$$

(1)

where SF was a fixed effect for a segregating factor observed in the M lines that must have been present in the ancestor, and was therefore not a product of mutation during the experiment (See SI in MCGUIGAN et al. 2014b). To prevent this factor from contributing to estimates of mutational (among-line) variance, we added a fixed effect to remove the mean difference in trait expression between the two groups of lines with the alternative forms of this segregating factor. Line was a random factor representing the among M line variance,
Rep was a random factor nested within lines representing the two replicate extractions per line, and the residual error, $\epsilon$, was the variance among the five probes per gene. To be consistent with the bivariate analyses (see below), we used the General-purpose optimization based on Nelder-Mead algorithm (method “optim”). From the Line variance component, we calculated the broad sense mutational heritability of gene expression for each gene according to $H^2_M = V_B^2 / 2t$ where $t$ is the number of mutation accumulation generations, here 27 (see details in McGUIGAN et al. 2014b). Of the 11604 gene expression traits analysed, 3385 showed among line variance in the M lines, with statistical support for mutational variance in 1035 traits, 533 of which remained significant at the 5% false discovery rate threshold. Selecting genes according to a significance threshold, particularly in large datasets, is prone to limit the detection of pleiotropy (HILL and ZHANG 2012). We did not wish effect size and subsequent type II error to affect the interpretation of our measure of function-specific pleiotropy. We therefore selected the entire subsample of 3385 genes showing non-zero mutational variance for further consideration. Second, the analysis of the G lines to estimate the standing genetic variance in each of the 3385 expression traits with non-zero mutational variance followed the same model, but removing the segregating factor effect. As we detected consistent differences in mean signal intensity among the five replicate probes of each gene in the G lines, we included a fixed effect for probe when analysing the G lines (McGUIGAN et al. 2014b).

**Expressed sequence tag (EST) annotation:**

To establish the function of the genes present on our microarrays, *D. serrata* ESTs were first classified as putative *Drosophila melanogaster* homologs. *D. melanogaster* was chosen as
the reference as it is a relatively well-annotated model organism that has been incorporated into many analysis tools. Homolog IDs were assigned following BLAST of the *D. serrata* ESTs to 12 *Drosophila* species’ genomes obtained from FlyBase (McQuilton *et al*. 2012) followed by mapping of the FlyBase gene IDs to homolog IDs from the ortholog database (Waterhouse *et al*. 2013), which allowed us to identify *D. melanogaster* homologs of the *D. serrata* genes even when they had a BLAST hit to a gene in a taxon other than *D. melanogaster*. The *D. melanogaster* gene’s annotations were then assigned to the *D. serrata* gene and used for enrichment analysis. To allow for highly divergent homologs to be identified, tblastx with a liberal e-value threshold of 10 was applied; median e-value was $1 \times 10^{-78}$. The method identified 10843 (93%) genes on the microarray as homologs of 9500 *D. melanogaster* genes. For some genes, several *D. serrata* ESTs corresponded to a unique *D. melanogaster* homolog, and we established that those *D. serrata* genes could be considered as independent (See Supplementary Information).

**Enrichment analyses and identification of functional groups**

We restricted our investigation of potential GO functions to those functional groups that were the best candidates to harbour function-specific mutational variance. If a functional group experienced a function-specific pleiotropic mutation, several (or perhaps all) of the genes that are part of that function will each exhibit univariate mutational variance. Therefore, we performed Gene Ontology term and KEGG pathway enrichment analyses that detected functional groups within the 3385 genes potentially affected by mutations (non-zero mutational variance) compared to all 11604 genes contained on the microarray. Of the 3385 variable genes, 2783 were identified as unique *D. melanogaster* homologs, of which 2604 were assigned a gene ID using DAVID 6.7 (Huang *et al*. 2009). Of the 9500 genes
contained on the microarray that were assigned a unique *melanogaster* genes ID from Flybase genes, 8978 were recognized and assigned an ID in DAVID. Therefore, we were able to apply the enrichment analysis to 2604 genes of interest (i.e., mutationally variable candidates for functional groups) against the reference background containing 8978 genes. We did not consider enrichment on *Cellular Component* GO domain for two reasons. First, our goal in this paper was to identify functional modules for which we had the greatest expectation of correspondence to pleiotropic variational modules, and while both *Biological Process* and *Molecular Function* domains are likely to capture multiple functions of a single gene (He and Zhang 2006; Su et al. 2010), co-localisation of gene products has a less direct link to definitions of pleiotropy. Second, gene groups identified through the *Cellular Component* GO domain might be more likely to be affected by Pearson’s rule of neighbourhood, where geographically closer elements are more likely to be correlated than distant elements (Whiteley and Pearson 1899), which is a known source of bias in interpretations of coincidence of functional and variational modules at the phenotypic level (Mitteroecker 2009). Therefore, we only looked for enrichment in *Biological Process* and *Molecular Function* in the GO terms, and analysed the list for GO terms and KEGG pathways with the ‘Functional Annotation Chart’.

Functional groups defined by GO terms typically contain a lot of genes in common with other groups because of parent-child relationship between GO terms, because the same set of genes can belong to different types of functional groups (i.e., a Molecular Function and a Biological Process), or because similar biological processes can be categorised in different regions of the gene ontology. This is a commonly recognised problem for using analyses of GO terms to identify pleiotropic (multifunctional) genes. Most methods developed to deal
with the lack of functional uniqueness or independence of GO terms use indices measuring the semantic similarity between terms that share common ancestors or children (see comparison of methods in Harispe et al. 2014), and more rarely the similarity between terms using their probability of sharing genes (Pritykina et al. 2015). Here, we limited gene sharing among our GO term defined functional groups to maximise our chances of detecting a difference in pleiotropic covariance within versus among functional groups. We achieved this using the following method. We started by selecting the group with the lowest enrichment \( p \) value when testing mutational variance enrichment (GO term: group Cuticle, \( p = 3.3e^{-11} \), Table 1). Retaining this group with the lowest enrichment \( p \) value, we then sequentially compared all groups with increasing enrichment \( p \) values, retaining groups only if they shared less than 10% of their genes with any of the previously retained (i.e., lower enrichment \( p \) value) groups. Fourteen GO terms with low levels of gene composition overlap were selected using this process including seven molecular functions and seven biological processes, spanning structural molecule activities (two groups), metabolic processes (three groups), developmental processes (two groups), responses to stimulus (two groups), or binding (two groups) (Table 1). However, one of the 14 identified groups failed a randomization test to determine if the observed mutational covariance was above the random expectation when the mutational covariance among traits was destroyed (See Supplementary Information). We removed this group from further analysis. We further identified two independent significantly enriched KEGG pathways, but they identified the same putative functional modules as identified by the GO terms enrichment analysis, and we only report GO terms results here (see Supplementary Information).
We tested whether the 434 genes assigned to the 13 functional groups were a representative sample of the mutational and standing genetic variance found in the entire set of mutationally variable genes. For each functional group, we used Wilcoxon signed-rank tests to compare the mutational and standing genetic variance of the genes contained within that group to 1000 sets of the same number of genes, randomly chosen from the 2951 genes with non-zero mutational variance that were not assigned to any of our 13 studied functional groups.

