HIGH-DIMENSIONAL VARIANCE PARTITIONING REVEALS THE MODULAR GENETIC BASIS OF ADAPTIVE DIVERGENCE IN GENE EXPRESSION DURING REPRODUCTIVE CHARACTER DISPLACEMENT

Elizabeth A. McGraw,[1] Yixin H. Ye,[1] Brad Foley,[2] Stephen F. Chenoweth,[1] Megan Higgie,[3] Emma Hine,[1] and Mark W. Blows[1,4]

[1]School of Biological Sciences, University of Queensland, Brisbane, QLD 4072, Australia

[2]Molecular and Computational Biology, University of Southern California, Los Angeles, CA 90089

[3]Division of Evolution, Ecology and Genetics, Research School of Biology, The Australian National University, Canberra, ACT 0200, Australia

[4]Email: m.blows@uq.edu.au

Although adaptive change is usually associated with complex changes in phenotype, few genetic investigations have been conducted on adaptations that involve sets of high-dimensional traits. Microarrays have supplied high-dimensional descriptions of gene expression, and phenotypic change resulting from adaptation often results in large-scale changes in gene expression. We demonstrate how genetic analysis of large-scale changes in gene expression generated during adaptation can be accomplished by determining high-dimensional variance partitioning within classical genetic experimental designs. A microarray experiment conducted on a panel of recombinant inbred lines (RILs) generated from two populations of *Drosophila serrata* that have diverged in response to natural selection, revealed genetic divergence in 10.6% of 3762 gene products examined. Over 97% of the genetic divergence in transcript abundance was explained by only 12 genetic modules. The two most important modules, explaining 50% of the genetic variance in transcript abundance, were genetically correlated with the morphological traits that are known to be under selection. The expression of three candidate genes from these two important genetic modules was assessed in an independent experiment using qRT-PCR on 430 individuals from the panel of RILs, and confirmed the genetic association between transcript abundance and morphological traits under selection.

KEY WORDS: gene expression, G. matrix, adaptation, reinforcement.

Understanding the genetic basis of adaptation remains a key goal of evolutionary genetics (Orr 2005; Hoekstra and Coyne 2007). Although adaptive change is usually associated with complex changes in morphological phenotype (Blows 2007), few genetic investigations have been conducted on adaptations that involve sets of high-dimensional traits (Albert et al. 2008). This is a par-

ticularly important limitation of evolutionary studies, as a full understanding of the evolution of a focal trait is unlikely to be gained in the absence of knowledge on how it interacts with the wider phenome that ultimately is comprised of a very large number of traits (Houle 2010). Pleiotropic genetic associations among multiple traits can cause focal traits to respond to selection in the

direction opposite to that favored by selection (Walsh and Blows 2009), or to stop traits from evolving in the presence of genetic variation and ongoing selection (Hine et al. 2011). The genetic independence of a focal trait from other traits, or the genetic independence among different sets of traits—which is often referred to as modularity (Cheverud 1996; Hansen 2003), are important determining factors of if, and how, adaptation will occur in a particular circumstance.

The investigation of pleiotropic associations among phenotypes is relatively uncommon in the application of high-throughput genomic technologies that generate vast amounts of data. Although marker-based QTL mapping approaches have been successful in identifying discrete regions of the genome that underlie divergence between populations in individual traits, they often do not directly consider pleiotropic relationships among multiple traits (Xu et al. 2005; Biswas et al. 2008), as distinct from mapping single traits and searching for nonoverlap of confidence regions to reject the hypothesis of pleiotropy. Similarly, microarrays have supplied high-dimensional descriptions of transcript abundance, which have traditionally been analyzed by the phenotypic identification of co-expressed networks of gene transcripts based on various clustering approaches. However, the genetic analysis of these high-dimensional expression phenotypes, as distinct from the phenotypic clustering of co-expressed transcripts, is more problematic (Kadarmideen et al. 2006), and has tended to concentrate on individual transcripts, rather than the genetic control of sets of co-expressed transcripts (Biswas et al. 2008). What has been lacking are statistical approaches that allow the multivariate analysis of the abundance of a large number of transcripts measured from classical genetic experimental designs to determine the extent of shared genetic control of gene expression.

The importance of addressing how pleiotropic effects of genes influence the evolution of high-dimensional expression phenotypes has been highlighted by studies that have found a very large number of expression profiles that differ between the sexes (Ranz et al. 2003; Gibson et al. 2004), developmental stages (White et al. 1999), and between strains and populations adapted to different environments (Franchini and Egli 2006; Ronald and Akey 2007; Lai et al. 2008; St-Cyr et al. 2008). It is difficult to reconcile changes of expression in such a large number of transcripts with the relatively modest number of QTL that are often found to underlie adaptations in single traits, even after taking into account the underpowered nature of QTL mapping. The large number of transcripts exhibiting a change in expression is therefore likely consequence of pleiotropic regulation of expression or physical linkage to some unknown extent (Chesler et al. 2005; Kadarmideen 2006; Biswas et al. 2008; Gilad et al. 2008; Litvin et al. 2009; Skelly et al. 2009). A recent genetic analysis of transcript abundance within a population of *Drosophila*

melanogaster indicated that the large number of differences in transcript abundance among genotypes were likely to be a consequence of a smaller number of modules of pleiotropically related expression phenotypes (Ayroles et al. 2009). This indicates that although phenotypic change resulting from adaptation may result in large-scale changes in gene expression, such changes may be accomplished through a modest number of regulatory genes that influence these pleiotropic networks.

The development of genetic analyses for high-dimensional phenotypes, particularly in the extreme case of large numbers of transcript abundances in systems genetics, has lagged behind our ability to generate these large datasets. High-dimensional genetic analysis of transcript abundances can be approached in at least two ways. First, transcript abundance phenotypes can be subjected to an ordination procedure to generate "eigentraits," linear combinations of the large number of expression traits that covary together (Biswas et al. 2008). This allows coregulation of phenotypes to be inferred, and the discovery of eQTL that are associated with large-scale regulatory changes when suitable markers have also been obtained. The extent to which the multivariate analysis of expression phenotypes in this manner will reflect the underlying genetic patterns of coregulation will depend on the contribution of environmental covariance among the phenotypes in question. For example, if the magnitude of an environmental correlation is much stronger (or weaker) than the genetic correlation between two transcripts, or if two correlations are of opposite signs in the more extreme case, a misleading picture of the genetic coregulation of transcript abundance will be given by the multivariate analysis of phenotypes. In contrast, multivariate genetic analysis of standard metric traits (Mezey and Houle 2005; Hines and Blows 2006; Meyer and Kirkpatrick 2008) explicitly removes the confounding influence of environmental covariance to directly model the multivariate genetic relationships among traits.

A second approach is to explicitly consider the partitioning of environmental and genetic covariance among expression phenotypes to remove the confounding influence of the environment on transcript abundance. This is a challenging task as the high-dimensional analysis needs to incorporate an experimental design that is more complex than measures of multiple phenotypes of individuals. In a recent example of such an approach, the genetic co-regulation of gene expression among inbred lines of *Drosophila* was inferred from patterns among bivariate genetic correlations that had been estimated by partitioning out the influence of environmental covariance among transcripts. These genetic correlations were arranged in a distance matrix, from which genetic modules of coexpressed genes were identified using clustering (Ayroles et al. 2009; Stone and Ayroles 2009).