**Mutational and standing genetic covariance within functional groups**

We determined the genetic covariance between each pair of expression traits using bivariate models within each functional group of genes in both the M and G lines separately. For each pair of genes within a functional group, we implemented the bivariate form of the model (1):

\[
\mathbf{Y} = \mathbf{\mu} + \mathbf{Z}_l \mathbf{\delta}_l + \mathbf{Z}_r \mathbf{\delta}_r + \mathbf{\varepsilon}
\]

where \( \mathbf{Z}_l \) and \( \mathbf{Z}_r \) are design matrices for the line and replicate within-line random effects. We modelled the covariance structure among traits at the line (\( \mathbf{\delta}_l \)) and replicate (\( \mathbf{\delta}_r \)) levels, using unstructured 2 X 2 covariance matrices, and \( \mathbf{\varepsilon} \) was a diagonal matrix containing the residual (among-probe) variances for each trait. The segregating factor and a probe level fixed-effect were included as before for the M line and G line analyses, respectively.

For each functional group, we used the among-Line univariate variance component (model 1) and covariance component (model 2) from the respective M and G lines to construct mutational (\( \mathbf{M} \)) and genetic (\( \mathbf{G} \)) variance-covariance matrices. Constructing the \( \mathbf{M} \) and \( \mathbf{G} \) matrices enabled us to test for the presence of shared genetic variance in any multivariate
trait combination (not just pairwise combinations) in the most efficient fashion by focusing on the presence of mutational or genetic variance in the major axes represented by the dominant eigenvectors of the respective matrices of each functional group.

Our goal was to detect whether the levels of mutational and standing genetic covariance found in our candidate functional groups were specific to that function and exceeded the typical levels of mutational and standing genetic covariance found in traits that were not identified by the enrichment analysis as sharing the same function. To achieve this, we compared the mutational or genetic variance in the dominant eigenvectors of a matrix ($m_{max}$ and $m_2$ for mutational variance, $g_{max}$ and $g_2$ for standing genetic variance; Table 2) of a given functional group, to the observed eigenvalues of a distribution of eigenvalues that represented biological variation characterised by the background level of mutational or genetic covariance among random sets of genes. We first created 150 datasets of 73 genes that were sampled at random from the subset of 2951 genes that showed mutational variance but were not assigned to any of the 13 functional groups under consideration. We did not know the functions of the genes included in these datasets. Thus, some of the genes included in each dataset may share a common function, but it was unlikely that they all had the same function. Moreover, none of the genes included in those datasets were assigned to any of the GO terms in Table 1. Sampling was done with replacement, so that some genes were included in several of the 150 data sets. For each of the 150 created datasets, we estimated $M_{bb}$ and $G_{bb}$ covariance matrices (Table 2) by applying the univariate (model 1) and bivariate (model 2) mixed models, as described above. To obtain $M_{bb}$ and $G_{bb}$ covariance matrices of a size that matched each functional group (i.e. the number of genes in each functional group, Table 1), we randomly selected the number of genes required for
each functional group from the 73 genes in each data set. Each draw of the appropriate
number of genes was sampled from the set of 73 genes with replacement, creating random
overlap between biological simulations of the 13 functional groups. The functional groups of
the same size (groups Chitin and Sensory, and groups Heme and Cuticle) were compared to
the same sample of biological background genes. A small number (seven) of the $2 \times 150 = 300$
$M_{bb}$ and $G_{bb}$ covariance matrices estimated in this fashion returned extreme outlying
eigenvalues ($>10^{10}$) after diagonalization, and we removed these data sets, resulting in 143
estimates of the biological background covariance in the M and G lines for each functional
group. Because several functional groups had average univariate genetic variances
significantly higher than the genes that were not selected in any functional groups (Table 1),
we could not directly compare the shared genetic variance contained in those functional
groups with the biological controls. Instead, we compared the proportion of variance
represented by each eigenvector, calculated by dividing each eigenvalue by the trace of the
matrix, to the 95% confidence interval of the proportion of variance for each vector
obtained from the biological controls. This comparison was used to determine whether the
genetic covariance found within functional groups was larger than among background
(random) genes, independent of the level of genetic variance displayed by each gene.

Selection on pleiotropic effects within functional groups

One of the primary evolutionary advantages of modularity has been predicted to be the
mitigation of the adverse effects of deleterious mutation through restricting the extent of
their pleiotropic effects. We would therefore predict that selection on the genetic variance
within functional groups should be weaker than selection among groups. We tested this
prediction using two approaches. First, we estimated $s$, the selection coefficient, from our
estimates of the mutational variance and standing genetic variance for gene expression,

using: \( s = \frac{V_M}{V_G} \) (Barton 1990; Houle et al. 1996). Specifically, we calculated \( s \) for the multivariate traits \( m_{\text{max}} \) and \( m_2 \) using:

\[
s = \frac{\lambda m_n / 54}{m_n^T G m_n}
\]

(3)

where \( \lambda m_n \) was the eigenvalue of the \( n \)th eigenvector of \( M \) divided by twice the number of generations to give the per-generation input of variance (Lynch and Walsh 1998) and \( m_n^T G m_n \) represents the projection of the normalized \( m_n \) vector through the standing genetic space of \( G \). To return \( s \) to the original log_{10} scale, we multiplied it by the ratio of phenotypic variances of \( m_n \) in the G and M lines. To do this, we used the linear equation of each of the \( m_n \) eigenvectors to generate phenotypic scores in both M and G lines. We kept the probe level of information by applying the linear equation of the eigenvectors five times; each of the five times we randomly chose one of the five probes that targeted each trait (without replacement). We analysed these phenotypic trait scores for the \( m_n \) index with model (1) to estimate the phenotypic variance associated with each \( m_n \) trait combination in the M and G lines. We repeated the same analyses for the groups of genes in the biological background. In three (out of 26) estimates for the two vectors in the observed data for the 13 functional groups, the calculation of \( s \) returned negative values, as \( m_n \) fell within the null space of the estimated \( G \) matrix. In these three cases, we bent \( G \) using nearPD (library Matrix, R 3.1.2) to obtain the closest positive definite matrix.

To directly estimate the amount of sampling error associated with the estimates of mutational and genetic variance in \( s \), it is necessary to take into account that the estimation of \( s \) can be inflated as a result of REML enforcing positive-definite constraints. To account
for this upward bias, we applied equation (3) to random pairs of the 50 estimates of $M_{se}$ and $G_{se}$ for each functional group representing the amount of sampling error generated by our experimental design for a given module size (See Supplementary information; Table 2). We then used these 50 estimates of $s$ representing sampling error to remove the magnitude of the inflating effect in our observed estimate of $s$ by taking the difference between the observed estimate and the median of these 50 estimates of sampling error (McGuigan et al. 2014a).

To estimate the 95% CIs of $s$ based on the biological background, we used the difference between the 143 observed biological background estimates and the corresponding median sampling error estimate obtained for each function specific group. Since the univariate estimates of mutational and genetic variance in the functional groups were on average higher than in genes included in the biological background (Table 1), the CIs for the biological background could be biased if the observed ratio of $\text{tr}(M)/\text{tr}(G)$ for a group differs from the set of 143 biological background estimates. For two groups (Chorion and Cuticle), the observed $\text{tr}(M)/\text{tr}(G)$ was higher than the biological background. In both cases, the lower and upper 95% CIs for these estimates are therefore likely to be smaller in magnitude than would occur if the traces of the observed matrices matched the traces of the matrices for the biological background (as shown by the negative lower confidence interval obtained in Chorion). Therefore, for those two groups, a lower observed estimate of selection than the confidence interval would be a conservative result.

As described above, estimates of $s$ are only informative of selection operating on the major axes of mutation, but selection may act instead on other dimensions. We therefore took a second approach to test the hypothesis that mutations acting within a functional group
were under weaker selection than mutations affecting functionally unrelated sets of traits.

We determined the level of overlap between $M$ and $G$ matrices using Krzanowski’s common subspaces approach (KRZANOWSKI 1979; AGUIRRE et al. 2014):

$$ S = A^T B B^T A $$ (4)

where $A$ and $B$ contain the two first eigenvectors of $M$ and $G$ as columns, respectively. The sum of the eigenvalues of the resultant matrix $S$ then ranges between 0 (complete orthogonality of the subspaces) and 2 (coincident subspaces). If selection is weak, $G$ is not predicted to have diverged from the structure generated by mutation, and a big overlap between $M$ and $G$ subspaces, and thus a higher $S$, is expected. This approach to estimating selection does not assume that it acts directly on the eigenvectors of $M$, but rather on any combination of expression traits contained within the first two dimensions of $M$ or $G$. Note that the upper and lower 95% random values of the sum of the eigenvalues of $S$ decrease with the size of the groups. Since we only considered the two first axes whatever the size of the group considered, it is expected that the proportion of the common subspace shared between two matrices in those two axes will be lower when those axes represent a smaller proportion of the total subspace.

RESULTS

Mutational and standing genetic variance in genes included in functional groups

The enrichment analysis of the 3385 genes showing non-zero mutational variance identified 13 independent candidate GO terms most likely to harbour function-specific mutational covariance. Enrichment for one term (Cuticle) passed the 5% fdr threshold. A total of 434
genes were assigned to the 13 functional groups, while the remaining 2951 genes with non-zero mutational variance had functions that did not fit our criteria to belong to candidate functions. Even though the enrichment analyses did not take into account the level of mutational or genetic variance, the median standing genetic variance in five of the 13 functional groups was significantly higher than expected from the other genes with non-zero mutational variance that were not assigned to our candidate functional groups (Table 1). Overall, the genes assigned to the 13 functional groups had significantly higher mutational variance (median $H_M^2 = 0.0013$ vs $0.0011$, median Wilcoxon $W$ for the 1000 tests $= 103,544$, median $p = 0.012$) and standing genetic variance (median $H_G^2 = 0.4469$ vs $0.3841$, median Wilcoxon $W$ for the 1000 tests $= 104,144$, median $p = 0.006$) than the genes not assigned to our functional groups (Table 1).

**Mutational and standing genetic covariance within functional groups**

We found little support for functionally related genes being universally affected by pleiotropic mutations. The functional group “structural constituents of chorion” was the only one of the 13 functional groups in which all genes assigned to that function had mutational variance above zero (Table 1, all eight genes from the DAVID term had above-zero mutational variance). Our enrichment analysis further showed that the functional group Chorion was the only GO term of the entire DAVID 6.7 repository for which all genes showed above-zero mutational variance.

We next tested whether functional modules were enriched in function-specific mutational covariance by determining if the observed mutational covariance within a functional module exceeded that of the biological background based on random, size-matched, sets of genes whose functions were not known. Only three functional groups showed levels of mutational...
covariance above that found among random sets of genes (captured by $M_{bb}$) (Pro-DNA, Heme, and Cuticle: Table 3). Furthermore, we found one functional group (NeuroT) that showed lower levels of mutational covariance than random groups of genes of the same size that did not belong to a common function (Table 3). Therefore, it appears that genes within a functional module are not typically affected by pleiotropic mutations to a greater extent than functionally unrelated sets of genes (Table 3).

To illustrate the variation in mutational pleiotropy represented by our functional groups, the mutational covariance structure of two groups of similar size (39 and 49 genes, groups Ion Tsp and Cuticle respectively) is graphically displayed in Figure 1. Most functional groups were similar to group Ion Tsp, where the mutational pleiotropy found in the first two axes did not exceed the level of pleiotropy found in random groups of genes of the same size (Table 3; Figure 1A), thus showing no sign for function-specific mutational covariance. In contrast, group Cuticle was an example of where mutational covariance exceeded the background level for $m_{max}$ (Table 3; Figure 1D). The difference in mutational covariance between groups Ion Tsp and Cuticle is clearly illustrated by the difference in the frequency of pairwise mutational correlations >0.99 within each of these functionally defined modules (Figure 1B, 1E).

Function-specific standing genetic covariance was even less common than function-specific mutational covariance. Only functional group Chorion showed higher levels of standing genetic covariance (on two axes) within the functional groups than what was seen among random sets of the same number of genes (captured by $G_{bb}$) (Table 3). In all other groups, the level of function-specific standing genetic covariance matched the typical levels of
standing genetic covariance found in the biological background as illustrated with groups Ion Tsp and Cuticle on Figure 1C and 1F.

**Stabilizing selection on mutational pleiotropic effects within functional modules**

To test whether mutations with pleiotropic effects on functionally related sets of traits were less deleterious than mutations with pleiotropic effects on sets of random traits we first estimated the selection coefficients ($s$) for both $m_{max}$ and $m_2$ for each of the 13 functional groups. One of the three $s$ obtained after bending $G$ (functional group Chorion) was extreme (1.306) as a consequence of the very small genetic variance in the direction of $m_{max}$; without groups with similar estimates of selection, we must consider this value as an outlier. The 25 remaining $s$ values for $m_{max}$ and $m_2$ ranged from 0.010 to 0.062, with a median $s$ of 0.024 and 0.022, respectively (Table 3). While these selection estimates are slightly stronger than the median reported for random sets of five expression traits (0.016; McGuigan et al. 2014a), the within functional module $s$ was within the range of what was typically observed for the same number of randomly chosen genes (i.e., estimates fall within the 95% CI estimated from $G_{bb}$ and $M_{bb}$; Table 3). Only one functional group, Cuticle, showed the predicted weaker within module selection for $m_{max}$ (Table 3).