Although the removal of the confounding influence of environmental covariance among transcripts in this way is a major advantage over the multivariate analysis of transcript

abundance phenotypes, two issues remain to be addressed before high-dimensional genetic analysis of transcript abundance can be implemented in a framework that shares all the advantages of standard multivariate genetic analysis. First, clustering is explicitly exploratory, lacking a hypothesis-testing framework that can be readily adapted to experimental designs with hierarchical levels, that are often required to partition phenotypic variation into genetic and environmental sources. Second, conversion of the data to a network comprised of vertices and edges based on a distance matrix of absolute pairwise genetic correlations was used to approximate the genetic covariance structure among multiple traits (Ayroles et al. 2009; Stone and Ayroles 2009), in place of the true genetic variance–covariance (**G**) matrix, the multivariate extension of bivariate genetic correlations that is modeled in standard multivariate quantitative genetics (Mezey and Houle 2005; Hines and Blows 2006; Meyer and Kirkpatrick 2008). Genetic information from the sign of bivariate genetic correlations, and hence the exact nature of how transcripts are coregulated, was lost as a consequence of this transformation. This precludes a formal determination of the genetic independence of expression across multiple transcripts, based on the true eigenstructure of the multivariate genetic relationships among transcripts. Ideally, the **G** matrix among a large number of transcript abundances needs to be directly estimated within an established hypothesis-testing framework so that the multivariate genetic relationships among transcripts can be fully characterized.

Using a well-characterized example of adaptation, we demonstrate how high-dimensional genetic analysis of gene expression can be accomplished by determining the modularity of the effect space of the among-genotype (genetic) variance in a multivariate linear model (Hine and Blows 2006). Reproductive character displacement is an adaptation that occurs when individuals from two different species that coexist encounter each other during mate choice, and suffer a fitness cost if they perceive an individual from the other species as a potential mate (Brown and Wilson 1956; Howard 1993). In this situation, reinforcing selection acts on the traits that are used by individuals in mate choice so that they evolve to avoid making such mistakes. Species of the *Drosophila serrata* complex use contact pheromones, comprised of cuticular hydrocarbons (CHCs), to identify potential mates. The CHCs of male *D. serrata* are under strong sexual selection as a consequence of female choice (Blows et al. 2004; Higgie and Blows 2008) and display reproductive character displacement in field populations where the closely related *D. birchii* is sympatric with *D. serrata* (Higgie et al. 2000; Higgie and Blows 2007). The reproductive character displacement evolves in experimental sympatry under laboratory conditions (Higgie et al. 2000), demonstrating that reinforcing selection is responsible for the divergence in CHCs among sympatric and allopatric *D. serrata* populations. Enzymes that are involved in the production of *Drosophila* CHCs

have been shown to have very high rates of evolution in gene expression between the sexes (Shirangi et al. 2009), suggesting that gene regulation may play a major role in the response to selection of these traits.

We present the results from a series of three genetic analyses. First, using a panel of recombinant inbred lines (RILs) generated from two populations of *D. serrata* that have diverged in response to reinforcing selection, we determined that the evolutionary response to selection was associated with changes in expression of a large number of gene transcripts, but that these changes were explained by a much smaller number of genetically independent changes in regulation. Second, we show that the two major genetic modules identified by the high-dimensional genetic analysis were genetically correlated with the morphological traits under reinforcing selection, suggesting that changes in transcript expression underlie the adaptive changes in morphology. Finally, we used Quantitative Reverse Transcription PCR (qRT-PCR) to provide independent experimental validation of the genetic association between transcript abundance and CHC phenotypes. The expression of three candidate genes, identified as playing a major role in the two important genetic modules by the multivariate genetic analysis, was shown to be genetically correlated with CHC expression.

Methods
RIL CONSTRUCTION

Eungella and Forster are two geographic locations along the east coast of Australia; the former in a sympatric region in which *D. birchii* is present, while *D. birchii* is not present at the latter allopatric region (Higgie et al. 2000; Higgie and Blows 2007). Two lines founded by a single inseminated female from mass-bred populations sourced from Eungella and Forster were made cytologically standard (inversion free), and inbred for 10 generations of full-sibling mating. A single male and female from both parental lines were used in reciprocal crosses to balance the maternal and paternal contributions from the two populations in the panel of RILs. F_2 full-sibling pairs were used to establish 101 RILs that were inbred by full-sibling mating for 17 generations, effectively reducing initial heterozygosity by >90%. From each of the two parental lines and 41 randomly selected RILs, 10 individual male flies were phenotyped for CHC profile and saved for subsequent qRT-PCR analysis after the microarray experiment.

For the microarray experiment, 15 of the 101 RILs were randomly selected for transcriptional profiling (details below). After several candidate genes had been identified by the microarrays, their expression was then examined using qRT-PCR analysis in the original two parental lines and in the 41 randomly selected RILs. These 41 RILs included eight of the RILs that were represented in the microarray experiment (Fig. 1).
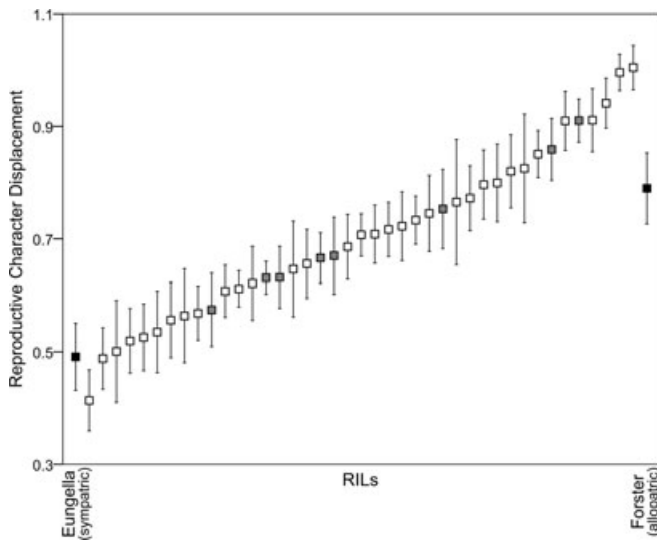
Figure 1. Reproductive character displacement of male cuticular hydrocarbons (CHCs) in sympatric and allopatric parental lines and RILs of *D. serrata*. We used the control population data from Higgie et al. (2000) that consisted of the CHC phenotypes from 20 males from each of three sympatric and three allopatric populations, to create a single, univariate trait of reproductive character displacement. We applied a multivariate hierarchical linear model to these data, with replicate geographic population nested within sympatry or allopatry (using Proc Mixed in SAS). The canonical variate at the sympatry/allopatry level from this model represented the linear combination of CHCs that differed most between sympatry and allopatry. The equation for the canonical variate was then applied to the male CHC data from the two parental lines and 41 RILs. Boxes represent line means and bars are 95% confidence intervals. Lines have been arranged along the *x*-axis to place sympatric (Eungella) and allopatric (Forster) parental lines at either end, with the RILs in an ascending order of *y*-axis value. The parental lines are significantly different from one another. The RILs that were also assayed in the subsequent microarray experiment are shown with gray boxes.

CHC PHENOTYPING

Ten males from each of the 41 RILs and two parental lines were collected when males were 6-day old post eclosion for a total sample size of 430. To collect the CHC samples, flies were individually immersed in 120 μL of hexane for 3 min before being vortexed for 1 min and then removed from the hexane. Immediately, the same individual flies were then placed in Trizol (Invitrogen, Carlsbad, CA) and prepared for qRT-PCR as below. The CHC samples were kept at $-20°C$ prior to analysis on an Agilent 7890A gas chromatograph that was fitted with an Agilent HP5 column of 30-m length, 250-μm diameter, and 0.10-μm film thickness. Using an Agilent 7693A autosampler, 1 μL of each sample was pressure-pulse injected into a 200°C splitless inlet. The hydrogen carrier gas flow started at 2.5 mL/min held for 3.7 min, and then ramped at 5 mL/min to a final flow of 5 mL/min. The oven tem-

perature started at 140°C held for 0.55 min, ramped at 100°C/min to 190°C, then ramped at 45°C/min to 320°C and was held for 1 min, for a total run time of 4.94 min. The flame ionization detector was set at 315°C.

The resulting 430 male CHC phenotypes were quantified in the same way as was done previously (Higgie et al. 2000; Higgie and Blows 2007, 2008). Briefly, for each male the relative areas of nine peaks were calculated using Agilent GC Chemstation version B.04.01. These peaks correspond to the compounds $5,9-C_{24}$, $5,9-C_{25}$, $9-C_{25}$, $9-C_{26}$, $2-Me-C_{26}$, $5,9-C_{27}$, $2-Me-C_{28}$, $5,9-C_{29}$, and $2-Me-C_{30}$ (Fig. 2 in Higgie and Blows 2007). To make the data suitable for multivariate statistical analyses, the nine relative peak areas were transformed into eight logcontrasts using $9-C_{26}$ as the divisor (for equation see Higgie and Blows 2007).

EXPRESSION PROFILING AND INITIAL DATA MANAGEMENT

Agilent oligonucleotide microarrays were designed using eArray Version 5.0 (Agilent Technologies, Inc., Santa Clara, CA) based on a *D. serrata* EST library (Frentiu et al. 2009). The $8 \times 15K$ format arrays contained the standard Agilent control set and three replicates of the 60 mer oligonucleotides representing each of the 3762 features. Four replicate pools of 20, 5-day-old males were collected from each of the 15 RILs for analysis (for a total of 60 hybridizations). RNA extraction, cDNA synthesis and labeling, hybridization and scanning procedures were as previously reported (Ye et al. 2009) with the exception that hybridizations were single color (Cy-3). Four replicates of each RIL were hybridized against eight chips in a partial block design, where chips (partial blocks) never received more one replicate from any particular RIL. The median signal intensity minus background fluorescence for each spot was log transformed and an average was calculated for each feature based on the three technical replicates per hybridization. All data have been deposited in ArrayExpress (http://www.ebi.ac.uk/microarray-as/ae/) under the accession no. A-MEXP-1759. *Drosophila melanogaster* orthologs of *D. serrata* ESTs were identified using FlyBase (www.flybase.org) and the associated annotations were used for functional analysis and tissue expression.

STATISTICAL ANALYSIS

The abundance of all 3762 transcripts was subjected to the univariate hierarchical mixed linear model:

$$y_{ijkl} = chip_i + RIL_j + rep_{k(j)} + \varepsilon_{l(ijk)},$$

where chip and RIL were treated as fixed effects, and replicate hybridizations (rep) within RILs was a random effect, implemented using the Mixed Procedure in SAS. Transcripts were then ranked according to the significance of the among-RIL effect. A Bonferroni correction for multiple comparisons gave a conservative

estimate of the number of transcripts that displayed significant among-RIL variation. Heritabilities for all transcripts were estimated from the same mixed model, but where RIL was changed to a random effect, and the proportion of total variance that was present at the RIL level was calculated.

We determined the number of statistically independent dimensions required to explain the among-RIL variation in transcriptional abundance in two steps using theory for multivariate effect spaces of linear models (Amemiya 1985; Anderson and Amemiya 1991; Hine and Blows 2006) which we outline below. First, we subjected the 10 most significant transcripts to this analysis to determine in a relatively low-dimensional situation, how many underlying genetically independent expression modules could be identified. Second, we repeated the analysis for the most significant 82 transcripts. We chose to use the first 82 genes in this analysis as detailed investigation showed that our approach began to lose power to detect significant dimensions if more transcripts were included as a consequence of the limited number of degrees of freedom available in the experiment (Fig. S1).

Hine and Blows (2006) presented a method for establishing which dimensions of the genetic variance–covariance matrix are statistically supported. Below, we summarize this method, which draws from three papers from the statistical literature. The first shows how a negative definite matrix can be partitioned into the sum of a nonnegative matrix and a negative definite matrix (Amemiya 1985). This work forms the basis for constructing the covariance matrix based on only those dimensions that receive statistical support. The second derives a test statistic, Y, to determine the dimensionality of a covariance matrix (Anderson and Amemiya 1991) based on the estimated quantiles of the distribution of Y, which is presented in the third paper (Amemiya et al. 1990).

The method presented by Amemiya (1985) manipulates mean square matrices obtained from a one-way multivariate ANOVA of p traits. The experimental design for our RIL experiment was also a one-way multivariate ANOVA, where variation among the means for each biological replicate within RILs represent the within-source error, and the variation among RILs is the between-source component. Hine and Blows (2006) show how the method can also be adapted for alternative experimental designs. In the one-way MANOVA, the estimator for the covariance matrix at the effect (RIL) level is

$$\mathbf{\Sigma}_{bb} = r^{-1}(\mathbf{m}_{bb} - \mathbf{m}_{ww}),$$

where $\mathbf{m}_{bb}$ and $\mathbf{m}_{ww}$ are the between group and within group $p \times p$ mean square matrices, respectively, and r is the coefficient of the variance components at the between group level. Amemiya (1985) notes that, in the univariate case, the variance estimate:

$$\sigma_{bb} = r^{-1}(m_{bb} - m_{ww})$$

will be in the parameter space (i.e., ≥ 0) if $m_{bb} - m_{ww} \geq 0$. He then introduces a value, $\hat{\lambda}$, such that

$$m_{bb} - \hat{\lambda}m_{ww} = 0. \tag{1}$$

If $\hat{\lambda} < 1$, $m_{bb} - m_{ww}$ is negative and the estimator for σ_{bb} is 0.

A multivariate version of (1) is

$$|\mathbf{m}_{bb} - \lambda\mathbf{m}_{ww}| = 0. \tag{2}$$

Solving (2) yields the vector $\mathbf{\lambda}$ of the p characteristic roots (λ_i) of $\mathbf{m}_{bb}$ in the metric of $\mathbf{m}_{ww}$. If all λ_i are ≥ 1, $(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ is nonnegative definite and the estimator for $\mathbf{\Sigma}_{bb}$ is in the parameter space. If some λ_i are < 1, Amemiya (1985) shows how to partition the matrix $(\mathbf{m}_{bb} - \mathbf{m}_{ww})$ into the sum of a nonnegative matrix, $(\mathbf{m}_{bb} - \mathbf{m}_{ww})_+$, and a negative definite matrix, $(\mathbf{m}_{bb} - \mathbf{m}_{ww})_-$, and derives a new, nonnegative definite estimator for $\mathbf{\Sigma}_{bb}$ based on this partition.