Using an alternate measure of selection that did not restrict selection to occur only along $m_{max}$ or $m_2$, revealed three groups showing a higher common subspace than expected (Table 3), consistent with weaker selection within these functional groups than among random sets of genes. Importantly, two of those three groups showed levels of mutational covariance higher than the 95% CI interval of the mutational covariance obtained in the biological background (Table 3), and the third group exhibited mutational variance close to
the upper CI (Bacterium: $\lambda m_{\text{max}}$ 0.630, upper CI 0.631), indicating that considerable mutational covariance was generated, which was not subsequently removed by selection.

**Stabilizing selection on mutational pleiotropic effects among functional modules**

Overall, mutational and genetic covariance was not typically greater than observed in the biological background, suggesting that much of the mutational covariance was not restricted to functional modules. To investigate this observation further, we estimated among-function covariance contained in the 26 trait combinations with the greatest within-function covariance (i.e., $m_{\text{max}}$ and $m_2$ or $g_{\text{max}}$ and $g_2$ within each of the 13 functional groups). This approach of reducing the within-module complexity to allow us to explore the among-module relationships is analogous to the eigengene network analysis approach developed by Langfelder and Horvath (2007).

We used the linear equations of the first two eigenvectors of $M$ and $G$ of each of the 13 functional groups to generate phenotypic scores for these new index traits. We applied the same univariate (model 1) and bivariate (model 2) mixed models to obtain two 26 x 26 covariance matrices ($M_{26}$ and $G_{26}$, Table 2) that contained any mutational or genetic covariance between functional groups that occurred among the major axes of variance within functional groups. We repeated this procedure on shuffled data to generate 50 $M_{26se}$ and $G_{26se}$ (Table 2) matrices representing sampling error to enable an assessment of the significance of the eigenvalues of $M_{26}$ and $G_{26}$ and extent of covariance among functional groups (See details on how we obtained randomised distributions in Supplementary Information and McGuigan et al. 2014 b).
The first two eigenvalues resulting from the diagonalization of $M_{26}$ exceeded the magnitude expected solely from sampling error (Fig. 2A), indicating that among functional group covariance was greater than expected from sampling error alone. These first two of the 26 eigenvalues of $M_{26}$ represented 72.4% of the total mutational variance contained in the 26 traits. Therefore, most of the mutational covariance found within functional groups was also shared among the functional groups; as we explicitly limited the opportunity for the same gene to belong to different functional groups, this observation reveals strong mutational pleiotropic links among functions. For $G_{26}$, again only the first two eigenvectors had eigenvalues larger than expected from sampling error (Fig. 2B), but in this case only represented 40.1% of the variance included in $G_{26}$. Therefore, there appears to be less pleiotropic covariance among functional groups in the standing genetic variance than that generated when new mutations first arise in the absence of selection.

Hine et al. (submitted) recently applied a high-dimensional Bayesian sparse factor (BSF: Runcie and Mukherjee 2013) analysis to estimate the mutational covariance across all 3385 genes in our set of 41 M lines and here we compare this broad-scale distribution of variance to our functional groups to gain further insight into the distribution of variance across traits. The BSF model identified 21 factors that displayed significant mutational heritability, and explained 46% of the total estimated mutational variance. These 21 factors combined had significant contributions from 1263 of the 3385 genes. Figure 3 shows how the genes contained in each functional group analysed here contributed significantly (average error rate < 0.005) to each high-dimensional factor. All of the 13 functional groups contained at least one gene that was significantly associated with at least one of the 21 high-dimensional factors; the 21 variational modules identified from the BSF model were typically associated
with more than one functional group. This comparison indicates that genes within the
functional groups are almost always part of a wider variational module, with much of the
mutational variance within functional groups shared across groups within much wider
variational modules, as suggested by the analysis of $M_{26}$.

As modularity is expected to reduce the strength of selection acting within functional
groups, the trait combinations highlighted as displaying strong covariance among functional
groups are expected to be under particularly strong selection. We estimated selection
against the first two axes of $M_{26}$ by estimating the Krzanowski $S$ subspace comparison of the
2-dimensional subspaces of $M_{26}$ and $M_{26}$ in the G lines. Surprisingly, the sum of the
eigenvalues of the $S$ matrix was 1.11, considerably higher than the average $S$ found within
functional modules, suggesting that selection has had little effect on the mutational
covariance that is originally widespread among functional groups.

**DISCUSSION**

A genotype–phenotype map in which pleiotropic effects are relatively restricted to groups
of functionally integrated traits has been predicted to be evolutionarily beneficial (WAGNER
1996). Here, we provide a test of whether variational (pleiotropic) and functional modules
coincide, and whether functional relationships among genes influence the strength of
stabilizing selection acting on new mutations. Enrichment analyses assigned 434
mutationally variable genes to 13 independent candidate Biological Processes or Molecular
Functions that also formed variational modules. In general, expressional levels of groups of
genes related by function (GO terms) rarely had higher covariance than genes assigned to
groups at random. Similarly, mutations jointly affecting functionally related genes were
rarely under weaker selection than mutations affecting random sets of genes. Our analyses revealed widespread pleiotropic effects as we found consistent covariance among functionally unrelated traits (biological background) and we observed that pleiotropic effects spanning functional groups generate much of the mutational variance within functional groups. We now consider both the implications of these results, and potential caveats on our interpretation.

**Pleiotropy and selection on gene expression traits**

We found weak evidence that variational modules were restricted to functional modules; in only four of 13 functional groups did covariance within functional groups exceed the background level of either mutational or genetic covariance found among random, size-matched, sets of genes. Furthermore, most (72%) of the mutational covariance established within functional groups was shared among the functional groups. Again consistent with previous analyses of these data (McGuigan et al. 2014a) and studies in other taxa (Denver et al. 2005; Lemos et al. 2005; Rifkin et al. 2005), we observed that expression traits were generally under relatively strong stabilizing (purifying) selection that typically matched levels of selection reported for life history traits (Houle et al. 1996). However, the strength of selection on mutations affecting traits within a functional group was not typically statistically distinguishable from mutations affecting random groups of genes, not associated with a common function; only three of the 13 functional groups exhibited significant signatures of weaker than background stabilizing selection.

To what extent might our results reflect misidentification of functional modules? The lack of support for higher mutational or standing genetic covariance within GO terms might reflect missing or inaccurate information on gene function. GO terms are incomplete descriptions
of functional groups (i.e., not all genes involved in that function have necessarily been identified), allowing the potential for strong covariance among randomly chosen genes to reflect undetected functional modules. Further, as many genes are attributed to functional groups without experimental evidence (for example, in the well-studied *Arabidopsis thaliana*, only 39% of annotated genes have had their function determined by experimental evidence: Rhee and Mutwil 2014), the potential exists for genes to be incorrectly assigned to functional modules (Clark and Radiojac 2011), which would result in underestimation of covariance within putative functional modules. However, the joint analyses of co-expression and GO terms have successfully identified gene function in many studies (e.g. Luo et al. 2007; Nayak et al. 2009; Ayroles et al. 2011; Proost and Mutwil 2017). Although we cannot exclude the possibility that current information on functional groups is so poor that functional and variational modules will not coincide, it seems likely that pleiotropy may generate covariance among functions in an unpredictable manner to a substantial extent.

**Functional groups within wider variational modules**

While pleiotropy clearly exists within functional groups, it is the widespread and extensive nature of pleiotropy among the GO term defined functional groups, and among random sets of expression traits in general (Mcguigan et al. 2014b; Blows et al. 2015) that is the overriding feature of the mutational and standing genetic covariance in these expression traits. The 21 variational modules of gene expression uncovered by the high-dimensional Bayesian sparse factor model applied by Hine et al. (submitted) provided a very useful framework within which to interpret the mutational covariance among functional groups. At one extreme, genes from functional group NeuroT displayed within group mutational covariance that was lower than in the biological background and only three of its genes
were independently part of a wider variational module, indicating that mutations in these
genes had highly specific effects. At the other extreme, genes from functional group
Endopep had genes contributing to 14 of the 21 high dimensional variational modules. The
comparison of our analyses of functional group Cuticle to the 21 BSF is of particular interest.
Functional group “structural constituent of cuticle” was the only group to meet the all
expectations for correspondence between variational and functional modules: (i) this was
the only functional group to show significant enrichment after FDR correction; (ii)
mutational variance in $m_{max}$ was higher than in the biological background, and (iii) there was
weaker selection on mutational variance than expected from the biological background.
Group Cuticle had 19 (over 1/3) genes significantly contributing to the second common
factor from the Bayesian analysis. This second factor spanned multiple functions as 11 of
our 13 functional groups had genes significantly contributing to it. Thus, what appeared in
group Cuticle as aligned functional and variational modularity may reflect the contribution
of a functional module to a larger variational module.

The sizes of the variational modules in gene expression were considerably larger than the
relatively small number of genes contained by the functional groups explored here. A
minimum average variational module size to explain the extent of mutational covariance
among small random sets of these traits was predicted by McGuigan et al. (2014b) to be 70
genes, but it was also suggested that some variational modules were likely to be much
larger than this. Analysis of the transcriptome-wide covariance structure of the standing
genetic variation in the G lines indicated the presence of one variation module that affected
a very large number of expression traits (Blows et al. 2015). Here, the analysis of $M_{26}$ was
also consistent with the existence of large variational modules.
How biologically specific a functional module is and whether its function is essential or localized might also help explain some of our observations. A first step towards accounting for the different roles of our 13 functional groups is to consider the structure of GO ontology. Analyses of GO terms as directed acyclic graphs (DAGs) (or gene or protein interaction networks more generally) have identified common topological features, and highlighted the functional or evolutionary significance of certain types of nodes (genes or gene products) or edges (interactions) (reviewed in Hu et al. 2016; Zhang et al. 2016). The 13 functional groups (individual GO terms) considered here varied considerably in the topology of their networks (DAGs), including in their degree (number of other GO terms to which they relate, Table 1). Under the hypothesis that selection on function drives evolution of pleiotropy to match functional groupings, GO terms embedded within broader functional networks (i.e., with high degree) might be expected to have reduced within-term covariance compared to GO terms with more specific functions (low degree). However, the estimated Spearman’s correlation between degree and the proportion of variance associated with each of the first two axes of mutational or standing genetic variance was mainly not in the predicted direction ($m_{\text{max}}$: $\rho = -0.20$; $m_2$: $\rho = 0.04$; $g_{\text{max}}$: $\rho = 0.09$; $g_2$: $\rho = 0.02$). These correlations were not statistically supported as distinct from zero ($m_{\text{max}}$: $p = 0.50$; $m_2$: $p = 0.90$; $g_{\text{max}}$: $p = 0.78$; $g_2$: $p = 0.94$), but it should be noted that we had very low power given a sample size of 13. Thus, whether or not the studied GO term belonged to a broader functional module might help explain our results, but analysis of more functional modules with a range of degrees will be needed to make a robust test of this. It is important to remember that the effects of the genes identified in our functional groups are not restricted to this group.
Indirect evidence for the embedding of a number of functional groups within larger variational modules comes from several published studies. In humans, Pickrell et al. (2016) combined GWAS studies on 42 traits or diseases and identified 341 loci that associated with multiple functionally unrelated traits, highlighting that pleiotropy spanned multiple functions. When considering clusters of co-expressed genes, Allocco et al (2004) found that in yeast, pairs of co-regulated genes belong to significantly closer GO terms than randomly chosen pairs of genes. However, many co-expressed and co-regulated genes belonged to very distant GO terms. Thus, although clusters of co-expression between unstudied genes and genes of known biological function has proven a powerful tool for the identification of putative function of unknown genes and of candidate genes for biological traits or functions of interest (e.g. Luo et al. 2007; Nayak et al. 2009; Ayyer et al. 2011; Proost and Mutwil 2017), there is nonetheless considerable evidence that pleiotropic variation is not strongly restricted to occur only within functional groups.

Selection on pleiotropic mutations

One of the key assumptions of theoretical predictions that variational modules should evolve to correspond to functional modules is that pleiotropic effects across functions will be associated with greater loss of fitness than pleiotropic effects within functions. Evidence for this has been equivocal, with some studies reporting that more highly connected genes, putatively highly pleiotropic genes functioning as hubs between different functional groups, are under stronger selection than genes with more functionally limited interactions (Jeong et al. 2001; Fraser et al. 2002; Krylov et al. 2003; Carter et al. 2004; Han et al. 2004; Butland et al. 2005; Zhao et al. 2007; Lin et al. 2015; Pirtykin et al. 2015), while other studies have
failed to detect any evidence that the strength of selection scales positively with the extent of pleiotropy, essentiality or network position (Pal and Hurst 2003; Hahn et al. 2004; Salathé et al. 2006; Jovelin and Phillips 2009; Kopp and McIntyre 2012). Here, we also obtain mixed support for this prediction. We observed the predicated significantly weaker selection within functional groups than among random sets of genes for only three of the 13 functional groups. Again, variation among functional groups in their network topology could account for variation in the strength of selection acting on functional groups, with biologically more specific groups expected to more closely match this prediction. The relationship between degree and selection was in the predicted direction for two of our indices of selection ($s_m^2$: $\rho = 0.38$, $p = 0.20$; Krzanowski $S$: $\rho = -0.31$, $p = 0.30$), but not for the other ($s_{max}^2$: $\rho = -0.15$, $p = 0.62$) and in no case were we able to support a significant relationship between selection and degree. Surprisingly, our results even suggested that selection against pleiotropic mutations affecting many functions may be lower than the selection observed within functional modules. The greater similarity of the mutational and genetic subspaces that represented the among functional group covariance, compared to within functional group covariance, is inconsistent with the evolution of modularity in response to deleterious pleiotropic mutation.