Hine and Blows (2006) extend this approach to construct a nonnegative definite genetic variance–covariance matrix based on only those dimensions of the effect space that are statistically supported, combining the partitioning of mean square matrices described above with Amemiya et al. (1991)'s work on determining the effective dimensionality of a covariance matrix. The effective dimensionality is determined through a nested series of hypothesis tests, starting with the null hypothesis that the effective number of dimensions (m) is less than or equal to the number (k) of $\lambda_i \geq 1$. This hypothesis can be accepted immediately, as there can be at most as many dimensions as there are $\lambda_i \geq 1$. The method iterates through null hypotheses that $m \leq k - 1$, $m \leq k - 2 \ldots m \leq 0$ until one of these null hypotheses is rejected. For example, if the null hypothesis that $m \leq k - 1$ is rejected, the effective dimensionality is k.

To test the null hypothesis that $m \leq b$, Amemiya et al. (1991) derive a test statistic,

$$Y = \sum_{i=b+1}^{k} f(\lambda_i),$$

where

$$f(\lambda) = -M \log \lambda + (M + N) \log \left(\frac{M\lambda + N}{M + N} \right).$$

Here, M and N are the degrees of freedom at the between-group and within-group levels, respectively. The distribution of Y for $q = p - m$ is presented in Table 1 of Amemiya et al. (1990).

Once the effective dimensionality has been established, it is then possible to construct the covariance matrix with only those dimensions that are statistically supported. We present only the essential steps to obtaining the reduced matrix, and refer readers

Table 1. Eigenanalysis of the genetic variance in transcript abundance among the first 10 transcripts.

Probe	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9	e_{10}
λ_i^1	27.844	10.862	6.429	4.725	2.959	2.124	1.086	0.479	0.000	0.000
1579	0.849	−0.345	0.015	−0.355	0.124	0.043	0.018	−0.015	−0.064	0.114
2315	0.201	0.106	0.031	0.496	0.080	0.483	0.593	0.294	0.113	0.102
3476	0.034	0.454	−0.586	−0.449	0.047	−0.140	0.360	0.208	−0.010	−0.231
3739	−0.100	0.243	0.572	−0.521	−0.234	0.231	−0.021	0.383	0.241	0.140
678	0.288	0.500	−0.017	0.134	0.110	0.048	−0.269	−0.405	0.631	−0.008
1030	−0.016	−0.278	−0.019	0.139	0.225	−0.554	−0.001	0.512	0.529	0.046
2624	−0.279	−0.083	0.091	−0.241	0.877	0.255	0.032	−0.112	0.016	0.007
1481	−0.193	−0.283	−0.552	−0.132	−0.181	0.414	−0.221	0.056	0.273	0.480
1786	−0.168	−0.376	0.091	−0.206	−0.230	−0.040	0.551	−0.476	0.390	−0.206
594	0.041	−0.221	−0.080	0.030	−0.057	0.380	−0.303	0.233	0.139	−0.793

[1]The eigenvalues (λ_i) of each eigenvector (e_i) are shown in the first row.

wishing to understand the somewhat lengthy statistical and linear algebraic background to Amemiya (1985).

In practice, the λ_i are easily obtained as the eigenvalues of

$$\mathbf{Lm}_{bb}\mathbf{L}^{\mathrm{T}}, \qquad (3)$$

where $\mathbf{L}$ is a lower triangular matrix and defined as the transpose inverse of $\mathbf{U}$ (upper triangular) which in turn is the Cholesky root of $\mathbf{m}_{ww}$. As the first step in determining $\hat{\mathbf{\Sigma}}_{bb}$, the eigenvectors of (3) are assembled as columns to form the matrix $\mathbf{Q}$. Then define the matrix

$$\mathbf{P} = \mathbf{L}^{-1}\mathbf{Q}.$$

The first m columns of $\mathbf{P}$ are assembled as the $p \times m$ matrix $\mathbf{P}_m$, and are associated with the m characteristic roots of $\mathbf{m}_{bb}$ in the metric of $\mathbf{m}_{ww}$ that received statistical support. Now let $\mathbf{\Lambda}_m$ be a diagonal matrix of the m significant λ_i, and $\mathbf{I}_{mm}$ be the $m \times m$ identity matrix. The reduced-rank covariance matrix that consists only of those m supported dimensions is then

$$\hat{\mathbf{\Sigma}}_{bb} = r^{-1}\mathbf{P}_m(\mathbf{\Lambda}_m - \mathbf{I}_{mm})\mathbf{P}_m^{\mathrm{T}},$$

Here, the first m eigenvectors of $\hat{\mathbf{\Sigma}}_{bb}$ represent the combinations of transcripts that are associated with each independent genetic module. We used the arbitrary cut-off of 70% of the largest coefficient (Table S1) of an eigenvector (Mardia et al. 1979) to determine which transcripts contributed strongly to each eigenvector. This approach indicated that two transcripts contributed strongly to each of the first two eigenvectors (see results), and three of these four candidate genes were chosen for amplification using qRT-PCR across the larger sample of 41 RILs.

To determine if CHCs were genetically correlated with the genetic modules, we first decomposed the variation among the eight CHCs into two principal components that explained 88.8% of the total variation in CHCs. This was required as a consequence of the limited number of degrees of freedom (Lai et al. 2008)

available for among-line hypothesis testing. We then applied a multivariate regression (implemented using the GLM procedure in SAS) using the RIL means for the first two genetic modules, calculated using the linear equations for each of the two eigenvectors, as the independent variables, and the RIL means for the first two principal components of the CHC means as the response variables.

qRT-PCR EXPRESSION OF CANDIDATE GENES

For each of the two parental lines and the 41 RILs, the RNA from 10 individual males was extracted using Trizol (Invitrogen) following the manufacturer's protocols and then treated with 2 μl of DNase I (Roche, Switzerland) for 30 min at 37°C to eliminate genomic DNA. Approximately 0.5 μg of total RNA was then reverse transcribed to generate cDNA using random primers and SuperScript III reverse transcriptase (Invitrogen) according to manufacturer's protocols. Primers were designed to amplify three of the four candidate loci associated with the first two genetic modules as identified by the microarray analysis (Table 1). Primers were as follows for each of the following *D. serrata* ESTs (Supp Table 1): CL481Contig1 (CG10514 ortholog, probe ID 3011), 171 Forward ACGGGGATGTGTGGACTAAC and 276 Reverse GGAGAGCCCCAGAAGGAATA; CL600Contig1 (*ninaD* ortholog), 460 Forward TCGTGCTGAAATTGATGAGG and 583 Reverse GGTGCCAACGGCTATAAGAA; and CL470Contig1 (*Amy-P* ortholog), 112 Forward ATCAGTTGCGGTACCTGTCC and 260 Reverse GTACTGCTTGGGCACCTTGT. Expression data were not obtained for one of the *D. serrata* ESTs, CL0Est000004994973G08 (CG10514 ortholog, probe ID 1579), associated with module 1 (Table 1). Like CL481Contig1, this EST was also orthologous to CG10514 in *D. melanogaster*. In fact these two ESTs shared the same sequence over 415 base pairs. CL481Contig1 possessed an additional 311 bp at the 5′ end

that were not present in CL0Est000004994973G08 and the latter possessed 70 bp at the 3' end that were not present in CL481Contig1. Although there is no evidence that the *D. melanogaster* ortholog CG10514 produces multiple transcripts, the two *D. serrata* ESTs acted independently and in opposition to one another with respect to module 1 as identified by the transcriptional profiles of the 15 arrays (Table S1). So although the primers above for CL481 effectively amplified the unique region in this EST, due to the short length and base composition, a parallel set of primers could not be developed that successfully amplified the unique region in CL0Est000004994973G08.