Collectively, our results confirm that while focusing on individual GO terms as functional modules can to some extent predict variational modularity, it will miss many important biological connections among functional modules. Considering higher-order interactions generated by broader variational modules spanning various functions will be a necessary part of understanding the evolution of genetic covariance.
ACKNOWLEDGEMENTS

We are grateful to John Stinchcombe for helpful discussions. We thank J. Thompson, C. Oesch-Lawson, D. Petfield, O. Sissa-Zubiate, F. Frentiu and Y.H. Ye for help with the MA experiment and data collection. We also thank Gabriel Marroig and an anonymous reviewer for their very helpful comments on the manuscript. This work was funded by the Australian Research Council and an Agreenskills + fellowship to J.M.C.
REFERENCES


JOVELIN, R., and P. C. PHILLIPS, 2009 Evolutionary rates and centrality in the yeast gene regulatory

network. Genome Biology 10.

KANEHISA, M., S. GOTO, M. HATTORI, K. F. AOKI-KINOSHITA, M. ITOH et al., 2006 From genomics to

chemical genomics: new developments in KEGG. Nucleic Acids Research 34: D354-D357.


KOPP, A., and L. M. McINTYRE, 2012 Transcriptional Network Structure Has Little Effect on the Rate of


KRYLOV, D. M., Y. I. WOLF, I. B. ROGOZIN and E. V. KOONIN, 2003 Gene Loss, Protein Sequence

Divergence, Gene Dispensability, Expression Level, and Interactivity Are Correlated in


KRZANOWSKI, W. J., 1979 Between-groups comparison of principal components. Journal of the


LANGFELDER, P., and S. HORVATH, 2007 Eigengene networks for studying the relationships between co-


expression levels are coupled in Drosophila and are independently associated with mRNA

abundance, protein length, and number of protein-protein interactions. Molecular Biology


LIN, C. Y., T. L. LEE, Y. Y. CHIU, Y. W. LIN, Y. S. LO et al., 2015 Module organization and variance in


LUO, F., Y. YANG, J. ZHONG, H. GAO, L. KHAN et al., 2007 Constructing gene co-expression networks and

predicting functions of unknown genes by random matrix theory. Bmc Bioinformatics 8.


**Figure 1:** Function specific genetic variance in two groups of similar sizes: Ion Tsp (panels A, B, and C) and Cuticle (panels D, E, and F). A and D: Proportion of the mutational variance contained in $m_{max}$ and $m_2$ against the distribution of the proportion of mutational variance in the non-function specific biological background ($M_{bb}$, Table 2). $m_{max}$ and $m_2$ are represented by black and grey vertical lines respectively. The density functions of the unspecific biological variance are represented for $bbm_{max}$ (black) and $bbm_2$ (grey) with a Gaussian smoothing kernel. The shaded area corresponds to the highest 2.5% of the distribution. B and E: Illustration of how many strong links among pair-wise traits there are within a functional group. Each circle represents a gene, the black lines represent mutational correlations between pairs of genes over 0.99 (this cut-off was chosen for the clarity of the illustration and has no role in our analyses or interpretation). C and F: Plot as for B and F, except graphing genetic rather than mutational variances.

**Figure 2:** Spectral distribution of the mutational variance contained in $M_{26}$ (A) and $G_{26}$ (B) (See Table 2). The dots are the variances associated with each eigenvector in a descending order. The grey shaded area represents the 95% CI of the 50 simulations estimating sampling error (se).

**Figure 3:** Contribution of the genes contained in the 13 GO terms functional groups to the 19 heritable phenotypic common factors identified in a Bayesian sparse factor (BSF) analysis (HINE et al. submitted). Each line represents a functional module (see Table 1) and each column a BSF. The number of genes within each functional module that also had a significant contribution to the BSF (5% fdr) is listed, with shading from light to dark indicating few to many genes.
**Table 1**: Description of the 13 selected functional modules. MF and BP describes whether the GO term is a Molecular Function or a Biological Process. Background is the count of all 8978 genes associated with Flybase Gene IDs recognised by DAVID that are associated with the corresponding GO term. Sample (FBgn) is the number of genes that are associated with the GO term and have H_m^2 > 0 when considering *D. melanogaster* FlyBase Gene IDs. Sample (genes) is the corresponding number of genes on the microarray; when multiple genes within the a functional module correspond to the same *D. melanogaster* homolog, the total number of genes in the functional module, listed under Sample (genes), is larger than the number of Flybase Gen IDs, listed under Sample (FBgn). Enrichment (p value) is the P value of enrichment for this term. Degree is the number of other GO terms to which a GO term relates, according to the GO term topology as revealed in Directed Acyclic Graphs. Overlap is the number of genes that can also be found in at least one other functional module. Genes with significant H_m^2 show the number of genes in the group with significant H_m^2 (P value). Median H_m^2 and H_g^2 show the median mutational and broad sense heritability respectively for the sampled genes of the group; median H_m^2 is calculated only for the genes that had non-zero mutational variance. Level of significance (* <0.05, ** <0.01, *** <0.001) given by the median p value of 1000 Mann-Whitney tests comparing the univariate variances of the genes of the groups to the same number genes, randomly chosen among that were not included in any functional module.