Quantitative PCR (qPCR) was performed on a Rotorgene 6000 (Corbett Life Science, Sydney, NSW) using Platinum®SYBR®Green (Invitrogen Inc, Carlsbad, CA) according to manufacturer's instructions. For each sample, a mastermix of 2 μl RNase-free water, 5 μl of SYBR Supermix, and 0.5 μl of each primer (10 μM) was added to 2 μl of cDNA. Three replicates were run for each sample. The cycling protocol was as follows; 1 cycle UDG incubation at 50°C for 2 min, 1 cycle *Taq* activation at 95°C for 2 min, 40 cycles of denaturation at 95°C for 5 s, annealing at 60°C for 5 s, extension at 72°C for 15 s, fluorescence acquisition 78°C, and 1 cycle of melt curve analysis from 68°C to 95°C in 1°C steps. Only one biological replicate per line was analyzed in a single qRT-PCR run for a total of 10 independent runs per gene. Each of the qRT-PCR runs was therefore a complete randomized experimental block in this experimental design. The effect of runs on the raw CT values for the candidate genes was first removed using one-way ANOVA, and the residuals obtained were then analyzed in conjunction with the CHC phenotypes of the same males as outlined in the main text.

Results
CHC PHENOTYPES

The reproductive character displacement in the eight CHC traits can be represented as a single phenotypic trait constructed from the linear combination of the eight CHCs that differ most between allopatric and sympatric natural populations (Higgie et al. 2000). The distribution of the phenotypic means of the reproductive character displacement trait in the parental lines and 41 RILs shows that the parental lines have extreme phenotypes for this trait (Fig. 1), indicating that the construction of the parental lines successfully captured the divergence in phenotype found in the natural sympatric and allopatric populations. RILs were predominantly distributed evenly between the sympatric and allopatric parents. However, six RILs displayed CHC phenotypes that were significantly more extreme than the parental line from Forster (allopatric) population, indicating transgressive segregation that may be attributable to either purely additive or epistatic effects.

TRANSCRIPT ABUNDANCE

The 15 RILs phenotyped for transcript abundance of 3762 ESTs revealed that 400 (10.6%) of these displayed significant among-RIL variation in expression levels after Bonferroni correction, indicating that a substantial proportion of the genome differed in transcript abundance between the two parental lines. The significance of these transcripts corresponds to a false discovery rate (Storey and Tibshirani 2003) of 2.4×10^{-5} (using the *q-value* R package). The distribution of heritability of transcript abundance was highly skewed (Fig. 2), in contrast to the approximately normal distribution of heritability seen within outbred natural populations (Skelly et al. 2009). Because the RILs segregate only for the genetic variation that was present between the original two parental inbred lines, a large number of loci are expected to be fixed for the same allele in both parents at those loci that have not diverged among the two parent populations. These loci will therefore exhibit little segregating variance among the RILs, resulting in the majority of transcripts having very low heritability.

Many of the top-ranked transcripts displayed two distinct clusters of RILs. In Figure 3A, we display a typical example of such a pattern using two of the top-ranked transcripts. This pattern is consistent with two alleles at a single locus underlying much of the genetic variation among RILs for each transcript. Importantly, RILs tended not to co-segregate for the same combinations of these putative allelic classes for different transcripts (not shown),
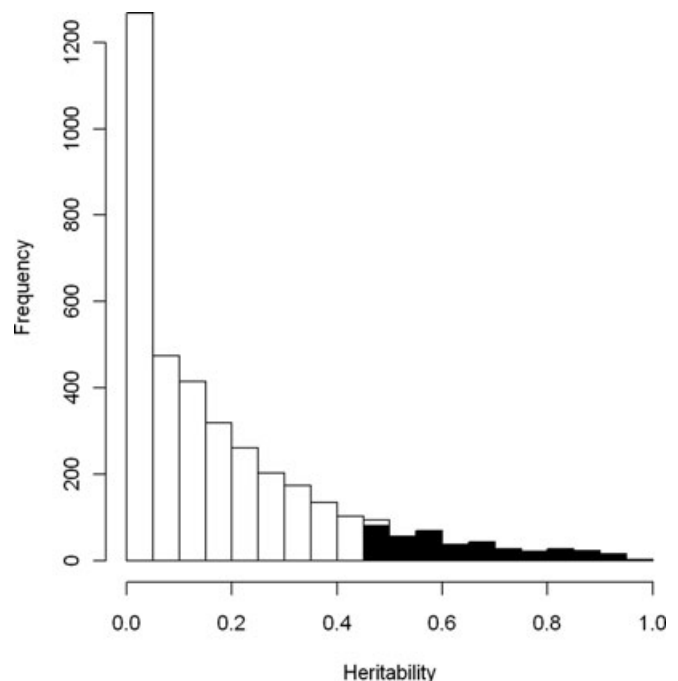

Figure 2. Distribution of heritability in transcript abundance among 15 RILs of *D. serrata*. The 400 of the 3762 heritabilities that remain significant after Bonferroni correction are shown in filled bars.
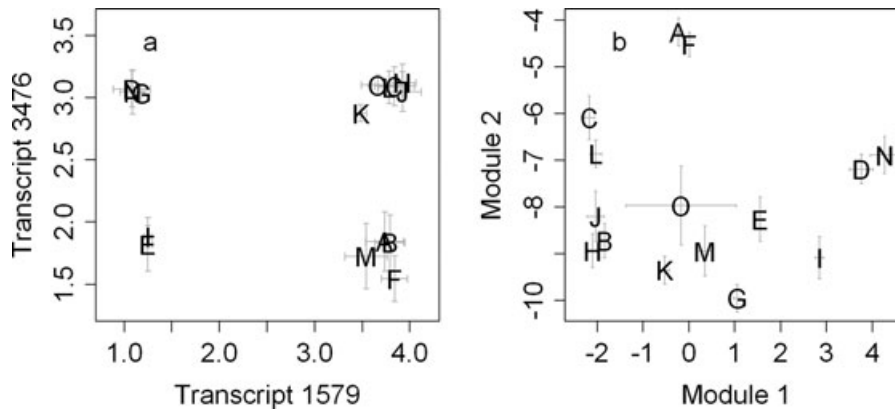
Figure 3. Typical patterns of segregation of gene expression among 15 RILs of *D. serrata*. (A) RIL means represented by capital letters and their 95% confidence intervals for the first and third transcripts that varied the most among RILs. Most transcripts that were highly divergent among the RILs show this pattern of segregation into two distinct groups, resulting in four groups when plotted in the two-dimensional space. (B) RIL means (±95% CIs) for the first two eigenvectors of the genetic variance among the RILs. Note how the RIL means do not fall into discrete groups as is the case with single transcripts.

suggesting that each RIL had a different combination of these putative alleles.

We first determined the dimensionality of the effect space of the among-RIL genetic variance (see Materials and Methods) in transcript abundance for the first 10 transcripts that displayed the most genetic variation among the RILs (Table 1). The reduced-rank **G** matrix ($\hat{\boldsymbol{\Sigma}}_{bb}$) had eight eigenvalues that explained a significant amount of the among-RIL genetic variance. The eigenvectors associated with these eigenvalues (Table 1) represented eight combinations of the 10 transcripts that are co-regulated by independent genetic modules. Each eigenvector tended to have contributions from a number of transcripts, indicating that each individual transcript is not regulated by its own independent module. In other words, it is unlikely that a single gene controls the expression of a single transcript for each of these transcripts that have diverged most among the two parental populations.

We then proceeded to determine the dimensionality of the genetic variance among the 82 transcripts that changed expression to the most significant degree among the RILs, which represented the limit for this experiment given our sample size (Fig. S1). In total, 12 significant modules explained 97% of the total estimated genetic variance among the 82 transcripts. The first two genetic modules explained 50% of the genetic variation in transcript abundance controlled by the 12 modules. In other words, half of all the genetic variation expressed in the 82 transcripts is accounted for by just two underlying genetic sources. In contrast to single transcripts, genetic modules did not display a pattern of simple biallelic segregation among the RILs (Fig. 3B). A greater number of discrete phenotypic combinations were present, suggesting that modules may represent combinations of genotypes at more than one locus.

Only a handful of transcripts were strongly influenced by each independent genetic module (Table S1). No more than seven transcripts contributed strongly to each factor, and only four transcripts contributed strongly to the first two modules. Furthermore, many (13 of 27) transcripts identified as being strongly influenced by an independent genetic module were also strongly influenced by more than one module (Table 1 and S1). These patterns, in conjunction with the fact that eight genetic modules could be identified within just the first 10 transcripts, suggest that adaptation may have targeted specific transcripts for a substantial change in regulation, but this has occurred in conjunction with smaller changes in a larger number (>400) of transcripts.

TRANSCRIPT-MORPHOLOGY GENETIC ASSOCIATIONS

To determine if the underlying genetic modules were associated with the major morphological traits known to be under reinforcing selection, we first correlated the RIL line means for the first two

Table 2. Genetic correlations between individual CHCs and the first two genetic modules of gene expression explaining 50% of the genetic variance in transcript abundance.

CHC	Module 1	Module 2
5, 9-C_{24}	0.419	−0.131
5, 9-C_{25}	0.404	−0.359
9-C_{25}	0.497	−0.131
2-Me-C_{26}	0.065	0.581
5, 9-C_{27}	0.159	0.673
2-Me-C_{28}	0.153	0.625
5, 9-C_{29}	0.151	0.672
2-Me-C_{30}	0.061	0.506

modules with the means for the CHCs (Table 2). The two modules had opposite effects; module 1 was more strongly associated with 5, 9-C_{24}, 5, 9-C_{25}, and 9-C_{25}, whereas module 2 was more strongly associated with the remaining CHCs and had negative associations with the first three. Multivariate multiple regression indicated a significant association between these first two principal components of CHCs (see Materials and Methods) and the first two genetic modules (Wilks' lambda = 0.383, $F_{4,22} = 3.38$, $P = 0.027$). The restricted number of RILs limited our ability to further explore the associations between the independent genetic modules and the CHC phenotypes. However, the opposing effects of the first two modules on CHCs, and their significant association with CHC expression indicated that the underlying genetic modules controlling gene expression identified by our analyses are either pleiotropically related or physically linked to those morphological traits under strong selection in the sympatric and allopatric parent populations.

The biological functions of the 82 most significant transcripts are diverse (Table S2). Most of the transcripts are of unknown function (28%) followed then by an association with either metabolism (14%) or proteolysis (12%). Genes associated with the top three genetic factors (Table S2) have previously been shown to exhibit male-biased expression (http://141.61.102.16:8080/sebida/index.php) and be involved with phototransduction, carbohydrate metabolism, and microtubule-based movement, respectively. In most cases, the link between the documented functional roles for the genes and the mating phenotypes studied here is not self-evident. This is commonly the case (Ayroles et al. 2009), given the nature of complex traits and the limitations to our functional knowledge. Regardless, several of the genes involved in the top 12 modules do appear in potentially relevant studies in *Drosophila* (Table S1). Mutants of Ade5, which is influenced by three independent factors, show increased male–male aggression (Edwards et al. 2009). The gene is highly expressed in the head, spermatheca and carcass (www.flyatlas.org). The amy-P locus (factors 2 and 3) exhibits elevated K_a/K_s ratios in closely related *Drosophila* species, which has been interpreted as the signature of directional selection acting during speciation (Civetta and Singh 1998). The expression of CG31148, jon65Aiii, and Obp99c is down regulated in females after mating (McGraw et al. 2004). Obp99c is also differentially expressed in genetic lines selected for either fast- or slow-mating responsiveness (Mackay et al. 2005). Dhc62B is a member of a family of genes encoding dynein heavy chains that are highly expressed in the testes and thought to play a role in sperm flagella assembly and motility (Rasmusson et al. 1994). Lastly, jon66ci shows phenotype plasticity and genotype by environment interactions with respect to olfactory behavior that could be involved with mating (Sambandan et al. 2008).

EXPRESSION OF INFLUENTIAL GENE CANDIDATES IN PARENTAL LINES AND EXPANDED LIST OF RILs

Independent expression analysis using qRT-PCR of three influential candidate genes in the top two genetic modules was significantly phenotypically associated with the eight CHCs traits across the 430 individuals subjected to both the qRT-PCR and GC analysis as determined by canonical correlation analysis. Canonical correlation analysis allows the variation among individuals in gene expression to be associated with the variation among the same individuals in CHC phenotype. Because there were three candidate genes, only three dimensions (canonical variates) of candidate gene expression could be associated with CHC phenotype (eight dimensions), All three testable dimensions displayed a significant association between gene expression and CHCs ($F_{24,1183.9} = 5.08$, $P < 0.0001$; $F_{14,818} = 3.61, P < 0.0001; F_{6,410} = 2.70, P = 0.014$). Because 10 replicate individuals from each of the 41 RILs were included in this experiment, a genetic correlation between candidate gene expression and CHC phenotype could be estimated for each of the three pairs of canonical variates. The variance component correlation at the among-RIL level in a multivariate mixed model was estimated using REML, and genetic correlations were tested for significance from zero using a log likelihood ratio test. The presence of significant genetic correlations for each of the pairs of canonical variates ($r_G = 0.592$, $\chi_1 = 15.45$, $P < 0.001$; $r_G = 0.383$, $\chi_1 = 5.97$, $P = 0.015$; $r_G = 0.316$, $\chi_1 = 3.73$, $P = 0.053$), confirmed the genetic association between the expression of these three genes and the variation in CHC expression.