<table>
<thead>
<tr>
<th>Group</th>
<th>GO ID</th>
<th>Ontology</th>
<th>Term</th>
<th>Background Sample (FBgn)</th>
<th>Sample (genes)</th>
<th>Enrichment (p value)</th>
<th>Degree</th>
<th>Overlap</th>
<th>Genes with significant H_m^2</th>
<th>Median H_m^2</th>
<th>Median H_g^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chorion</td>
<td>GO:0005213</td>
<td>MF</td>
<td>structural constituent of chorion</td>
<td>8</td>
<td>8</td>
<td>0.0011</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0.0016</td>
<td>0.0000 *</td>
</tr>
<tr>
<td>Amino A</td>
<td>GO:0000096</td>
<td>BP</td>
<td>sulfur amino acid metabolic process</td>
<td>14</td>
<td>9</td>
<td>0.0222</td>
<td>12</td>
<td>0</td>
<td>4</td>
<td>0.0014</td>
<td>0.4453</td>
</tr>
<tr>
<td>NeuroT</td>
<td>GO:0030594</td>
<td>MF</td>
<td>neurotransmitter receptor activity</td>
<td>29</td>
<td>15</td>
<td>0.0245</td>
<td>10</td>
<td>0</td>
<td>5</td>
<td>0.0016</td>
<td>0.3840</td>
</tr>
<tr>
<td>GT</td>
<td>GO:0015020</td>
<td>MF</td>
<td>glucuronosyltransferase activity</td>
<td>25</td>
<td>15</td>
<td>0.0049</td>
<td>17</td>
<td>0</td>
<td>10</td>
<td>0.0018</td>
<td>0.6255 *</td>
</tr>
<tr>
<td>Pro-DNA</td>
<td>GO:0065004</td>
<td>BP</td>
<td>protein-DNA complex assembly</td>
<td>35</td>
<td>18</td>
<td>0.0086</td>
<td>15</td>
<td>0</td>
<td>2</td>
<td>0.0009</td>
<td>0.3198</td>
</tr>
<tr>
<td>Bacterium</td>
<td>GO:0009617</td>
<td>BP</td>
<td>response to bacterium</td>
<td>51</td>
<td>25</td>
<td>0.0033</td>
<td>7</td>
<td>4</td>
<td>15</td>
<td>0.0025</td>
<td>0.4621</td>
</tr>
<tr>
<td>Chitin</td>
<td>GO:0008061</td>
<td>MF</td>
<td>chitin binding</td>
<td>47</td>
<td>28</td>
<td>0.0001</td>
<td>2</td>
<td>0</td>
<td>10</td>
<td>0.0013</td>
<td>0.4438</td>
</tr>
<tr>
<td>Sensory</td>
<td>GO:0007606</td>
<td>BP</td>
<td>sensory perception of chemical stimulus</td>
<td>62</td>
<td>32</td>
<td>0.0002</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>0.0013</td>
<td>0.5777</td>
</tr>
<tr>
<td>Ion Tsp</td>
<td>GO:0030001</td>
<td>BP</td>
<td>metal ion transport</td>
<td>91</td>
<td>35</td>
<td>0.0373</td>
<td>12</td>
<td>2</td>
<td>10</td>
<td>0.0007</td>
<td>0.3364</td>
</tr>
<tr>
<td>Heme</td>
<td>GO:0020037</td>
<td>MF</td>
<td>heme binding</td>
<td>99</td>
<td>43</td>
<td>0.0037</td>
<td>2</td>
<td>1</td>
<td>24</td>
<td>0.0017</td>
<td>0.5043 *</td>
</tr>
<tr>
<td>Cuticle</td>
<td>GO:0042302</td>
<td>MF</td>
<td>structural constituent of cuticle</td>
<td>59</td>
<td>43</td>
<td>&lt;0.0001</td>
<td>3</td>
<td>0</td>
<td>21</td>
<td>0.0014</td>
<td>0.5022</td>
</tr>
<tr>
<td>Cell Fate</td>
<td>GO:0045165</td>
<td>BP</td>
<td>cell fate commitment</td>
<td>133</td>
<td>59</td>
<td>0.0001</td>
<td>44</td>
<td>2</td>
<td>11</td>
<td>0.0009</td>
<td>0.2568 *</td>
</tr>
<tr>
<td>Endopep</td>
<td>GO:0004252</td>
<td>MF</td>
<td>serine-type endopeptidase activity</td>
<td>143</td>
<td>62</td>
<td>0.0004</td>
<td>7</td>
<td>3</td>
<td>31</td>
<td>0.0013</td>
<td>0.6143 ***</td>
</tr>
</tbody>
</table>

Genes included in functional groups: 386, 434

Biological background – genes with non-zero mutational variance that were not assigned to any of the 13 functional group: 2397, 2951
Table 2: Definitions of the used mathematical terms.

<table>
<thead>
<tr>
<th>Name</th>
<th>Quantifies</th>
<th>Number of datasets/ permutations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>m&lt;sub&gt;max&lt;/sub&gt; or g&lt;sub&gt;max&lt;/sub&gt;</td>
<td>Eigenvectors of most (max) and second most (2) mutational (m) and genetic (g) variance</td>
<td>Estimated by diagonalising the respective M or G matrix</td>
<td>Phenotypic scores were estimated for individuals for the trait combinations described by the first and second major axes of within function mutational or standing genetic variance for each of the 13 functional groups (i.e., 26 traits). These scores were then analysed as per observed traits to estimate the 26 trait variance covariance matrix, which was then diagonalised.</td>
</tr>
<tr>
<td>M&lt;sub&gt;26&lt;/sub&gt; or G&lt;sub&gt;26&lt;/sub&gt;</td>
<td>Mutational (M) or standing genetic (G) covariance among functional groups</td>
<td>50</td>
<td>Data shuffled among lines, disrupting the mutational or genetic covariances between traits while retaining the observed levels of variance for each individual gene expression trait</td>
</tr>
<tr>
<td>m&lt;sub&gt;2&lt;/sub&gt; or g&lt;sub&gt;2&lt;/sub&gt;</td>
<td>Estimated by diagonalising the respective M or G matrix</td>
<td>143</td>
<td>Uses random sets of genes taken from the list of genes that were not assigned to one of the 14 functional groups. Each of the 117 datasets was then subsampled to the same number of genes as in the respective functional group.</td>
</tr>
<tr>
<td>M&lt;sub&gt;26&lt;/sub&gt;se or G&lt;sub&gt;26&lt;/sub&gt;se</td>
<td>Sampling error within the among functional group mutational (M&lt;sub&gt;26&lt;/sub&gt;) or standing genetic (G&lt;sub&gt;26&lt;/sub&gt;) covariance matrices</td>
<td>50</td>
<td>Data shuffled among lines, disrupting the mutational or genetic covariances between traits while retaining the observed levels of variance for each composite trait present in M&lt;sub&gt;26&lt;/sub&gt; and G&lt;sub&gt;26&lt;/sub&gt;</td>
</tr>
</tbody>
</table>
Table 3: Proportion of variance of the first and second eigenvectors of \( \mathbf{M} \) and \( \mathbf{G} \) and measures of selection within functional groups (note that because there are some negative eigenvalues, the denominator used in these proportions, namely the trace, can be lower than the amount of variance found in the first two axes of \( \mathbf{M} \) or \( \mathbf{G} \), resulting in sums of proportions greater than 1 in group Chorion, Amino A, and Pro-DNA). Within each functional group, Krzanowski \( \mathbf{S} \) compares the 2- dimensional subspaces defined by \( \mathbf{m}_{\text{max}} \) and \( \mathbf{m}_2 \) in the space of \( \mathbf{M} \) and \( \mathbf{G} \). Confidence intervals correspond to the 2.5\textsuperscript{th} and 97.5\textsuperscript{th} percentile eigenvalues, as a proportion, observed for the 123 \( \mathbf{M}_{gb} \) and \( \mathbf{G}_{gb} \). Significance levels are * > 95%, ** > 99%, and *** outside confidence intervals. CI for two sets of two groups (Chitin and Sensory, and Heme and Cuticle) were the same as groups contain the same number of genes.