Discussion

We have shown that a multivariate approach to the genetic analysis of high-dimensional transcript abundances successfully isolates genetically independent sets of co-regulated transcripts that in turn are genetically correlated with the major morphological changes that have occurred through selection. The nature of regulatory changes during adaptation is controversial, with the importance of different regulatory mechanisms (Hoekstra and Coyne 2007), and the extent of pleiotropic regulatory control (Skelly et al. 2009) is unclear. The genetic variation that segregates among the RILs in the 82 transcripts that displayed the greatest levels of genetic variance was explained by a small number of genetic modules. Although most of the genetic variation in single transcripts segregated in a pattern that superficially resembled single biallelic loci, 97% of this genetic variation was accounted for by only 12 genetically independent modules. In addition, only the first two of these modules were required to explain 50% of the total genetic variance in the 82 transcripts.

The overriding impression of changes in gene regulation in response to reinforcing selection on the CHCs is that the vast

majority of regulatory changes are likely to have occurred as a consequence of divergence in a small number of trans-regulatory loci or through a limited number of cis-regulatory changes for physically linked genes. Recent global expression QTL (eQTL) mapping studies have indicated that a very large number of eQTL may underlie variation in gene expression and that multiple eQTL often affect each transcript, but that the proportion of variation accounted for by each eQTL is very low (West et al. 2007). Our multivariate analysis suggests that such numerous eQTL may either play only a very minor role during a response to selection, or alternatively may simply be overestimated as a result of the highly correlated nature of expression phenotypes (Kadarmideen et al. 2006).

Although a large number of transcripts have diverged between the parental lines of the two populations, relatively few transcripts were strongly affected by the genetic modules uncovered by our multivariate approach. These influential transcripts had a diverse range of putative functions. This array study and others like it (Civetta et al. 1998; McGraw et al. 2004; Mackay et al. 2005; Edwards et al. 2009;) are revealing novel functional roles for genes in complex behaviors, such as mating, whose function has, previously only, narrowly been defined by annotation or mutational analysis. A key future question to address is whether the large number of minor changes in expression play a functional role in an adaptive event, or if much of this elaborate co-regulation may represent nonadaptative (or even deleterious) pleiotropic consequences (Lynch 2007) of the putative trans-regulatory genotypes that underlie each independent genetic module. A response to selection is almost always accompanied by deleterious correlated responses to selection in unselected traits (Falconer and Mackay 1996), and such elaborate co-regulation provides one mechanism that could explain the almost ubiquitous nature of this observation.

A highly pleiotropic model of gene regulation as an explanation for the modules that have responded to selection is consistent with the role ascribed to regulatory networks in the evolution of animal form (Carroll 2008). It should be emphasized however that it is unlikely that either a cis-regulatory change at a single locus defines a module, or that such modules act on completely independent regulatory networks. Modules segregated in a fashion consistent with combinations of genotypes at more than one locus contributing to each module, and expression of the same transcript was governed by a number of different, genetically independent modules. These patterns suggest that epistatic interactions between loci controlling gene regulation (West et al. 2007; Gjuvsland et al. 2007) may be an important component of the adaptive response. Phenotypic differences generated by the interaction between specific genotypes have been shown to be associated with expression differences in a large number of transcripts (Dworkin et al. 2009). Unfortunately, our experimental design precludes a statistical partitioning of the potential effects of such interactions as additive and epistatic effects are confounded in the variance component estimates from our RILs.

The genetic dimensionality underlying multiple phenotypes is a vital component to understanding how populations respond to selection and the mechanisms by which genetic variation is maintained in natural populations (Walsh and Blows 2009). With the advent of new statistical approaches to the analysis of high-dimensional phenotypes, it is becoming clear that the number of phenotypes that can be measured on organisms far exceeds the number of independent genetic factors that explain the genetic variation in these phenotypes (Walsh and Blows 2009). Systems genetic analysis of extreme high-dimensional gene expression data is likely to continue to suffer from computational limitations associated with the application of restricted maximum likelihood based mixed-model approaches (Kadarmideen et al. 2006). The approach presented here provides a way of obtaining insights into the modular genetic control of complex changes in gene expression without the need for mixed-model convergence for such high-dimensional problems (Hine and Blows 2006).

LITERATURE CITED
Albert, A. Y. K., S. Sawaya, T. H. Vines, A.K. Knecht, C. T. Miller, B. R. Summers, S. Balabhadra, D. M. Kingsley, and D. Schluter. 2008. The genetics of adaptive shape shift in stickleback: pleiotropy and effect size. Evolution 62:76–85.

Amemiya, Y. 1985. What should be done when an estimated between-group covariance matrix is not nonnegative definite? Am. Stat. 39:112–117.

Anderson, T. W., and Y. Amemiya. 1991. Testing the dimensionality in the multivariate analysis of variance. Stat. Prob. Lett. 12:445–463.

Ayroles, J. F., M. A. Carbone, E. A. Stone, K. W. Jordan, R. F. Lyman, M. M. Magwire, S. M. Rollmann, L. H. Duncan, F. Lawrence, R. R. H. Anholt, et al.. 2009. Systems genetics of complex traits in *Drosophila melanogaster*. Nat. Genet. 41: 299–307.

Biswas, S., J. D. Storey, and J. M. Akey. 2008. Mapping gene expression quantitative trait loci by singular value decomposition and independent component analysis. BMC Bioinform. 9:244.

Blows, M. W. 2007. A tale of two matrices: multivariate approaches in evolutionary biology. J. Evol. Biol. 20:1–8.

Blows, M. W., S. F. Chenoweth, and E. Hine. 2004. Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. Am. Nat. 163:329–340.

Brown, W. L., and E. O. Wilson. 1956. Character displacement. Syst. Zool. 5:49–65.

Carroll, S. B. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134:25–36.

Chesler, E. J., L. Lu, S. Shou, Y. Qu, J. Gu, J. Wang, H. C. Hsu, J. D. Mountz, N. Baldwin, M. A. Langston, et al. 2005. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. Nat. Genet. 37:233–242.

Civetta, A. and R. S. Singh. 1998. Sex-related genes directional sexual selection and speciation. Mol. Biol. Evol. 15:901–909.

Cheverud, J.M. 1996. Quantitative genetic analysis of cranial morphology in the cotton-top (*Saguinus oedipus*) and saddle-back (*S. fuscicollis*) tamarins. J. Evol. Biol. 9:5–42.

Dworkin, I., E. Kennerly, D. Tack, J. Hutchinson, J. Brown, J. Mahaffey, and G. Gibson. 2009. Genomic consequences of background effects on *scalloped* mutant expressivity in the wing of *Drosophila melanogaster*. Genetics 181:1065–1076.

Edwards, A. C., L. Zwarts, A. Yamamoto, P. Callaerts, and T. F. Mackay. 2009. Mutations in many genes affect aggressive behavior in *Drosophila melanogaster*. BMC Biol. 7:29.

Falconer, D. S., and T. F. C. Mackay. 1996. Introduction to quantitative genetics. Addison Wesley Longman, Harlow, Essex, UK.

Franchini, A. G., and T. Egli. 2006. Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions. Microbiology 152:2111–2127.

Frentiu, F. D., M. Adamski, E. A. McGraw, M. W. Blows, and S. F. Chenoweth. 2009. An expressed sequence tag (EST) library for *Drosophila serrata* a model system for sexual selection and climatic adaptation studies. BMC Genom. 10:40.

Gibson, G., R. Riley-Berger, L. Harshman, A. Kopp, S. Vacha, S. Nuzhdin, and M. Wayne. 2004. Extensive sex-specific nonadditivity of Gene Expression in *Drosophila melanogaster*. Genetics 167:1791–1799

Gilad, Y., S. A. Rifkin, and J. K. Pritchard. 2008. Revealing the architecture of gene regulation: the promise of eQTL studies. Trend. Genet. 24:408–415.

Gjuvsland, A. B., B. J. Hayes, S. W. Omholt, and O. Carlborg. 2007. Statistical epistasis is a generic feature of gene regulatory networks. Genetics 175:411–420.

Hansen T. F. 2003. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. BioSystematics 69:83–94.

Higgie, M., and M. W. Blows. 2007. Are traits that experience reinforcement also under sexual selection? Am. Nat. 170:409–420.

———. 2008. The evolution of reproductive character displacement conflicts with how sexual selection operates within a species. Evolution 62:1192–1203.

Higgie, M., S. F. Chenoweth, and M. W. Blows. 2000. Natural selection and the reinforcement of mate recognition. Science 290:519–521.

Hine, E., and M. W. Blows. 2006. The effective dimensionality of the genetic variance-covariance matrix. Genetics 173:1135–1144.

Hine, E, K. McGuigan, and M.W. Blows. 2011. Natural selection stops the evolution of male attractiveness. Proc. Natl. Acad. Sci. USA 108:3659–3664.

Hoekstra, H. E., and J. A. Coyne. 2007. The locus of evolution: evo devo and the genetics of adaptation. Evolution 61:995–1016.

Howard, D. J. 1993. Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. Pp. 46–69 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process. Oxford Univ. Press, New York.

Houle, D. 2010. Numbering the hairs on our heads: The shared challenge and promise of phenomics. Proc. Natl. Acad. Sci. USA 107:1793–1799.

Kadarmideen, H. N., P. von Rohr, and L. L. G. Janss. 2006. From genetical genomics to systems genetics: potential applications in quantitative genomics and animal breeding. Mammal Genom. 17:548–564.

Lai, Z., N. C. Kane, Y. Zou, and L. H. Rieseberg. 2008. Natural variation in gene expression between wild and weedy populations of *Helianthus annuus*. Genetics 179:1881–1890.

Litvin, O., H. Causton, B.-J. Chen, and D. Pe'er. 2009. Modularity and interactions in the genetics of gene expression. Proc. Natl. Acad. Sci. USA 106:6441–6446.

Lynch, M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity. Proc. Natl. Acad. Sci. USA 104:8597–8604.

Mackay, T. F., S. L. Heinsohn, R. F. Lyman, A. J. Moehring, T. J. Morgan, and S. M. Rollmann. 2005. Genetics and genomics of *Drosophila* mating behavior. Proc. Natl. Acad. Sci. USA 102:6622–6629.

Mardia, K. V., J. T. Kent, and J. M. Bibby. 1979. Multivariate analysis. Academic Press, London.

McGraw L. A., G. Gibson, A. G. Clark, and M. F. Wolfner. 2004. Genes regulated by mating sperm or seminal proteins in mated female *Drosophila melanogaster*. Curr. Biol. 14:1509–1514.

Meyer, K., and M. Kirkpatrick. 2008. Perils of parsimony: properties of reduced-rank estimates of genetic covariance matrices. Genetics 180:1153–1166.

Mezey, J. G., and D. Houle. 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. Evolution 59:1027–1038.

Orr, H. A. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Ranz, J. M., C. I. Castillo-Davis, C. D. Meiklejohn, and D. L. Hartl. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science 300:1742–1745.

Rasmusson, K., M. Serr, J. Gepner, I. Gibbons and T. S. Hays. 1994. A family of dynein genes in *Drosophila melanogaster*. Mol. Biol. Cell 5: 45–55.

Ronald, J. and J. M. Akey. 2007. The evolution of gene expression QTL in *Saccharomyces cerevisiae*. PloS ONE 2:e678.

Sambandan, D., M. A. Carbone, R. R. H. Anholt, and T. E. C. Mackay. 2008. Phenotypic plasticity and genotype by environment interaction for olfactory behavior in *Drosophila melanogaster*. Genetics 179:1079–1088.

Shirangi, T.R., H. D. Dufour, T. M. Williams, and S. B. Carroll. 2009. Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. PloS Biol. 7:e1000168.

Skelly, D.A., J. Ronald, and J. M. Akey JM. 2009. Inherited variation in gene expression. Annu. Rev. Genomics Hum. Genet. 10:313–332.

St-Cyr, J., N. Derome, and L. Bernatchez. 2008. The transcriptomics of life-history trade-offs in whitefish species pairs (*Coregonus* sp.). Mol. Ecol. 17:1850–1870.

Stone, E.A., and J. F. Ayroles. 2009. Modulated modularity clustering as an exploratory tool for functional genomic inference. PloS Genet. 5: e1000479

Storey, J.D., and R. Tibshirani. 2003. Statistical significance for genome-wide experiments. Proc. Natl. Acad. Sci. USA 100: 9440–9445.

Walsh, B., and M. W. Blows. 2009. Abundant genetic variation + strong selection = multivariate genetic constraints: a geometric view of adaptation. Ann. Rev. Ecol. Evol. Syst. 40:41–59.

West, M. A. L., K. Kim, D. J. Kliebenstein, H. van Leeuwen, R. W. Michelmore, R. W. Doerge, and D. A. St. Clair. 2007. Global eQTL mapping reveals the complex genetic architecture of transcript-level variation in Arabidopsis. Genetics 175:1441–1450.

White, K. P., S. A. Rifkin, P. Hurban, and D. S. Hogness. 1999. Microarray analysis of *Drosophila* development during metamorphosis. Science 286:2179–2184.

Xu, C. W., Z. K. Li, and S. Z. Xu. 2005. Joint mapping of quantitative trait loci for multiple binary characters. Genetics 169:1045–1059.

Ye, Y. H., S. F. Chenoweth, and E. A. McGraw. 2009. Effective but costly evolved mechanisms of defense against a virulent opportunistic pathogen in *Drosophila melanogaster*. PLoS Pathog. 5:e10000385.

Associate Editor: L. Moyle

Supporting Information

The following supporting information is available for this article:

Figure S1. Relationship between the number of transcripts included in an analysis and the number of significant genetic modules identified.

Table S1. Eigenvectors of the 82-dimensional effect space representing 12 genetic modules that account for a significant amount of the genetic variance among the RILs.

Table S2. Putative functions of the gene products that contribute strongly to the 12 genetic modules.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.