<table>
<thead>
<tr>
<th>Group</th>
<th>( \lambda m_{\text{max}} ) [CI]</th>
<th>( \lambda m_2 ) [CI]</th>
<th>( \lambda g_{\text{max}} ) [CI]</th>
<th>( \lambda g_2 ) [CI]</th>
<th>Krzanowski ( \mathbf{S} ) [CI]</th>
<th>( s_{m_{\text{max}}} ) [CI]</th>
<th>( s_{m_2} ) [CI]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chorion</td>
<td>0.67 [0.42; 0.88]</td>
<td>0.37 [0.18; 0.45]</td>
<td>0.69*** [0.28; 0.56]</td>
<td>0.50*** [0.18; 0.35]</td>
<td>0.91 [0.17; 1.31]</td>
<td>1.306*** [0.002; 0.177]</td>
<td>0.012 [-0.010; 0.070]</td>
</tr>
<tr>
<td>Amino A</td>
<td>0.71 [0.41; 0.79]</td>
<td>0.30 [0.17; 0.48]</td>
<td>0.42 [0.26; 0.55]</td>
<td>0.25 [0.18; 0.33]</td>
<td>0.51 [0.14; 1.18]</td>
<td>0.011 [0.005; 0.369]</td>
<td>0.010 [-0.006; 0.089]</td>
</tr>
<tr>
<td>NeuroT</td>
<td>0.58 [0.35; 0.75]</td>
<td>0.17*** [0.18; 0.43]</td>
<td>0.30 [0.22; 0.43]</td>
<td>0.22 [0.17; 0.28]</td>
<td>0.32 [0.08; 0.94]</td>
<td>0.022 [0.005; 0.176]</td>
<td>0.012 [-0.007; 0.195]</td>
</tr>
<tr>
<td>GT</td>
<td>0.44 [0.35; 0.68]</td>
<td>0.39 [0.18; 0.43]</td>
<td>0.25 [0.22; 0.41]</td>
<td>0.20 [0.16; 0.27]</td>
<td>0.57 [0.08; 0.87]</td>
<td>0.028 [0.005; 0.180]</td>
<td>0.022 [-0.009; 0.075]</td>
</tr>
<tr>
<td>Pro-DNA</td>
<td>0.66 [0.34; 0.73]</td>
<td>0.42* [0.16; 0.40]</td>
<td>0.34 [0.21; 0.38]</td>
<td>0.16 [0.15; 0.26]</td>
<td>0.15 [0.07; 0.78]</td>
<td>0.033 [0.005; 0.139]</td>
<td>0.031 [-0.010; 0.096]</td>
</tr>
<tr>
<td>Bacterium</td>
<td>0.63 [0.31; 0.63]</td>
<td>0.23 [0.18; 0.37]</td>
<td>0.30 [0.18; 0.32]</td>
<td>0.22 [0.14; 0.22]</td>
<td>0.84*** [0.05; 0.61]</td>
<td>0.042 [0.008; 0.279]</td>
<td>0.013 [-0.012; 0.128]</td>
</tr>
<tr>
<td>Chitin</td>
<td>0.56 [0.32; 0.60]</td>
<td>0.33 [0.17; 0.36]</td>
<td>0.22 [0.18; 0.32]</td>
<td>0.17 [0.14; 0.22]</td>
<td>0.27 [0.04; 0.59]</td>
<td>0.015 [0.009; 0.218]</td>
<td>0.023 [-0.003; 0.126]</td>
</tr>
<tr>
<td>Sensory</td>
<td>0.49 [0.32; 0.60]</td>
<td>0.23 [0.17; 0.36]</td>
<td>0.21 [0.18; 0.32]</td>
<td>0.19 [0.14; 0.22]</td>
<td>0.14 [0.04; 0.59]</td>
<td>0.024 [0.009; 0.218]</td>
<td>0.019 [-0.003; 0.126]</td>
</tr>
<tr>
<td>Ion Tsp</td>
<td>0.47 [0.30; 0.60]</td>
<td>0.28 [0.17; 0.36]</td>
<td>0.23 [0.17; 0.29]</td>
<td>0.18 [0.13; 0.20]</td>
<td>0.31 [0.05; 0.55]</td>
<td>0.062 [0.012; 0.417]</td>
<td>0.025 [-0.019; 0.130]</td>
</tr>
<tr>
<td>Heme</td>
<td>0.46 [0.31; 0.57]</td>
<td>0.40** [0.18; 0.33]</td>
<td>0.17 [0.15; 0.27]</td>
<td>0.13 [0.13; 0.18]</td>
<td>0.50* [0.04; 0.47]</td>
<td>0.051 [0.016; 0.443]</td>
<td>0.022 [-0.019; 0.185]</td>
</tr>
<tr>
<td>Cuticle</td>
<td>0.63** [0.31; 0.57]</td>
<td>0.22 [0.18; 0.33]</td>
<td>0.23 [0.15; 0.27]</td>
<td>0.14 [0.13; 0.18]</td>
<td>0.61*** [0.04; 0.47]</td>
<td>0.014* [0.016; 0.443]</td>
<td>0.021 [-0.019; 0.185]</td>
</tr>
<tr>
<td>Cell Fate</td>
<td>0.46 [0.30; 0.52]</td>
<td>0.27 [0.16; 0.32]</td>
<td>0.19 [0.15; 0.26]</td>
<td>0.15 [0.12; 0.18]</td>
<td>0.25 [0.04; 0.41]</td>
<td>0.020 [0.018; 0.635]</td>
<td>0.037 [-0.017; 0.238]</td>
</tr>
<tr>
<td>Endopep</td>
<td>0.31 [0.31; 0.52]</td>
<td>0.22 [0.17; 0.32]</td>
<td>0.16 [0.15; 0.25]</td>
<td>0.13 [0.11; 0.18]</td>
<td>0.31 [0.04; 0.38]</td>
<td>0.016 [0.012; 0.591]</td>
<td>0.044 [-0.037; 0.111]</td>
</tr>
<tr>
<td>Phenotypic common factor</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Chorion</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino A</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NeuroT</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GT</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro-DNA</td>
<td></td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Bacterium</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chitin</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensory</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Ion Tsp</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Heme</td>
<td>6</td>
<td>6</td>
<td>13</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Cuticle</td>
<td>2</td>
<td>19</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell Fate</td>
<td>5</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endopep</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Functional groups: