

Environmental and genetic perturbations reveal different networks of metabolic regulation

Anthony J Greenberg^{1,*}, Sean R Hackett^{1,3}, Lawrence G Harshman² and Andrew G Clark¹

¹ Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA and ² School of Biological Sciences, University of Nebraska-Lincoln, Lincoln, NE, USA

³ Present address: Graduate Program in Quantitative and Computational Biology, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA

* Corresponding author. Department of Molecular Biology and Genetics, Cornell University, 101 Biotechnology Building, Ithaca, NY 14853, USA.
Tel.: +1 607 255 1707; Fax: +1 607 255 4698; E-mail: ajg67@cornell.edu

Received 25.7.11; accepted 25.10.11

Progress in systems biology depends on accurate descriptions of biological networks. Connections in a regulatory network are identified as correlations of gene expression across a set of environmental or genetic perturbations. To use this information to predict system behavior, we must test how the nature of perturbations affects topologies of networks they reveal. To probe this question, we focused on metabolism of *Drosophila melanogaster*. Our source of perturbations is a set of crosses among 92 wild-derived lines from five populations, replicated in a manner permitting separate assessment of the effects of genetic variation and environmental fluctuation. We directly assayed activities of enzymes and levels of metabolites. Using a multivariate Bayesian model, we estimated covariance among metabolic parameters and built fine-grained probabilistic models of network topology. The environmental and genetic co-regulation networks are substantially the same among five populations. However, genetic and environmental perturbations reveal qualitative differences in metabolic regulation, suggesting that environmental shifts, such as diet modifications, produce different systemic effects than genetic changes, even if the primary targets are the same.

Molecular Systems Biology 7: 563; published online 20 December 2011; doi:10.1038/msb.2011.96

Subject Categories: metabolic and regulatory networks

Keywords: Bayesian model; G matrix; hierarchical model; metabolic network

Introduction

System-level approaches to biological questions can yield important insights inaccessible to traditional reductionist methods (Misarović, 1968; Kitano, 2002). Properties of biological networks, such as the influence of their architecture on system robustness (Kitano, 2004) and control (Kacser and Burns, 1973; Fell, 2005), as well as evolutionary change in the individual components and the architectures themselves, are of great fundamental interest. The utility of systems approaches for other biological disciplines lies in the possibility to predict the results of experimental or natural perturbations. For example, while developing effective therapies, we would like to know the unintended consequences of administering a drug that is thought to target specific cellular processes. When studying fundamental evolutionary processes, it is important to account for constraints on the effectiveness of natural selection as a result of pleiotropic effects of naturally arising mutations.

Progress in systems biology is predicated on the accurate description of the relevant networks. Discovery of regulatory networks typically proceeds by conducting measurements (e.g., microarray analyses to assess transcript abundance) in a variety of conditions. Gene co-regulation is measured using some variant of the correlation coefficient (Eisen *et al.*, 1998;

Ayroles *et al.*, 2009). The system is usually perturbed by varying growth conditions (Eisen *et al.*, 1998) or sampling a set of engineered (Capaldi *et al.*, 2008) or naturally occurring mutations (Schadt, 2005; Rockman, 2008; Ayroles *et al.*, 2009). Frequently, co-regulated genes are sorted into modules to aid functional annotation or identify sets of genes involved in a particular process (Eisen *et al.*, 1998; Ayroles *et al.*, 2009). Genetic perturbations have also been used to estimate directed causal relationships in networks (Wagner, 2001; Rockman, 2008; Logsdon and Mezey, 2010). Sometimes, genetic perturbations are guided by networks discovered by varying environmental states (Amit *et al.*, 2009). It is unclear, however, whether the networks discovered through environmental perturbations are the same as the ones identified genetically.

Pairwise correlations describe both direct and indirect relationships. A matrix of these correlations, in addition to summarizing associations, can be used to predict the deviation of outcomes of system changes from those expected under independence of variables, an idea that has been used by biometricians for more than a century (Pearson, 1903). If we add variance information to the correlations, we get a covariance matrix. In particular, the matrix of additive genetic covariances (the G matrix, comprising the genetic correlation matrix and information on additive genetic variability of characters) has long been used to predict how phenotypic

selection within a generation translates to between-generation evolutionary change (Lande, 1980b; Lande and Arnold, 1983; Arnold, 1992).

The matrix-based approach to predicting how regulatory network structure will affect the outcome of system perturbation is quite appealing due to its relative mathematical and computational simplicity. However, it can only be used if a set of conditions is met. First, the nature and magnitude of changes we wish to study should be similar to the set of conditions used to ascertain the correlation matrix. Second, the matrix itself has to remain invariant under perturbation. The first condition is often under our control, and so is relatively easy to meet. The second is a property of the system and has to be established before we can use the correlation matrix for predictive purposes. A number of methods have been developed for estimation and comparison of G matrices (McGuigan, 2006). However, despite many theoretical and empirical studies, no general rules emerged that would indicate when to expect these matrices to be refractory to perturbation (Turelli, 1988; McGuigan, 2006). By combining direct molecular analyses of pathway components and traditional biometric approaches, we can both identify regulatory networks and use them to predict outcomes of directional changes in biological systems.

We set out to investigate the utility of covariance matrices among molecular traits as quantitative attributes for inference in systems biology. Rather than identify detailed network topology based on partial correlations, we were interested in capturing direct as well as indirect relationships, since both kinds of covariation provide information that is useful for prediction (Lande, 1980b). To accomplish this, we chose to focus on intermediary metabolism of the fruit fly, *Drosophila melanogaster*, as a model by assaying enzyme activities. Quantitative assays of a number of enzyme activities have already been established in flies (Laurie-Ahlberg *et al*, 1980; Clark and Keith, 1989). Most investigations of regulatory network structure have relied on measurements of mRNA levels. However, enzyme activities have a more direct impact on physiology, enhancing our power to predict organismal phenotypes. Moreover, natural genetic variation in enzyme activities has fitness consequences, often underlying adaptation (Clark, 1989; Berenbaum *et al*, 1996; Eanes, 1999). Abundant genetic and genomic resources are available for *D. melanogaster* and will allow experimental tests of hypotheses generated using our approach.

To estimate variance–covariance matrices among enzyme activities, we need a set of perturbations. Natural genetic variation, sampled using a set of wild-caught lines, can be a source of such perturbations. However, correct estimation of variance–covariance matrices from a sample of lines presents significant challenges (Lande and Arnold, 1983; McGuigan, 2006). Precision of estimates requires large sample sizes, while the separation of various genetic and environmental sources of variation hinges on proper experimental design. We therefore developed a crossing scheme that maximizes the number of lines sampled while still allowing the partitioning of the sources of phenotypic variance (Greenberg *et al*, 2010). In this study, we extend the univariate Bayesian statistical methods we previously implemented for this experimental design (Greenberg *et al*, 2010) to the analysis of multiple enzyme activities and estimation of their covariances.

We measured activities of 19 enzymes (see Figure 2), along with five physiological variables. This number of parameters is smaller than our sample of lines, allowing for reasonably precise estimates of covariance matrices, while covering the major pathways of intermediary metabolism (with the exception of β -oxidation). We sampled lines from five populations of *D. melanogaster* and conducted measurements in both males and females. Our Bayesian statistical methods allowed us to implement a fully probabilistic approach to the characterization of metabolic regulation networks. Using our model, we assessed variability of and dependence among metabolic parameters, and partitioned environmental and genetic sources of variation. This analysis found that the environmental co-regulation network is largely invariant among populations, and the genetic one is only slightly less so. Environmental and genetic correlations are largely the same in males and females, although there are notable differences. While we see measurable similarities between sets of correlations among physiological parameters induced by genetic and environmental perturbations, these networks are different in important ways. These results suggest that system-wide end point effects of genetic and environmental changes will be different, even if their primary targets are the same.

Results

Data structure and parameter estimation

We sampled 92 inbred lines from five *D. melanogaster* populations (see Materials and methods and Greenberg *et al*, 2010 for details) and crossed them to control for the effects of inbreeding (Figure 1A). Each cross was replicated in three blocks, picking two samples of five flies of each sex, weighing them, and homogenizing them for enzyme assays (see Figure 1B). We assayed each sample for activities of 19 enzymes (Figure 2 and Materials and methods for the list of enzymes, their EC numbers and abbreviations we use in the text), in addition to four metabolite pools: total triglyceride levels (TRI), glycogen (GLG), total protein (tPRT), and mitochondrial protein (mPRT). We measured important enzymes in most pathways of intermediary metabolism, focusing on reactions that were well established and could be used in a high-throughput format given technical, time, and budget constraints. We collected 1632 samples in each sex and performed a total of over 75 000 enzyme and metabolite assays. We include the data in Supplementary information.

The crossing scheme we employed maximizes the number of lines that can be assayed, given constraints on the total size of the experiment (Greenberg *et al*, 2010). It includes inbred lines themselves (gray diagonal boxes in Figure 1A), a round-robin set of crosses within each population and a nearly balanced set of between-population crosses. Our experimental design requires that we take into account different cross types, measurement errors and outliers in enzyme assays, and the hierarchical structure of the data. To estimate parameters from this data set, we previously developed a set of hierarchical Bayesian models (Greenberg *et al*, 2010). We further showed that the model that assumes a Student's *t*-distribution among replicates performs better than

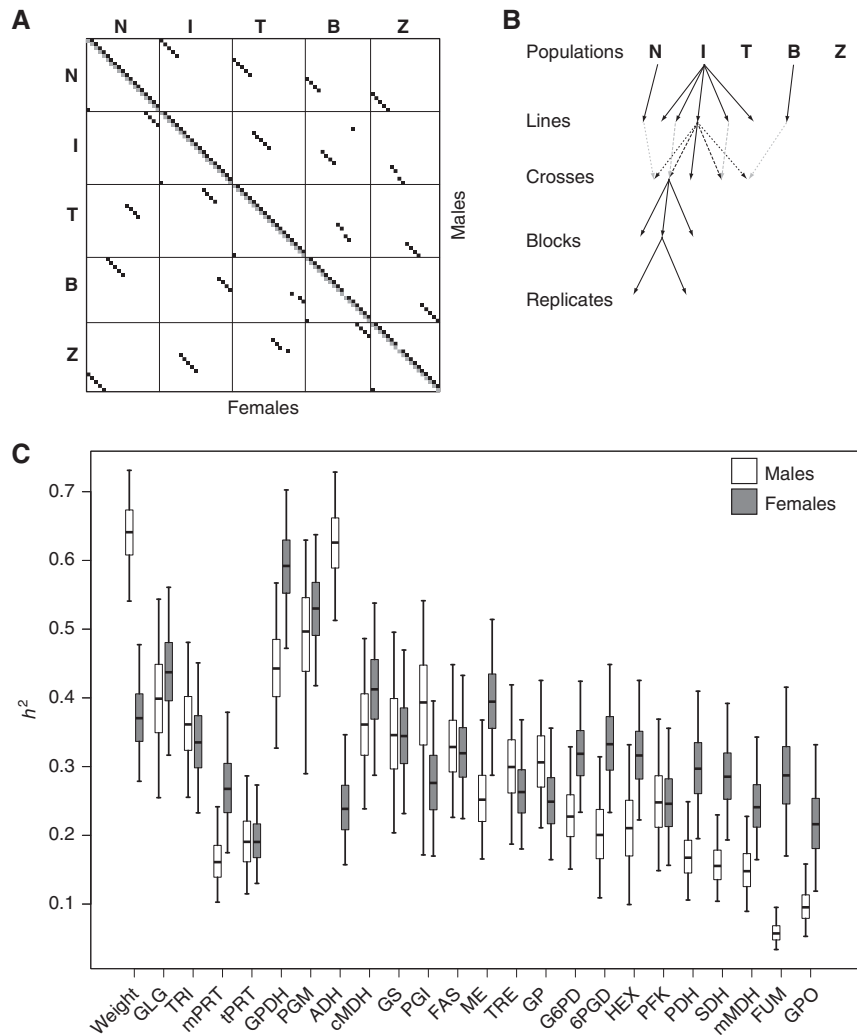


Figure 1 Data structure and heritability estimates. **(A)** The diallel crossing scheme. Male parents are listed top to bottom, female parents—left to right. Black boxes indicate the crosses we performed. Gray boxes mark the 'selfed' inbred lines. **(B)** A depiction of the data structure showing the replication scheme and nesting of variables. **(C)** Narrow-sense heritability (h^2) estimates for all parameters in males and females. For each estimate, the whiskers indicate the 95% credible interval, the box—the inter-quartile range, and the thick black line—the median of the samples from the posterior distribution. The data are arranged in the order of decreasing sex-averaged h^2 separately for physiological variables and enzymes.

comparable Gaussian Bayesian and maximum-likelihood models when outlier observations are present. Such outliers are evident in our data, and are due to occasional aberrant reactions. The number of outliers was variable among enzyme assays, but generally did not exceed 1% of the samples. Eliminating unusual observations from the data set would involve deriving some criterion for their identification, and setting some arbitrary threshold for their exclusion. In contrast, our modeling approach reduces the influence of these observations, while keeping them in the data.

Our previously developed model analyzes each enzyme separately, under the assumption that each enzyme activity is independent of others. This assumption is clearly violated in the present set of experiments, confirming previous findings (Laurie-Ahlberg *et al*, 1980; Clark, 1989) that enzyme activities are correlated. Taking advantage of the fact that we measured each enzyme activity in the same sample of flies, we extended the previous univariate model to a full multivariate treatment.

Thus, we are estimating variance–covariance matrices at each level of the hierarchy (Figure 1B). We lay out the model in the relevant section of Materials and methods and Supplementary Text 1. Briefly, we employed multivariate normal distributions for modeling sample parameters, such as line means. Inverse variance–covariance matrices were drawn from Wishart distributions, with diffuse Wishart priors (Gelman *et al*, 2004). The exceptions were replicates, which we modeled using multivariate Student's t -distributions, and population means, which we assumed independently Student's t -distributed for each enzyme. We implemented a Gibbs sampler (Gelman *et al*, 2004) to approximate posterior distributions of parameters.

Levels of variation in enzyme activity

Variance–covariance matrices summarize two distinct aspects of the biological system: the variability of each parameter and

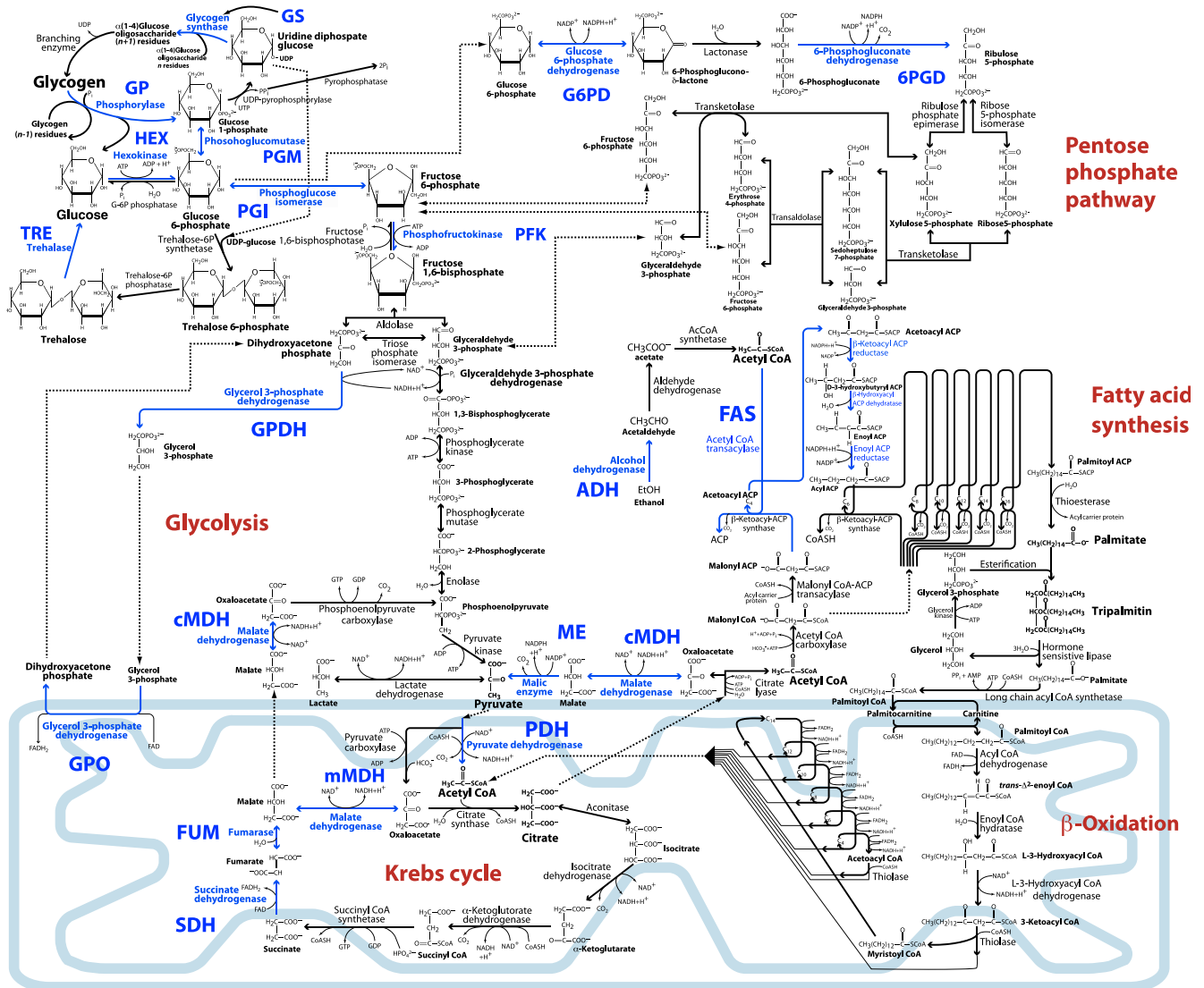


Figure 2 A scheme of intermediary metabolism. Enzymes assayed in this study are marked in blue, together with the abbreviations we used in the text.

the degree of dependence among parameters. Variability is reflected in the variances, located on the diagonal of a covariance matrix. Non-independence is measured by correlations, which are covariances standardized by the products of relevant standard deviations. We examined these two facets separately. First, we looked at variation in enzyme activity at each level of the hierarchy (Figure 1C; Table I).

We first examined narrow-sense heritability (h^2 ; the fraction of total phenotypic variation that is due to additive genetic variance; Lynch and Walsh, 1998) for each enzyme and physiological variable (Figure 1C). This parameter cannot be estimated from inbred lines alone. It has an important role in quantitative genetics, because it reflects the portion of total phenotypic variation that parents contribute to their offspring, and thus is an important factor in determining the potential of a character to respond to selection (Lynch and Walsh, 1998). We observed a variety of heritability levels, from low to moderately high. The range appears wider in males than in females. However, it is important to note that we likely

overestimated h^2 because our samples represent averages of five individuals, and our model has a slight upward bias when true heritability is low (Greenberg *et al*, 2010). Weight and ADH activity in males, and GPDH activity in females (medians of posterior distributions > 60%) display the highest h^2 . Only a small proportion of total variance is due to additive genetic effects for mitochondrial enzymes in males, particularly GPO and FUM (medians of posterior distributions 10% or smaller). Most physiological variables show little evidence of sex differences in h^2 , with substantial overlap in 95% credible intervals (Bayesian analogs of 95% confidence intervals; Figure 1C). The exceptions are weight and ADH activity, where heritability in males is much higher than in females, and fumarase, whose heritability in females is substantially higher than in males.

Since heritability is a ratio of the genetic and total variance, differences in h^2 can be due to disparities in the level of any component of phenotypic variation. We therefore examined variances at each level to elucidate the sources of disparities in

Table I Coefficients of variation for each level of the data hierarchy

	Replicate	Block	SCA	GCA
<i>Weight</i>				
♂	0.023 (0.021, 0.026)	0.033 (0.028, 0.038)	0.026 (0.020, 0.031)	0.065 (0.054, 0.079)
♀	0.030 (0.027, 0.033)	0.087 (0.079, 0.097)	0.006 (0.003, 0.011)	0.072 (0.060, 0.088)
<i>TRI</i>				
♂	0.301 (0.274, 0.331)	0.133 (0.087, 0.183)	0.065 (0.034, 0.104)	0.274 (0.222, 0.338)
♀	0.241 (0.220, 0.264)	0.115 (0.071, 0.171)	0.031 (0.020, 0.050)	0.218 (0.178, 0.270)
<i>GLG</i>				
♂	0.090 (0.081, 0.100)	0.102 (0.089, 0.117)	0.080 (0.070, 0.092)	0.188 (0.158, 0.228)
♀	0.099 (0.089, 0.110)	0.116 (0.100, 0.135)	0.090 (0.078, 0.104)	0.238 (0.200, 0.289)
<i>tPRT</i>				
♂	0.203 (0.186, 0.223)	0.100 (0.079, 0.126)	0.059 (0.049, 0.072)	0.133 (0.109, 0.165)
♀	0.176 (0.161, 0.193)	0.096 (0.076, 0.123)	0.055 (0.045, 0.067)	0.115 (0.094, 0.143)
<i>mPRT</i>				
♂	0.156 (0.143, 0.171)	0.065 (0.047, 0.085)	0.035 (0.027, 0.044)	0.082 (0.067, 0.102)
♀	0.157 (0.144, 0.171)	0.067 (0.049, 0.090)	0.046 (0.034, 0.062)	0.126 (0.100, 0.160)
<i>ADH</i>				
♂	0.171 (0.152, 0.190)	0.093 (0.062, 0.126)	0.049 (0.026, 0.077)	0.280 (0.235, 0.342)
♀	0.766 (0.672, 0.873)	0.215 (0.124, 0.348)	0.097 (0.050, 0.175)	0.508 (0.405, 0.642)
<i>ME</i>				
♂	0.199 (0.178, 0.222)	0.084 (0.051, 0.119)	0.040 (0.022, 0.057)	0.137 (0.111, 0.169)
♀	0.205 (0.187, 0.226)	0.079 (0.050, 0.115)	0.030 (0.018, 0.048)	0.209 (0.173, 0.258)
<i>cMDH</i>				
♂	0.059 (0.052, 0.067)	0.057 (0.048, 0.067)	0.040 (0.034, 0.048)	0.099 (0.082, 0.121)
♀	0.359 (0.316, 0.409)	0.359 (0.282, 0.462)	0.189 (0.152, 0.237)	0.558 (0.456, 0.693)
<i>FAS</i>				
♂	0.223 (0.201, 0.246)	0.131 (0.088, 0.172)	0.050 (0.017, 0.094)	0.198 (0.161, 0.245)
♀	0.235 (0.213, 0.257)	0.088 (0.044, 0.139)	0.020 (0.010, 0.048)	0.192 (0.156, 0.240)
<i>PGI</i>				
♂	0.084 (0.077, 0.091)	0.030 (0.020, 0.042)	0.015 (0.008, 0.025)	0.089 (0.073, 0.110)
♀	0.083 (0.075, 0.092)	0.036 (0.023, 0.052)	0.013 (0.007, 0.024)	0.095 (0.078, 0.117)
<i>PFK</i>				
♂	0.131 (0.120, 0.143)	0.048 (0.031, 0.067)	0.017 (0.010, 0.028)	0.092 (0.074, 0.116)
♀	0.149 (0.134, 0.164)	0.100 (0.073, 0.135)	0.020 (0.012, 0.035)	0.127 (0.103, 0.159)
<i>G6PD</i>				
♂	0.645 (0.585, 0.714)	0.133 (0.069, 0.221)	0.059 (0.025, 0.110)	0.384 (0.308, 0.484)
♀	0.674 (0.607, 0.753)	0.189 (0.094, 0.315)	0.056 (0.026, 0.115)	0.502 (0.413, 0.618)
<i>SPGD</i>				
♂	0.246 (0.212, 0.313)	0.150 (0.105, 0.210)	0.043 (0.026, 0.066)	0.183 (0.148, 0.229)
♀	0.184 (0.163, 0.207)	0.100 (0.067, 0.134)	0.031 (0.019, 0.050)	0.175 (0.143, 0.218)
<i>PGM</i>				
♂	0.103 (0.093, 0.113)	0.066 (0.050, 0.083)	0.019 (0.011, 0.030)	0.155 (0.130, 0.188)
♀	0.087 (0.079, 0.095)	0.048 (0.034, 0.064)	0.016 (0.009, 0.029)	0.152 (0.127, 0.184)
<i>GP</i>				
♂	0.880 (0.785, 0.998)	0.173 (0.082, 0.320)	0.065 (0.027, 0.136)	0.638 (0.512, 0.809)
♀	0.459 (0.419, 0.505)	0.101 (0.038, 0.194)	0.016 (0.008, 0.039)	0.286 (0.230, 0.358)
<i>GS</i>				
♂	0.062 (0.055, 0.071)	0.074 (0.066, 0.084)	0.066 (0.059, 0.075)	0.159 (0.134, 0.191)
♀	0.079 (0.071, 0.089)	0.095 (0.082, 0.115)	0.076 (0.067, 0.089)	0.186 (0.157, 0.225)
<i>TRE</i>				
♂	0.139 (0.124, 0.155)	0.081 (0.060, 0.104)	0.044 (0.034, 0.057)	0.122 (0.099, 0.151)
♀	0.115 (0.104, 0.128)	0.071 (0.053, 0.095)	0.039 (0.031, 0.049)	0.105 (0.085, 0.131)
<i>HEX</i>				
♂	0.127 (0.115, 0.140)	0.064 (0.044, 0.085)	0.018 (0.010, 0.030)	0.099 (0.080, 0.124)
♀	0.137 (0.124, 0.151)	0.076 (0.050, 0.103)	0.024 (0.013, 0.042)	0.132 (0.107, 0.163)

Table I Continued

	Replicate	Block	SCA	GCA
<i>GPDH</i>				
♂	0.147 (0.128, 0.167)	0.104 (0.075, 0.134)	0.073 (0.051, 0.095)	0.186 (0.151, 0.230)
♀	0.132 (0.112, 0.154)	0.073 (0.048, 0.104)	0.048 (0.026, 0.078)	0.243 (0.202, 0.298)
<i>GPO</i>				
♂	0.471 (0.431, 0.516)	0.080 (0.043, 0.137)	0.040 (0.018, 0.073)	0.163 (0.126, 0.211)
♀	0.522 (0.472, 0.578)	0.208 (0.123, 0.327)	0.056 (0.032, 0.100)	0.333 (0.264, 0.422)
<i>PDH</i>				
♂	0.183 (0.168, 0.200)	0.061 (0.041, 0.083)	0.033 (0.025, 0.042)	0.096 (0.079, 0.118)
♀	0.146 (0.133, 0.160)	0.089 (0.062, 0.126)	0.031 (0.022, 0.042)	0.137 (0.113, 0.168)
<i>FUM</i>				
♂	0.376 (0.344, 0.413)	0.100 (0.064, 0.144)	0.036 (0.025, 0.048)	0.105 (0.082, 0.135)
♀	0.237 (0.216, 0.261)	0.110 (0.076, 0.155)	0.052 (0.037, 0.075)	0.201 (0.158, 0.254)
<i>mMDH</i>				
♂	0.220 (0.203, 0.239)	0.066 (0.044, 0.093)	0.034 (0.022, 0.048)	0.105 (0.083, 0.132)
♀	0.203 (0.186, 0.222)	0.093 (0.058, 0.139)	0.041 (0.027, 0.059)	0.140 (0.113, 0.176)
<i>SDH</i>				
♂	0.941 (0.826, 1.085)	0.365 (0.290, 0.466)	0.228 (0.190, 0.278)	0.474 (0.384, 0.594)
♀	1.100 (0.889, 1.404)	0.965 (0.759, 1.269)	0.628 (0.500, 0.821)	1.337 (1.036, 1.791)

Numbers in parentheses represent 95% credible intervals. Coefficients of variation are standard deviations divided by the corresponding mean. Specific combining ability (SCA) is the deviation of cross means from mid-parent values and reflects dominance effects.

heritability. The units of measurement for each enzyme are different, and to some extent arbitrary, precluding direct comparison of variances among enzymes. We therefore calculated coefficients of variation (CV), which are ratios of standard deviations and mean values, and bring all measurements to a natural standard scale (Haldane, 1949; Houle, 1992; Hansen and Houle, 2008). We find (Table I; see Supplementary Table S1 for unscaled variances) that CV in general combining ability (the statistical parameter that measures variance among line means and estimates additive genetic variation; Sprague and Tatum, 1942; Lynch and Walsh, 1998) are quite similar among enzymes, and are generally in the range of 10–25%. For example, the differences in h^2 of ADH activity between males and females are largely due to increased replicate CV in females. Likewise, low heritability of mitochondrial enzyme activity is due to relatively high replicate CV. Measurements of weight stand out for their low overall variability. On the other side of the spectrum, SDH activity appears to be particularly variable, with environmental and genetic standard deviations on the order of the mean.

Genetic and environmental correlation networks are distinct

Having documented substantial environmental and genetic variation in enzyme activities and physiological variables in our data, we turned to the analysis of correlations among the characters we measured. We wanted to evaluate the ability of pleiotropic mutations segregating in natural populations to induce covariation in enzyme activities. To accomplish this, we estimated the matrix of correlations among the deviations of line means from their respective population means. This

matrix is the correlation component of the G matrix, i.e., the portion of the genetic covariance that parents contribute to their offspring. The G matrix occupies an important place in quantitative genetics of multivariate traits (Lande, 1980b; Lande and Arnold, 1983) as the predictor of the extent and direction of selection. Uncontrolled environmental affects, such as fluctuations in temperature and food quality among replicates, can also affect regulators of metabolism. This can potentially result in concerted changes in activities of several enzymes, leading to environmentally induced correlations among them. We sought to estimate these correlations and compare them with those induced by mutations. Our experimental design allows us to estimate two kinds of environmental covariance matrices: those among replicates and among blocks (Figure 1B). The relationships we estimate are between deviations of replicate values from the corresponding block means, and block means from cross means. To decide which of these environmental covariance matrices better reflects biological effects, we performed a simulation study (see Supplementary Text 2 for details). We simulated data with the same environmental and genetic correlation matrices and checked how well the estimated matrices matched true values. Our results show that the replicate covariance matrix is by far the best at capturing true correlations. We therefore focused on it for the rest of the study as reflecting environmental effects.

Our estimates of the genetic and environmental correlation matrices for both sexes are listed in Supplementary Tables S2 and S3. We illustrate the relationships graphically in Figure 3A and B. We did not attempt to separate direct and indirect relationships because both kinds of associations affect system behavior (Lande, 1980b). Our initial focus is on the results for males, since variation in germline size has a smaller role in males than in females.

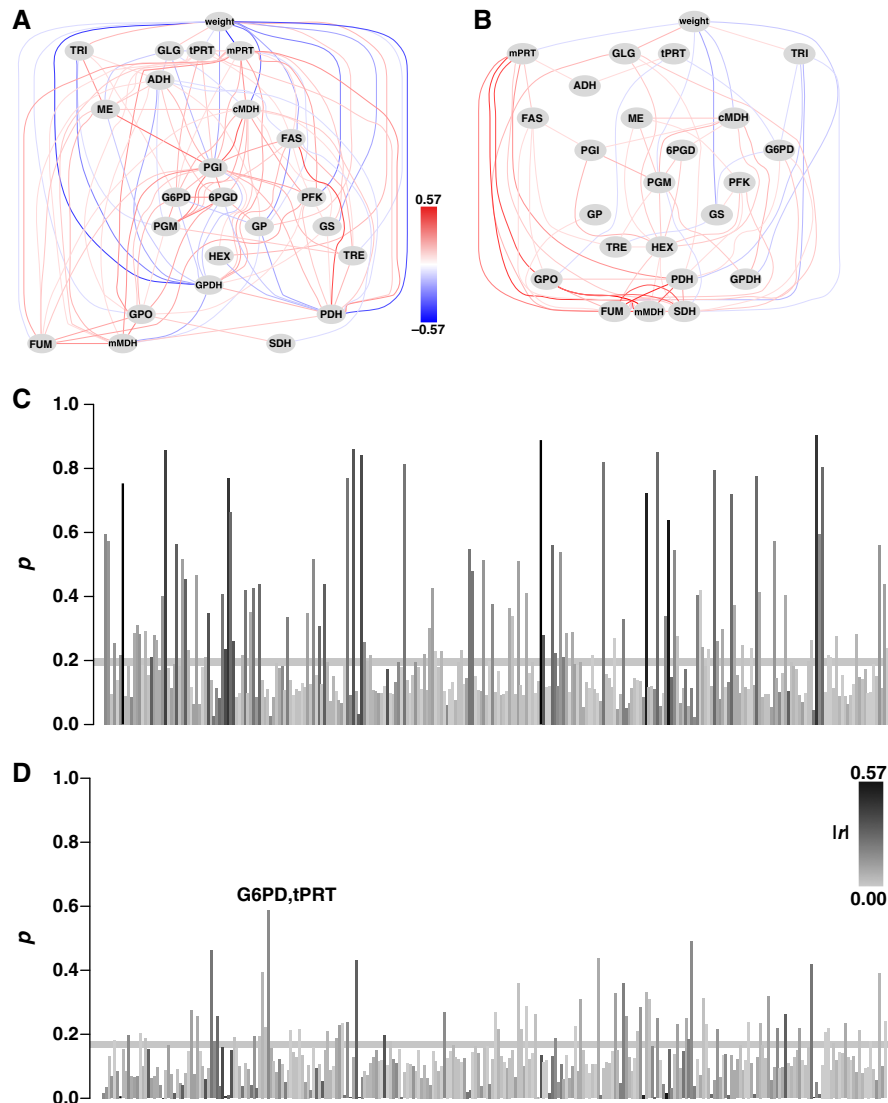


Figure 3 Correlation network structure. **(A)** The genetic correlation network estimated from male data. Gray ovals represent the variables and the lines depict correlations among them. The color of each line reflects the median posterior value of the correlation coefficient. The color distribution is described by the bar to the right. For clarity, only correlations with absolute values > 0.1 are depicted. **(B)** The environmental correlation network in males. This panel is organized the same as **(A)**, with the same color scheme. Differences in node arrangement reflect disparities in network structure. **(C)** Posterior probabilities of a given correlation being present in both the genetic and environmental matrices, and retaining the same sign. The color of each bar reflects the absolute value of the genetic correlation; the distribution of colors is depicted in **(D)**. The individual correlations are arranged along the X axis. This arrangement is identical in all figures, row by row of the upper triangle of the correlation matrices as represented in Supplementary Tables S2 and S3. The horizontal gray bar reflects the range of expected values for two randomly permuted matrices. **(D)** Posterior probabilities that a given correlation is present in both correlation matrices and has switched sign. The plot is arranged as **(C)**.

If we look at credible intervals of individual correlations, we notice that the majority of them contain zero, with only 19 genetic two-tailed posterior P -values based on these intervals coming under the conventional 5% threshold (63 environmental correlations meet this threshold; Supplementary Tables S2 and S3). However, simulations of independent data and permutations of our data set (see Materials and methods for details) reveal that these individual P -values are misleading. For example, a permutation that destroys genetic correlations, but maintains all the rest of the data features, results in a correlation matrix that has no two-tailed posterior P -values < 0.95 . While the reason for this effect is not entirely clear, at least some of it is due to the fact that individual correlations are

not independent, since not all combinations of them are mathematically possible (Gelman *et al*, 2004). For illustration purposes (Figure 3A and B), we draw only correlations with absolute values > 0.1 , and use colors to represent the corresponding edges proportionally to their value. However, when performing quantitative data analyses we sample from matrix distributions, thus taking as full account as possible of the uncertainty of their estimation and avoiding the use of arbitrary significance cutoffs. This is the essence of the Bayesian approach to statistical inference.

Looking at the correlation graphs, there appear to be substantial differences in structure between the genetic (Figure 3A) and environmental (Figure 3B) correlation

matrices. We wanted to test this quantitatively. A logical place to start was to summarize in one number some measure of closeness between these matrices. A widely used statistic of this sort is an element-wise correlation between (upper triangles) two matrices. If corresponding matrix elements are similar, this correlation is close to one, and it should be zero if the elements are no more likely to be the same than what we would expect by chance. A similar idea is behind the widely used Mantel's test of matrix similarity (Mantel, 1967). To assess the uncertainty of our estimates of this correlation (as well as all the other statistics described in the text), we sampled from the matrix distributions estimated using our hierarchical model (see Materials and methods for details and Supplementary Figure S1 for illustration). The element-wise correlation coefficient between the genetic and environmental matrices is 0.235 (0.126, 0.354; here and elsewhere in this report, the numbers in parentheses are 95% credible intervals), higher than the -0.0004 (-0.119 , 0.114) value we would expect for unrelated matrices ($P=0.0030$). The latter value was obtained by randomly permuting elements in the upper triangle of each matrix and recalculating the correlation (see Materials and methods for details). To test if this result is driven by a particular subset of the data, we performed the same comparison on matrices estimated after removing data for each population in turn. This procedure is similar to a standard jack-knife. We find virtually no difference in the resulting estimates of matrix similarity. Element-wise correlation ranged between 0.232 (0.141, 0.319) when the Tasmania population is removed and 0.313 (0.223, 0.398) when leaving out Zimbabwe, strongly suggesting that our matrix comparisons are not influenced by outlier data points. The 0.235 value is about half the correlation we get if we simulate data with the environmental and genetic matrices fixed equal to each other (Supplementary Text 2). Furthermore, we expect much higher correlations between matrices drawn from the same distribution (0.525 (0.411, 0.636) for the genetic and 0.827 (0.786, 0.862) for the environmental matrix, one-tailed $P<0.0001$; see the Matrix Comparison section of Materials and methods for details).

Dependence among variables insures that outcomes of perturbations to the system will differ from the initial displacement values even if we assume linear change and the existence of an end point steady state. Covariance matrices are linear operators that allow us to predict the final levels of all measured components after perturbation (Lande, 1980b). This observation suggests a matrix comparison method (Cheverud, 1996). It proceeds by first numerically generating unit-length random vectors of perturbation ('random skewers' Cheverud, 1996; see Materials and methods for details). These can be thought of as initial changes to the levels of each system component (level of each enzyme and physiological variable). In the original formulation, genetic covariance matrices were used to predict responses to selection, and the random skewers were thought of as potential selection gradients (Cheverud, 1996). We generalize these ideas to any set of perturbations that attempts to move the system from its current state. To predict the outcome of a chosen perturbation, we multiply the initial vector by a covariance matrix. By comparing the length and direction of the resulting vector with the original random skewer, we can gauge how much the

dependence among parameters affects system behavior. Furthermore, if we multiply the same initial vector separately by two matrices, the angle between the two resulting vectors reflects the salience of the differences in composition of these matrices to their predicted systemic effect. By repeating the process with many random skewers and samples from our covariance matrix distributions, we can probabilistically assess and compare the effects of our matrices on an average perturbation (see Materials and methods for a detailed explanation).

Both the environmental and genetic correlation matrices rotate vectors of initial perturbation to a similar extent. The angle between the environmental vector and the original one is 30.9° (21.4° , 40.4°), while it is 43.0° (32.3° , 53.3°) for the genetic vector. Covariance matrices rotate the initial vectors even more (60.9° (52.6° , 68.1°) and 59.8° (49.2° , 68.8°) for the environmental and genetic matrices, respectively), highlighting the importance of accounting for variation as well as correlation when predicting the outcomes of perturbations. The correlation matrices did not appreciably change the length of the initial vectors (length of the rotated environmental vector was 1.13 (0.81, 1.74; $P=0.4805$) and of the genetic one 1.32 (0.80, 2.23; $P=0.2818$), probabilities indicating deviations from initial values of 1.00). In contrast, vectors resulting from multiplication by covariance matrices were much shorter than the initial unit length (environmental: 0.32 (0.14, 0.48; $P<0.0002$); genetic: 0.16 (0.07, 0.31; $P<0.0002$)). Absolute lengths of the resulting vectors depend on the values of the diagonal elements of covariance matrices, and are therefore hard to interpret directly. However, scaling the covariance matrices by trait means did not measurably change the angle or length of transformed vectors (not shown).

Having established that dependence among physiological variables affects the direction of change in the system, we wanted to know if the genetic and environmental correlation matrices guide the initial perturbations to the same state. The angle between vectors obtained by transforming random skewers by the two correlation matrices is 43.5° (30.8° , 58.6° ; $P<0.0002$). This is clearly higher than the angle we see in simulated data (see Supplementary Figures S2 and S3 in Supplementary Text 2), as well as what we would expect for matrices drawn from the same distribution (38.0° (24.6° , 55.9°) for the genetic and 17.2° (11.2° , 25.0°) for the environmental matrix, one-tailed $P=0.0128$), although the difference is not large. Using covariance matrices, we get a similar result (38.5° (20.2° , 65.6° ; $P<0.0002$)). This suggests that environmental and genetic perturbations produce different systemic effects, even if the initial changes are the same. Each matrix defines a direction of change that is easiest to achieve ('line of least resistance'; Schluter, 1996). This direction coincides with the first principal component vector (Schluter, 1996). The angle between these vectors for the environmental and genetic matrices is 37.0° (18.2° , 81.6° ; $P<0.0002$), suggesting that the lines of least resistance are different in the environmental and genetic regulatory networks.

The above methods reveal the degree of overall matrix similarity. We wanted to explore the behavior of individual relationships between enzymes in more detail. Are some associations retained in both matrices? Do some correlations

switch signs? To answer these questions, we needed a way to identify which correlations were ‘present’ in the matrices, and with what sign, while avoiding the use of arbitrary cutoffs. To accomplish this, at each iteration of the Markov chain, we permuted variables so as to construct a ‘null’ matrix that reflected the expectation when no correlations were present (see Materials and methods for details and Supplementary Figure S1 for illustration). We then calculated the empirical probability that each real correlation did not come from a distribution of 276 null correlations. Each real correlation was stochastically accepted as ‘present’ with this probability and its sign was noted. We repeated this procedure at each step of the Markov chain, generating sets of probabilistically assigned correlation presences or absences. Using this method, we estimated that there were 161 (141, 185) non-zero genetic and 170 (151, 189) environmental correlations, out of 276 possible.

Given the probabilistic presence/absence and sign designations, we can test how many correlations retain the same direction in both the environmental and genetic matrices. Consistent with the element-wise correlation analysis discussed above, the number of elements that are present and retain the same sign (64 (49, 80)) is slightly larger than that expected for two unrelated matrices (54 (40, 69); $P=0.1816$ based on randomly reassigning correlation identities in each matrix). Additionally, the number of sign switches (37 (26, 50)) is somewhat lower than expected by chance (46 (36, 60); $P=0.2066$).

We further looked at each enzyme pair separately, recording how many times, out of the 10 000 Markov chain samples, the correlation between them switches sign, remains the same, is present in one but not the other matrix, or is absent in both. We then compared the incidences of each kind with frequencies for permuted enzyme pair identities to detect deviations from random expectation. The frequency of event incidence per chain sample can be viewed as the posterior probability of that event. The range of probabilities we see after permuting matrix elements delimits the distribution of P -values when no true positives are present and therefore reflects false discovery rates (see the Multiple Testing section of Materials and methods for details). Plots of posterior probabilities of sign identity (Figure 3C) reveal that a number of relationships retain their presence and sign in both matrices at a rate far higher than that expected by chance. These are the correlations with the highest absolute values (Figure 3C). Of the top five such correlations, two involve weight (with GPDH, $P=0.8885$ and cMDH, $P=0.8580$; expected range of P -values, calculated as described in Materials and methods, is 0.1841 to 0.2070),

and are negative in both matrices (see Figure 3A and B and Supplementary Tables S2 and S3). The top retained between-enzyme correlations are mMDH–GPO ($P=0.9057$), PGM–PGI ($P=0.8601$), PGM–6PGD ($P=0.8429$), mMDH–FUM ($P=0.8048$), and FUM–GPO ($P=0.7755$). All of these correlations are positive in both matrices.

None of the five correlations most likely to have switched signs between the genetic and environmental matrices involve weight. These are G6PD–tPRT (positive in the genetic matrix, negative in the environmental one; $P=0.5880$), PDH–HEX (negative genetic, positive environmental; $P=0.4914$), PGI–GLG (negative genetic, positive environmental; $P=0.4612$), GPO–tPRT (positive genetic, negative environmental; $P=0.4389$), and PGM–G6PD (positive genetic, negative environmental; $P=0.4317$). The expected range of P -values for sign switching is between 0.1578 and 0.1793.

Population invariance of correlation matrices

Having documented differences between the genetic and environmental correlation matrices, we wanted to see if they are also population specific. We therefore estimated environmental and genetic correlations within each population and compared them with matrices calculated using the whole data set. We find high element-wise correlations for both kinds of matrices in each population (left side of Table II). Despite these strong similarities, the random skewers method reveals notable differences in the direction of transformed vectors (right side of Table II). However, these differences are within the range expected for matrices drawn from the same distribution (the smallest angle for a population is 40.1° ($24.1^\circ, 65.9^\circ$, $P=0.3315$) for the genetic and 15.6° ($8.7^\circ, 26.3^\circ$; $P=0.1197$) for the environmental matrices. The P -values indicate the probability that an angle between a population-specific and the overall matrix is the same as the angle between matrices drawn from the same distribution; see Materials and methods for details). Looking at the probability that relationship presence and sign is preserved (Supplementary Figure S3), the majority of correlations with even a moderate absolute value retain the same sign in the population-specific and full-sample matrices, while sign switches are rare (Supplementary Figure S4). The biggest source of discrepancy is the absence in the population-specific matrix of some correlations present in the overall one. This is expected, since we have fewer samples and thus less power to detect correlations when we confine ourselves to a single population.

Table II Element-wise matrix correlations and random skewer rotation angles between each population-specific and the overall correlation matrix

Population	Correlations		Rotations	
	Genetic	Environmental	Genetic	Environmental
N	0.662 (0.563, 0.744)	0.787 (0.754, 0.817)	35.9 (22.0, 57.7)	22.5 (14.8, 32.7)
I	0.660 (0.561, 0.741)	0.730 (0.691, 0.766)	36.1 (22.1, 57.4)	23.6 (15.7, 34.8)
T	0.642 (0.544, 0.727)	0.751 (0.707, 0.792)	36.6 (22.7, 57.1)	22.5 (14.7, 33.2)
B	0.571 (0.455, 0.675)	0.638 (0.691, 0.738)	40.6 (25.2, 61.9)	25.5 (17.0, 36.5)
Z	0.651 (0.552, 0.736)	0.805 (0.776, 0.830)	36.8 (22.7, 57.4)	21.3 (14.4, 32.1)

Numbers in parentheses represent 95% credible intervals. Rotations are in degrees.

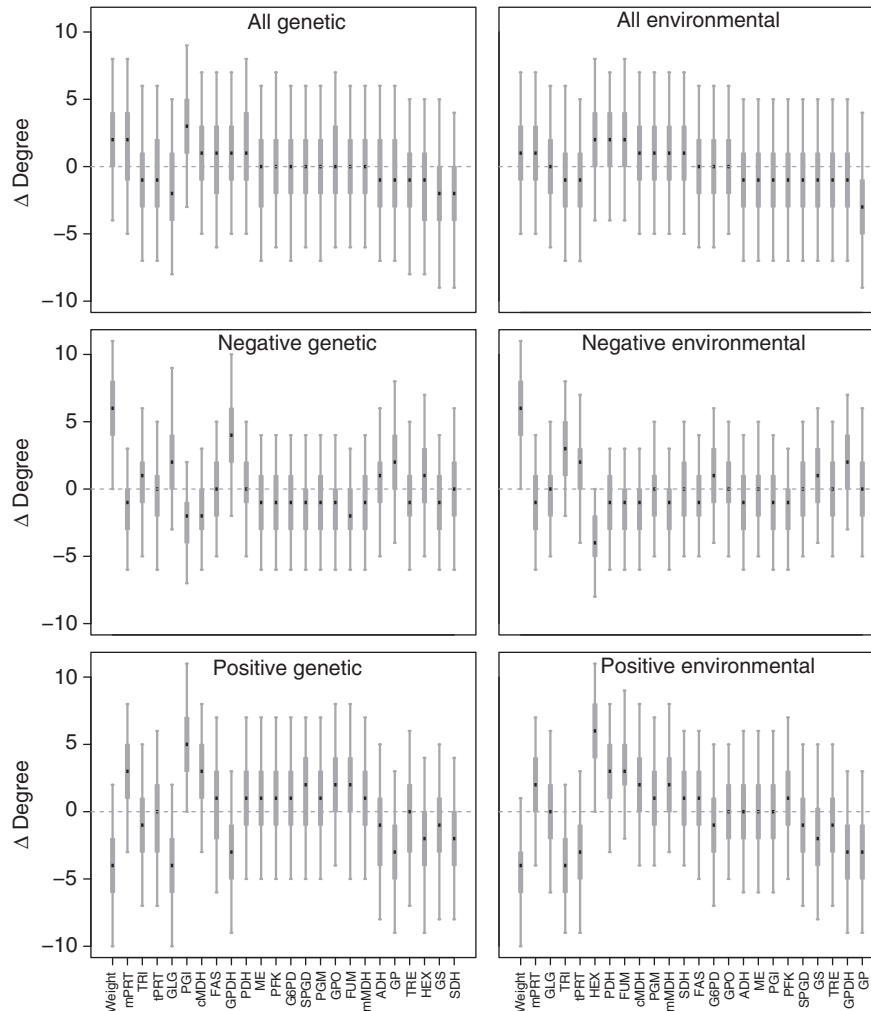


Figure 4 Degree distributions of the genetic and environmental correlation networks. Degree distributions for the genetic correlation network are on the left and the environmental ones on the right. The top two panels depict the sum of positive and negative correlations, the middle panels, only negative correlations, and the bottom panels, only positive correlations. The posterior distributions are shown as in Figure 1C. For each parameter, the value of degree expected under the assumption of a uniform degree distribution (see Materials and methods for details) is subtracted from the measured value and the difference plotted on the Y axis. Enzymes and physiological variables are sorted separately in the descending order of the posterior most likely value of total degree.

Enzyme connectivity patterns

By probabilistically identifying correlation presence as described above, we created samples of enzyme co-regulation networks. We used these samples to characterize the importance of each enzyme and physiological variable to these networks. To quantify importance, we calculated network centrality statistics for each enzyme, and compared these with randomly rewired networks to identify features that are unlikely to occur by chance. A common measure of centrality is ‘degree,’ or the number of connections a given node (in our case, an enzyme or a physiological variable) has with other nodes (Barabási and Oltvai, 2004). We calculated this statistic for every node in each sampled network, and subtracted the value expected for a network with randomly assigned connections. The results are shown in Figure 4 (top pair of plots). It appears that both co-regulation networks have flat degree distributions among nodes, fairly close the expectation for an Erdős-Rényi network (Barabási and Oltvai, 2004). Only two variables, weight

and PGI enzyme activity (weight and HEX activity for the environmental network), show any evidence of departure from expectation under a uniform degree distribution. This becomes clearer when we distinguish positive and negative correlations. Weight has a disproportionate number of negative correlations (Figure 4, middle panels; total number is 11 (6, 15) in both networks), while PGI and HEX have an unusually elevated number of positive correlations (Figure 4, lower panels; total number is 14 (9, 17) for PGI in the genetic network and 15 (11, 19) for HEX in the environmental network). Because we calculate enzyme activities per unit weight (see Materials and methods), the elevated number of negative correlations is likely due to allometric scaling of some enzyme activities with weight, and thus is probably not a feature of the regulatory network *per se*. The observation that PGI and HEX are hubs in the genetic and environmental co-regulation networks of enzymes, however, deserves further analysis.

We were concerned that allometric scaling with weight induces spurious correlations among enzyme activities

(Pearson, 1896–1897; Hayes, 2001). We therefore calculated correlations among enzymes and physiological variables, given their relationships with weight (Supplementary Figure S2). After correcting for correlations with weight, HEX is no longer a hub in the environmentally determined network. In contrast, the evidence for PGI as a hub becomes, if anything, stronger.

To check that the observation of high centrality of PGI is not dependent on the particular statistic used, we calculated node betweenness (Freeman, 1977) for each enzyme and physiological variable. Betweenness is the number of shortest paths between all pairs of nodes in a network that pass through the node under consideration. We find that the distribution of enzyme betweenness is virtually identical to that of degree (not shown), and PGI again emerges as the most central enzyme. To see how PGI's position in the co-regulation network relates to its place in the network of metabolic reactions, we constructed an enzyme-centric graph of these reactions, based on the general information of central metabolism (Salway, 2003) as well as insect-specific sources (Chefurka, 1965; Sacktor, 1970; Figure 2). Overall, there is only a weak relationship between enzyme degree or betweenness in the co-regulation network and these variables in the reaction network ($r=0.168$ ($-0.248, 0.542$), $P=0.4370$ for degree and 0.288 ($-0.163, 0.694$), $P=0.2384$ for betweenness). Furthermore, only 11 (7, 15) pairs of enzymes that share a metabolite are also correlated, exactly the number expected by chance. Despite this general lack of correspondence between the reaction and co-regulation networks, PGI has the highest degree (8) in the network of reactions, and the highest betweenness (502.0) among enzymes whose activity we measured, and thus occupies a central position in both networks.

Sex differences in enzyme activities and correlations

The analyses we presented so far were largely based on the data collected from males. We now consider the differences between sexes. First, we looked at sexual dimorphism in enzyme activities. Since these activities are measured in arbitrary units, we divided the sex differences by male means, thus estimating a fractional deviation of female values from male ones (Figure 5A). To calculate mean sex differences, we subtracted the male from the female values for each line mean, and took the average across all lines. Employing the same procedure across population rather than line means yields virtually identical results (not shown). Sexual dimorphism in enzyme maximal rates is widespread, with only three enzymes (FAS, GPDH, and SDH) showing no evidence of it (Figure 5A). Females appear to have lower levels of most enzymes, but are heavier and have higher triglyceride levels.

Having established extensive differences in mean enzyme activities between males and females, we wanted to know if these disparities persist in each line. To assess this, we calculated between-sex correlations (across all line means) for each enzyme and physiological variable. A high correlation would imply that lines with high levels in males are also likely to have high activities in females. No correlation would

suggest no such relationship, and would indicate a form of genotype by sex interaction. Finally, a negative correlation would imply a possible sex conflict in enzyme activity (Fisher, 1930; Lande, 1980a). We see variation in between-sex correlation strength across physiological variables (Figure 5B). Correlations range from close to zero (cytoplasmic protein levels and SDH activity) to around 0.6 (e.g., PGM and GPDH), but are never negative. Most correlations are between 0.2 and 0.4, suggesting moderate sex-by-genotype interaction.

Our results suggest that males and females differ in their metabolic states, as measured by the constellation of enzyme activities and metabolite levels, and the difference varies somewhat from line to line. Are the genetic and environmental co-regulation networks also distinct between the sexes? To answer this question, we compared both kinds of correlation matrices using the methods outlined above. Looking first at the genetic matrix, we observe a moderate element-wise correlation between males and females (0.336 (0.214, 0.461); $P<0.0002$), while the angle between transformed random skewers is 45.4° ($30.4^\circ, 64.1^\circ$). Furthermore, the presence and sign of 69 (53, 86) correlations are retained, which is more than expected by chance (51 (37, 66); $P=0.0216$), while relatively few correlations switch signs (31 (20, 43) versus 46 (33, 60), $P=0.0422$). At the level of individual relationships, it is clear that the majority of strong correlations retain their sign with high probability (Figure 5C), while only four switch signs (Figure 5D). Interestingly, all of these sign switches involve the cytoplasmic MDH enzyme (correlations with PGM ($P=0.6063$), 6PGD ($P=0.5976$), PFK ($P=0.5532$) and PGI ($P=0.5098$) are positive in males and negative in females).

Similarities between the male and female environmental correlation matrices are just as pronounced. The element-wise correlation is 0.473 (0.398, 0.547), $P<0.0002$, and the angle between transformed skewers is smaller than for genetic matrices (29.5° ($19.7^\circ, 41.0^\circ$)). Despite overall similarity, the number of presence and sign retentions, at 68 (54, 83), is only slightly larger than expected by chance (57 (43, 73), $P=0.1370$). The number of sign switches (37 (26, 49) versus 45 (33, 58), $P=0.2864$) is not much smaller than expected, although only moderately strong correlations switch signs (Figure 5F). Intriguingly, three out of top five sign switches involve cMDH, in concordance with the pattern we see between genetic matrices (correlations with PGM ($P=0.8887$), PGI ($P=0.7671$) and HEX ($P=0.7097$) are positive in males and negative in females).

Discussion

We set out to describe the *D. melanogaster* metabolic regulation network by sampling genetic and environmental perturbations, measuring activities of 19 enzymes and quantifying 5 physiological variables. We achieved the necessary perturbations by crossing 92 inbred lines from 5 populations (Greenberg *et al*, 2010), and replicating this set of crosses to assess the effects of uncontrolled environmental variation (such as fluctuations in food quality, humidity, and ambient temperature). We conducted the experiment on males and females, performing >75 000 reactions in 3264 samples.

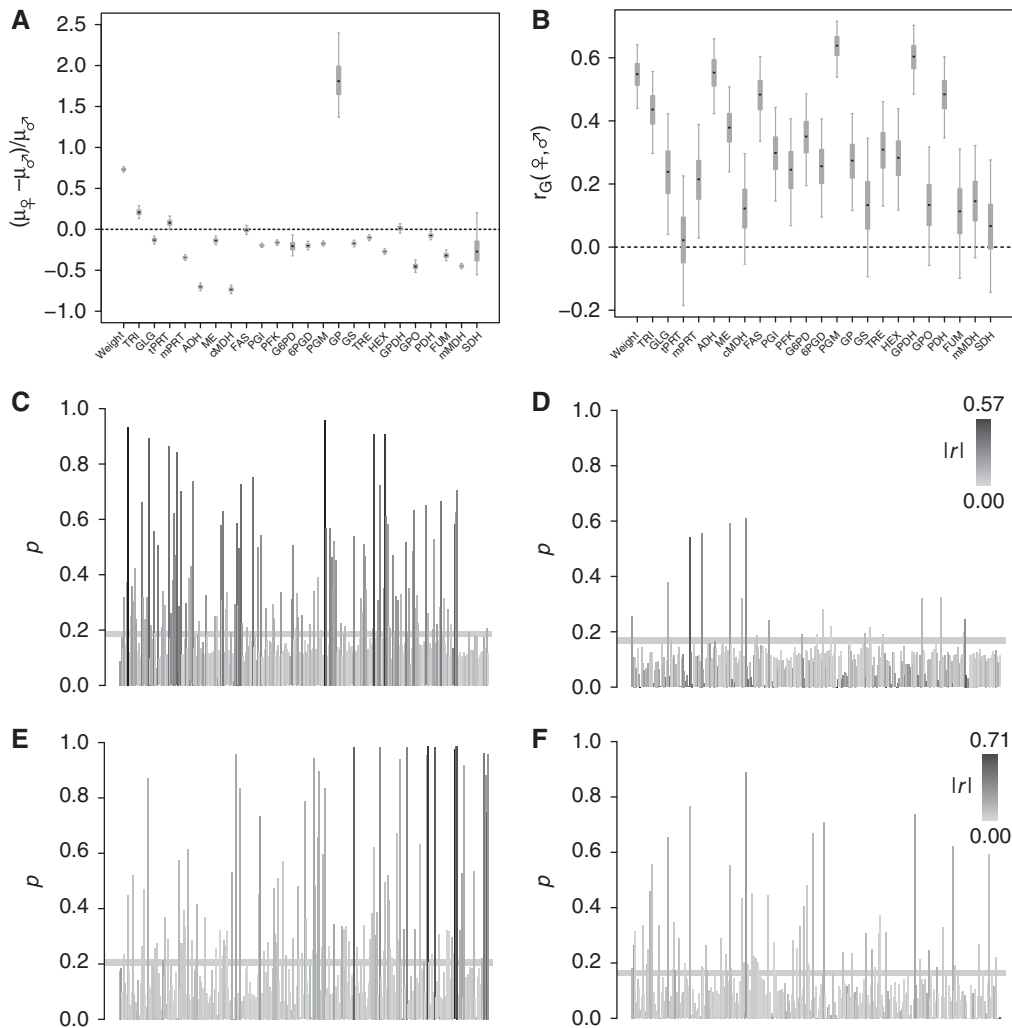


Figure 5 Sex differences in enzyme levels and correlation network structure. **(A)** Sex differences in enzyme levels (among-population means), standardized by male values. **(B)** Across-sex correlations for each variable. Correlations were calculated using breeding values for each line. **(C, D)** Posterior probabilities of a given genetic correlation being present and retaining (C) or switching (D) sign between sexes. The plotting arrangements are the same as those in Figure 3C and D. **(E, F)** The same as (C) and (D), but for environmental correlations.

To analyze this rich, but complicated, data set we developed multivariate Bayesian hierarchical models. This combination of experimental design and analysis tools allowed us to separately construct fully probabilistic descriptions of metabolic regulatory networks that are revealed by environmental and genetic perturbations, and compare these networks. We wanted to capture the direct as well as indirect relationships that are useful for predictive purposes (Lande, 1980b). Of particular interest is the possibility of predicting levels of end point metabolites and physiological variables from enzyme activities.

By performing crosses among lines, we were able to assess the performance of each genotype in a set of genetic backgrounds and control the effect of inbreeding. Thus, we could estimate the proportion of the genetic variation that parents contribute to their offspring and calculate narrow-sense heritabilities and G matrices, parameters that have an important role in predicting responses to selection (Lande, 1980b; Lynch and Walsh, 1998). Previous comprehensive analyses of natural variation in enzyme activities

(Laurie-Ahlberg *et al*, 1980; Clark, 1989) considered only inbred chromosome substitution lines. They were thus forced to lump all components of genetic variation together and ignore the possible effects of inbreeding. Like these previous studies, we find appreciable genetic variation in most enzyme activities and physiological variables. However, our estimates of heritability are generally several-fold higher. The earlier analyses relied on ANOVA and maximum-likelihood-based mixed models with point estimates of enzyme rates for each replicate. In contrast, we incorporated uncertainty in activity estimates, and used a multivariate Student's *t*-model to moderate the influence of outliers. These differences in modeling approaches likely account for the discrepancies in heritability estimates (Greenberg *et al*, 2010). Based on the analyses of simulated data, we previously concluded that in the presence of experimental noise and outliers our Bayesian hierarchical model out-performs maximum-likelihood approaches (Greenberg *et al*, 2010). The latter generally underestimated narrow-sense heritability, often drastically.

Given the differences in experimental design, line choice, and estimation procedures, it is difficult to compare our among-enzyme correlation estimates with previously reported results. We do observe a previously widely reported positive correlation between the activities of G6PD and 6PGD (Bijlsma, 1980; Wilton *et al*, 1982; Clark, 1989). This correlation is likely maintained due to toxicity of an intermediate compound in the pentose phosphate shunt (Hughes and Lucchesi, 1977). We observe this positive correlation in the genetic but not in the environmental matrix (Figure 3A and B; Supplementary Tables S2 and S3). Using a probabilistic method of picking values that are different from those expected under independence of enzyme activities, we estimate that more than half of all possible pairs of physiological parameters are correlated. In contrast, only 6% of all possible relationships are realized in the network of metabolic reactions. Our strategy of measuring a moderate number of variables in a highly replicated set of samples appears to yield high power to detect even weak correlations, revealing a high degree of dependence among enzyme activities. Our results suggest that scoring correlations above some threshold (e.g., satisfying a set false discovery rate) for network discovery risks missing some, perhaps even most, real relationships.

Comparing correlation matrices in a biologically relevant way is not a simple task. We implemented variants of traditional summary methods: element-wise correlations (Mantel, 1967) and Cheverud's random skewers test (Cheverud, 1996). While the former appears to highlight similarity, the latter statistic is quite sensitive to even small differences in matrix composition. We also developed a probabilistic method of tracking which individual correlations are retained in two matrices and which switch sign.

Neither the genetic nor the environmental correlations differ much from population to population. We are therefore reasonably certain that the relationships we identified are not driven by a small number of unusual observations. This is further confirmed by extensive similarity of both sets of correlations across sexes. While there is measurable similarity between the environmental and genetic co-regulation networks, important differences are apparent. We see an appreciable number of sign switches, and distributions of the number of correlations each enzyme has with others is clearly different. Finally, the environmental and genetic correlation matrices rotate vectors of initial perturbation in different directions. Although our simulations show that some of that latter difference reflects modeling noise, it seems clear that systemic effects of environmental and genetic changes differ. It is not immediately obvious why this occurs. While the environmental fluctuations do not involve DNA sequence changes, mutations still exert their effects through modifications of RNA, protein, or metabolite levels. One possibility is that the majority of genetic variants in our set persisted in nature long enough to be captured by our survey. In contrast, the environmental perturbations need only be consistent with viability over one generation in relatively favorable laboratory conditions. Therefore, at least some of the genetic correlations are perhaps preserved by selection, possibly for the maintenance of metabolite pools (Clark, 1989). An example of this is the well-known correlation between G6PD and 6PGD. It appears to be favored by selection due to toxicity of an

intermediate, and we see it in our genetic correlation matrix. However, this relationship is absent in the environmental matrix. Selective forces thus might work to modify the intrinsic regulatory architecture, since the genetic correlations are not simply a subset of the environmental ones. Verifying this hypothesis and determining the mechanisms of network modification by natural selection is a fruitful avenue for future inquiry.

To further examine the architecture of co-regulation networks, we turned our attention to the properties of the network nodes, i.e., enzymes and physiological variables. We find that the distributions of degree values among nodes are very close to uniform. This is not surprising, since our networks are highly saturated, with the number of connections exceeding half of all possible links. Thus, no variable has much scope to be unusually highly connected. Nevertheless, we do find that the PGI enzyme participates in more positive associations than expected by chance. While we see no overall relationship between the co-regulation and reaction networks, PGI appears to be a hub in both. It has long been recognized that this enzyme is located at a cross-roads among many pathways (Watt, 1977). We quantified this by constructing a graph representation of intermediary metabolism, and found that indeed PGI has the highest degree in this subnetwork of reactions. Another measure of network centrality, betweenness, is defined as the number of shortest paths between all pairs of nodes that pass through the node under consideration. This parameter is perhaps a better reflection of the number of fluxes that go through a given enzyme. While there are four enzymes in the network of reactions that have higher betweenness than PGI, among the ones we measured this enzyme has by far the highest betweenness. We speculate that stabilizing selection on ratios of metabolite pools requires tight co-regulation of fluxes through several pathways, leading to coordination of PGI activity with rates of many other enzymes.

Weight has a particularly high number of negative correlations. Moreover, females, which are almost twice as heavy as males on average, display lower activities of most enzymes per weight than do males. This likely reflects allometric scaling of some enzyme activities and metabolite levels with weight. While our data cannot address the origin of this behavior, a plausible explanation is that metabolically active tissues are underrepresented in heavier flies, particularly in females (Djawdan *et al*, 1997). Regardless of the exact mechanism, PGI retains its status as a hub in the genetic network after we take into account each enzyme's correlation with weight.

We find evidence of sexual dimorphism in activities of most enzymes. To assess the extent of sex-by-genotype interaction, we calculated cross-sex genetic correlations for each enzyme and physiological variable. These correlations range from zero to slightly above 0.6. In contrast, most cross-sex correlations reported for whole-organism phenotypes, such as morphology or life history, are > 0.6 (Poissant *et al*, 2010). Thus, it appears that one cannot reliably predict line mean ranks in females from the values in males for most enzymes. This is a common interpretation of the more abstract sex-by-genotype interaction terms typically estimated using traditional regression approaches (Lynch and Walsh, 1998). The larger the divergence of a cross-sex genetic correlation from one, the more scope there is for natural selection to evolve sexual

dimorphism. Therefore, it appears that the sex differences we see can be further widened by selection.

Despite extensive differences in mean activities, among-enzyme correlations stay largely the same between males and females. Interestingly, a number of correlations involving cytoplasmic MDH change sign between sexes in both the environmental and genetic network. This enzyme is part of the malate-aspartate shuttle that transports reducing equivalents from the cytosol to the mitochondria, and is also involved in supplying the malic enzyme (ME) with malate for NADPH production (Figure 2). The latter molecule is crucial for fatty acid synthesis. Our results suggest that cMDH fits differently into the metabolic regulatory network in males and females. Perhaps, energy and metabolite requirements for egg production lead to a distinct balance between respiratory generation of ATP and fat deposition in females.

Our results suggest that even very similar correlation matrices transform initial changes of metabolic state in measurably different ways. Thus, if we are to use correlation matrices to predict the outcome of directional system disturbances, we must insure that conditions of the matrix ascertainment and experimental perturbation are the same. In particular, given the differences between the environmental and genetic correlation matrices, it appears that a dietary shift that affects a set of enzymes would have a different system-wide effect than a mutation with similar primary targets.

The approach we presented here integrates direct measurements of metabolic network components (i.e., enzyme activities and metabolite levels), experimental design to separately assess sources of variation, and mathematical tools for modeling variation of and dependence among parameters. The analyses we performed open a path from network discovery to the use of system-level information to predict outcomes of experimental or natural perturbations.

Materials and methods

D. melanogaster lines and rearing

Lines

We selected 92 *D. melanogaster* lines derived from single wild-caught females (isofemale lines), from 5 populations (Beijing, Netherlands, Ithaca (New York), Tasmania, and Zimbabwe), and inbred them for 13 generations by sib mating. We detailed the provenance of the lines in an earlier publication (Greenberg *et al*, 2010).

Crosses

The 92 inbred lines were used to generate two sets of 92 crosses. One set of crosses was between consecutively named lines from the same population (round robin). The other set was performed between lines derived from different populations, with lines from each population participating in crosses with all the other populations (see Figure 1 and Greenberg *et al*, 2010 for details). We reared 5–7-day-old progeny of each cross (including 'selfed' inbred lines) at 25°C and a 12-h light-dark cycle on Cornell's standard medium (1% agar, 8.3% dextrose, 8.3% brewers yeast, 0.0415% phosphoric acid, 0.415% propionic acid). We weighed two replicates of five male or female flies from each block (Figure 1B) using a precision balance (Sartorius CP2P) and placed them in 2.0 ml 96-well plates (ABgene, AB-0661). The plates were frozen at –80°C. Three blocks utilizing this design were prepared for a total of 36 plates. In summary, we sampled 92 lines from 5 populations, performed 272 crosses in a total of 816 blocks and 1632 replicates separately in males and females.

Enzyme assays

Chemicals were obtained from Sigma-Aldrich and Fisher Scientific.

Homogenate preparation

We thawed deep-well plates in a 4°C cold room and suspended our samples in 1 ml cold (–20°C) homogenization buffer (0.01 M KH₂PO₄, 1 mM EDTA, pH 7.4). After adding a Zn-plated BB (Daisy Outdoor Products), we homogenized the samples at 600 Hz for 30 s using a Geno/Grinder 2000 (BT&C/OPS Diagnostics). Following a 5-min spin at 300 g (Eppendorf 5804 centrifuge with DWP-2 rotor used for all centrifugation), we aliquoted the supernatant into ten UV-clear 360 µl 96-well plates (Corning 3635) and five 250 µl 96-well plates (BD Falcon 353912) using a Hydra 96 (model 1029-41-3, Robbins Scientific), 20 µl of supernatant per well.

Mitochondrial extraction

After a 15-s vortex of the homogenization plate and a 5-min spin at 300 g, we transferred 340 µl of homogenate to a 0.5-ml 96-well plates. We then spun these plates at 2250 g for 40 min to pellet mitochondria and aliquoted 20 µl of supernatant into a UV-clear plate to assay cytosolic malate dehydrogenase. We removed 300 µl of supernatant and resuspended the pellets in cold mitochondrial resuspension buffer (0.01 M KH₂PO₄, 1 mM EDTA, 0.5% (w/v) Triton X, pH 7.4) and vortexed thoroughly to resuspend the mitochondrial pellets. After a 5-min incubation at 4°C, we aliquoted the mitochondrial extracts into two UV-clear 96-well plates and four non-UV-clear 96-well plates using a Hydra 96. All plates were stored at –80°C.

Cytosolic enzyme assays

We list the enzymes and metabolites we assayed, the abbreviations we used in the text, the EC number for each enzyme and assay conditions.

Glucose-6-phosphate dehydrogenase (G6PD; EC 1.1.1.49; modified from Clark and Keith, 1989): 3.5 mM glucose-6-phosphate, 0.2 mM NADP, and 18.8 mM MgCl₂ in 20 mM Tris-HCl (pH 8.0), assayed at 340 nm for 20 min.

Phosphoglucosyltransferase (PGM; EC 2.7.5.1; assay described in Stam and Laurie-Ahlberg, 1982): 0.83 mM glucose-1-phosphate, 0.05 mM glucose-1,6-bisphosphate, 0.5 mM NADP, 3.1 U/ml G6PD (Sigma G-8404), and 1 mM MgCl₂ in 20 mM Tris-HCl (pH 8.0), assayed at 340 nm for 8 min.

Phosphoglucose isomerase (PGI; EC 5.3.1.9; assay described in Stam and Laurie-Ahlberg, 1982): 2.3 mM fructose-6-phosphate, 0.38 mM NADP, and 1.37 U/ml G6PD in 20 mM Tris-HCl (pH 8.0), assayed at 340 nm for 8 min.

Hexokinase (HEX; EC 2.7.1.1; assay described in Stam and Laurie-Ahlberg, 1982): 5 mM dextrose, 0.5 mM NADP, 0.36 mM ATP, 0.23 U/ml G6PD, and 0.2 mM MgCl₂ in 20 mM Tris-HCl (pH 8.0), assayed at 340 nm for 10 min.

ME (EC 1.1.1.40; assay modified from Clark and Keith, 1989): 30 mM L-malate, 0.68 mM NADP, and 50 mM MgCl₂ in 250 mM Tris-HCl (pH 7.4), assayed at 340 nm for 10 min.

Alcohol dehydrogenase (ADH; EC 1.1.1.1; assay described in Stam and Laurie-Ahlberg, 1982): two parts of 2 mM EDTA in 10 mM KH₂PO₄ (pH 7.4), one part of 180 mM ethanol, 1.4 mM NAD⁺, and 0.9 mM EDTA in 40 mM glycine (pH 9.5), assayed at 340 nm for 10 min.

Glycerol-3-phosphate dehydrogenase (GPDH; EC 1.1.1.8; assay modified from Clark and Keith, 1989): 0.4 mM dihydroxyacetone phosphate, 0.15 mM NADH, and 5 mM EDTA in 20 mM Tris-HCl (pH 8.0), assayed at 340 nm for 10 min.

6-Phosphogluconate dehydrogenase (6PGD; EC 1.1.1.44; assay described in Clark and Keith, 1989): 0.3 mM 6-phosphogluconate, 0.3 mM NADP, and 18.8 mM MgCl₂ in 20 mM Tris-HCl (pH 8.0), assayed at 340 nm for 20 min.

Total and mitochondrial protein (tPRT and mPRT; modified from Lowry *et al*, 1951): we added 160 µl of 0.025 g/ml Lowry reagent (Sigma L-3540) to each sample. After a 20-min incubation at 20°C, we further added 20 µl of Folin reagent (Sigma F9252). After another 30-min

incubation at 20°C, absorption was followed for 5 min at 610 nm. For total protein, we included 0.04 and 0.004% BSA standards in each plate. For mitochondrial protein, we included 0.004 and 0.0004% BSA standards.

Fatty acid synthase (FAS; EC 2.3.1.85; modified from Nepokroeff *et al.*, 1975): we added 200 µl of a solution containing 0.033 mM acetyl-CoA, 0.1 mM malonyl-CoA, and 0.2 mM NADPH in 20 mM Tris-HCl (pH 8.0) to samples, assayed at 340 nm for 40 min.

Triacylglycerides (TRI; assay modified from the Sigma serum triglyceride kit procedure): first, we added 160 µl of free glycerol reagent (F-6428, resuspended in 80 ml per bottle) to each sample. After a 5-min incubation at 20°C, we followed absorption at 540 nm for 20 min to assay the quantity of free glycerol. Next, we added 20 µl of triglyceride reagent (Sigma T-2449, resuspended in 40 ml per bottle) to each sample and immediately measured absorption at 540 nm for 20 min. Plates were run with a 5-µl glycerol standard solution (Sigma G-7793).

Glycogen (GLG; assay described in Clark and Keith, 1989): we added 200 µl of a solution containing 0.1 U/ml amyloglucosidase (Sigma A-7420), 0.04 mg/ml o-dianisidine dihydrochloride and 10 g/l PGO enzyme preparation (Sigma P-7119) to each sample. Following a 30 min incubation at 37°C, we measured absorption at 450 nm for 5 min. We included a 4 and 46 µg standard in each plate.

Glycogen phosphorylase (GP; EC 2.4.1.1; assay described in Clark and Keith, 1989): we added 200 µl of a solution containing 10 mg/ml glycogen, 0.0012 mM glucose-1,6-bisphosphate, 0.57 mM NADP, 0.24 U/ml PGM (Sigma P-3397), and 0.85 U/ml G6PD in 38 mM triethanolamine to each sample, assayed at 340 nm for 30 min.

Trehalase (TRE; EC 3.2.1.28; assay described in Clark and Keith, 1989): we added 200 µl of a solution containing 0.9 mM trehalose, 1 mM ATP, 0.9 mM NADP, 1.4 U/ml HEX (Sigma), 2.5 U/ml G6PD, and 0.25 mM MgCl₂ in 20 mM Tris-HCl (pH 8.0) to each sample, assayed at 340 nm for 10 min.

Glycogen synthase (GS; EC 2.4.1.11; assay described in Clark and Keith, 1989): we added 20 µl of a solution containing 4.4 mM UDP glucose, 17.9 mM glucose-6-phosphate, and 8.9 mM MgCl₂ in 50 mM Tris-maleate (pH 7.0) to each sample. After a 5-min incubation at 20°C, we incubated the plates at 65°C for 5 min to stop polymerization. Next, we added to the samples 20 µl of a solution containing 21 mM phosphoenolpyruvate, 5.9 U/ml pyruvate kinase, 85 mM KCl in dH₂O. After a 5-min incubation at 37°C, we further added 0.15% w/v 2,4-Dinitrophenylhydrazine in 2 N HCl. A 5-min incubation at 37°C was followed by the addition of 160 µl 2 N NaOH and absorbance was followed for 5 min at 450 nm.

Phosphofructokinase (PFK; EC 5.3.1.9; assay modified from Clark and Keith, 1989): we added 200 µl of a solution containing 1.2 mM fructo-6-phosphate, 1.6 mM ATP, 0.2 mM NADH, 0.66 U/ml aldolase, 4 U/ml triosephosphate isomerase, 0.2 U/ml GPDH, 1 mM MgCl₂, 78 mM KCl, and 0.8 mM KCN in 100 mM Tris-HCl (pH 8.0) to each sample, assayed at 340 nm for 40 min.

Mitochondrial enzymes

Malate dehydrogenase (cytosolic and mitochondrial; cMDH and mMDH; EC 1.1.1.37; assay described in Stam and Laurie-Ahlberg, 1982): we added 200 µl of a solution containing 0.25 mM oxaloacetic acid and 0.34 mM NADH in 20 mM TAPS and 20 mM PIPES buffer (pH 8.0) to each sample, assayed at 340 nm for 10 min.

Succinate dehydrogenase (SDH; EC 1.3.5.1; assay described in Stam and Laurie-Ahlberg (1982)): we added 200 µl of a solution containing 20 mM succinic acid, 2 mM KCN, 72 µM dichloroindophenol (DCIP), and 25 mM HEPES in 50 mM KH₂PO₄ (pH 7.1) to each sample, assayed at 600 nm for 10 min.

FAD-dependent GPDH (GPO; 1.1.5.3; assay described in Stam and Laurie-Ahlberg, 1982): we added 200 µl of a solution containing 73 mM α-glycerol-3-phosphate, 72 µM DCIP, and 25 mM HEPES in 50 mM KH₂PO₄ (pH 7.1) to each sample, assayed at 600 nm for 10 min.

Pyruvate dehydrogenase (PDH; EC 1.2.4.1; assay described in Hinman and Blass, 1981): we added 200 µl of a solution containing 10 mM sodium pyruvate, 2 mM thiamine pyrophosphate, 2.5 mM NAD⁺, 0.1 mM CoA, 0.6 mM iodinitrotetrazolium violet, 6.5 µM phenazine methosulfate, 0.3 mM dithiothreitol, 1 mg/ml BSA, and 1.0 mM MgCl₂ in 50 mM KH₂PO₄ (pH 7.8) to each sample, assayed at 500 nm for 15 min.

Fumarase (FUM; EC 4.2.1.2; assay described in Stam and Laurie-Ahlberg, 1982): we added 200 µl of a solution containing 10 mM DL-malate in 65 mM NaH₂PO₄ (pH 7.8) to each sample, assayed at 250 nm for 30 min.

Data processing and scaling

We performed the enzyme and metabolite assays in 96-well plates. For kinetics, we measured optical densities as the reactions progressed. For end point reactions, we conducted several measurements after assay completion. We used a Spectramax M2 plate reader from Molecular Devices (Sunnyvale, CA) and handled the raw data collection using the manufacturers' proprietary software. We then exported the data as text files for further processing. We wrote custom Perl scripts to parse the data files, and conducted all subsequent analyses in R (R Development Core Team, 2008).

As a first step, we ran standard regressions of optical density against time to estimate V_{max} for each enzyme in each sample. We then examined plots of maximal rate values for a given enzyme separately in males and females. If outliers were present, we examined kinetic plots of individual reactions to determine if an obvious aberration had occurred. If it did, we checked if a similar pattern emerged in any other reactions from the same plate. We manually deleted offending data points only if concordant patterns of aberration happened in more than one well in a plate. Most outlier observations were kept in the data.

After initial data cleaning, we reran the regressions to obtain slope (V_{max}) estimates. However, some reactions saturated and a number accelerated after an initial lag. To capture the linear portion of each slope, we sequentially deleted up to a third of points from either the beginning or the end of a curve (sometimes both). We reran the regressions for each modified curve, and picked the one with highest model R^2 . We recorded which, if any, points were deleted. We saved the resulting V_{max} estimates, their degrees of freedom and standard deviations for data analysis. For end point reactions, we similarly checked for outliers, but since we were not estimating slopes, we did not perform the sequential-deletion procedure.

To measure triglyceride levels, we used a standard procedure that involves first measuring free glycerol, and then repeating the measurement after lipase treatment to assess the amount of glycerol released from triglyceride stores (see previous section for details). However, the first glycerol measurement is influenced by eye pigment and other molecules in the adult fly (Al-Anzi and Zinn, 2010). Its biological relevance is therefore questionable. Hence, we present only the difference between the lipase-treated and non-treated extracts. We stored these two sets of data separately (see the Supplementary information), and calculated the difference in the model.

Raw V_{max} estimates were in OD units per second. Due to differences in reaction rates, conditions and compounds whose concentration change was measured, these raw values varied over half a dozen orders of magnitude from enzyme to enzyme. To avoid numerical problems from having such disparate data types in the same model, we multiplied each V_{max} value and standard deviation for a given enzyme by 10 raised to an appropriate power so that all means of all variable estimates were within an order of magnitude of each other.

A complicating factor in measuring enzyme activities and metabolite levels is the variation in adult body weight. In particular, systematic variation in weight among cross types and populations (not shown), as well as sexes (Figure 5A), makes direct comparisons of raw enzyme activities among these groups difficult to interpret. One would like to correct for the effect of weight. We reasoned that enzyme activities and metabolite levels per unit weight are the relevant physiological variables for the present study. Since we directly measured weight of the flies we used for enzyme assays, we simply divided all estimates of V_{max} and their standard deviations by the weight of the relevant replicate sample. However, this procedure is not perfect in the cases when enzyme activities scale allometrically with weight. We do find some evidence of this in our data: a number of scaled enzyme activities negatively correlate with weight (Figure 3A and B), and females (that are generally almost twice as heavy as males) appear to possess lower activities of most enzymes per unit weight than males (Figure 5A). Furthermore, correlations between weight-corrected physiological parameters do not necessarily imply correlations among non-corrected

variables (Pearson, 1896–1897; Hayes, 2001). However, the other common procedure for accounting for weight—using enzymes as covariates in the model (Packard and Boardman, 1988; Clark, 1989)—is, if anything, more problematic for our data set. The problem is that given the highly structured nature of our experiment, we would have to perform a large number of regressions within each grouping, leaving us with little power to detect (and correct for) relationships between enzyme activities and weight. Furthermore, performing within-group regressions does not correct for systematic between-group differences in weight. Pooling groups is also problematic because it risks creating spurious relationships (an instance of Simpson's paradox; Simpson, 1951). We concluded that estimating per-weight enzyme activities and metabolite levels is an acceptable approach under the circumstances, as long as the interpretations of the results take this feature of the data into account.

Multivariate hierarchical model

We previously described a set of hierarchical Bayesian models that take advantage of our experimental design to analyze enzyme kinetic data (Greenberg *et al*, 2010). Using extensive simulations and some real data, we showed that models employing Student's *t*-distributions among replicates estimate sample parameters (e.g., line means, and generally parameters commonly referred to as 'fixed effects' in traditional mixed-effect model literature; Gelman, 2005; Greenberg *et al*, 2010) with high accuracy and precision. Population parameter (i.e., variances or 'random effects'; Gelman, 2005; Greenberg *et al*, 2010) estimates are generally less precise, but are better than other Bayesian and maximum-likelihood models we tried and are reasonably accurate even in the face of varied assay quality and outlier observations (Greenberg *et al*, 2010). However, these models only deal with one physiological parameter at a time. We have measured 24, and would like to model them jointly because they are not independent. We therefore extended our earlier models to incorporate multivariate observations. Since model structure and basic methodology are directly analogous to the univariate models (Gelman *et al*, 2004), we refer the reader to our previous publication (Greenberg *et al*, 2010) for the details and implementation examples. We provide target distributions and the sampling scheme in Supplementary Text 1. Here, we describe the salient differences.

Rather than use the various available R packages to sample from multivariate distributions, we wrote our own functions. We used the standard algorithm (e.g., Appendix A of Gelman *et al*, 2004) to sample from multivariate normal distributions and the well-established Odell and Feiveson (1966) scheme to construct random Wishart samples. To sample from multivariate Student's *t*, we implemented the parameter-expansion scheme of van Dyk and Meng (2001), which includes the univariate algorithm we used before (Greenberg *et al*, 2010) as a special case.

Since we have more observations than parameters at each level of the hierarchy, it was possible to model inverse-covariance matrices (also known as concentration matrices) using Wishart distributions. In the univariate case flat reference priors on variances performed well. The multivariate extension of the flat variance prior is the Jeffreys prior (Box and Tiao, 1973; Gelman *et al*, 2004). We therefore tested the Jeffreys prior for inverse-covariance matrices. This resulted in very diffuse distributions (spanning several orders of magnitude) for all matrix components, forcing us to consider other options. First, we implemented the Yang and Berger (1994) uninformative reference prior. This prior is not conjugate to the Wishart distribution, and therefore required Metropolis hit-and-run sampling to construct posterior distributions. This implementation resulted in much smaller credible intervals for covariance matrix elements, but frequently exhibited poor mixing, especially after convergence. We therefore turned to conjugate Wishart priors, with diagonal (all off-diagonal elements equal to zero) matrices and one degree of freedom, the smallest possible number that still results in a proper distribution (Gelman *et al*, 2004). Using a diagonal matrix for the prior amounts to pre-supposing no correlations among variables. All diagonal values were the same across variables, but we systematically varied them over five orders of magnitude to test if the choice of prior affected inference. The results presented here are from models where the prior

diagonal elements were equal to one. The values of these elements did not affect distribution spread or median; however, the more extreme values had a slightly adverse effect on mixing. Medians of posterior distributions of covariance matrix elements were very similar to those we estimated with the two reference priors. Choice of prior thus did not affect parameter estimates, but had a pronounced effect on the numerical properties of MCMC samplers. As before (Greenberg *et al*, 2010), we let each cross type in each sex have a different replicate and block covariance matrix. The analyses presented here refer to environmental matrices estimated for within-population crosses. We will detail the effects of inbreeding and between-population crossing elsewhere.

Posterior correlations between some variance components and regression coefficients for cross-type resulted in poor convergence in the univariate model. This prompted us to implement a parameter-expansion scheme (Gelman *et al*, 2004; Greenberg *et al*, 2010). Here, we sampled cross-type regression coefficients jointly for all physiological variables using multivariate normal distributions. This made the parameter-expansion step unnecessary.

Because all priors were conjugate, it was possible to construct a Gibbs sampler (Gelman *et al*, 2004) to estimate posterior distributions of parameters. We ran the sampler for 10 000 iterations of burn-in. In all, 62 500 iterations followed, saving every 50th sample, for 1250 posterior distribution samples per chain. We ran eight chains, recovering a total of 10 000 samples from posterior distributions of all parameters. We monitored convergence and mixing by calculating the Gelman–Rubin diagnostic (Gelman and Rubin, 1992) and the output of the summary function, both implemented in the coda R package (Plummer *et al*, 2009). The summary function provides estimates of chain standard error that are based on treating each Markov chain as a time series. In our experience, failure of this estimate flagged by the function can be used as a first-pass diagnostic of convergence problems. Sometimes, this failure occurs despite acceptable convergence, but we have not seen the converse. The parameter space was too large to manually examine convergence of each variable. Therefore, to further check the output, we broke it up into a set of parameter classes: each type of inverse-covariance matrix was in its own class; cross-type regression coefficients, and population, line and cross means were each in a separate class. We randomly selected 200 parameters from each class and plotted the time series and distribution views of Markov chains representing the selected parameters, as implemented in the plot function of the coda package. We inspected each plot, and confirmed that convergence was complete in each case, and the Markov samples mixed properly.

The initial Markov chains represent the primary variables in our model: the sample parameters, such as cross and line means, and population parameters (covariance matrices). Given these samples, it is possible to construct distributions of any combination of the primary parameters, as long as all values are picked from the same Markov chain step. For example, to estimate narrow-sense heritability we took samples of inverse-covariance matrices for replicates, blocks, crosses, and lines at each saved Markov chain step. By inverting these matrices and taking the diagonal values, we obtained variances of each physiological variable. Narrow-sense heritability is then the ratio of the among-line variance to the sum of all variances (Lynch and Walsh, 1998; Greenberg *et al*, 2010). Because this process is repeated at each step of the Markov chain, we obtain distributions of h^2 depicted in Figure 1C.

Correlation matrix analyses

To estimate correlation matrix distributions, we inverted the concentration matrices and swept out the variances. Performing this operation at each step of the Markov chain results in samples of correlation matrices. We present 95% credible intervals calculated from these individual distributions of each correlation coefficient in Supplementary Tables S2 and S3. We next wanted to know what correlation distributions one should expect when no real relationships exist in the data. We approached this question in two ways. First, we simulated 24 sets of 92 independent normal variables to mimic the sample size for genetic correlations in our data. We calculated the correlation matrix among these 24 independent sets and repeated the

calculation 10 000 times. This simulation thus yielded distributions of correlation coefficients expected for independent random variables. We computed medians of these distributions (analogous to the medians of posterior we estimated for our data) and two-tailed probability that each distribution includes zero. We found that the maximal absolute correlation estimate was 0.004, while the minimal P -value was 0.9649. This suggests that individual correlation P -values smaller than that are unlikely to occur by chance. This simulation does not take into account the complexity of our data, and was used merely to provide a rough idea of what to expect under ideal circumstances.

To gain more precise estimates of null P -value distributions, we turned to data permutations. For example, for genetic correlations, at each step of the Markov chain we randomly reassigned values to line names for every physiological variable (assignments to physiological variables were maintained). The permuted data were then used to sample a 'null' genetic concentration matrix using the same procedure as for the real inverse-covariance matrix. We performed this permutation for each inverse-covariance matrix, and calculated correlations from them as described above. As a result, for each MCMC sample from the posterior distribution, we obtained a set of correlations that we expect to see in a data set identical to ours, but with no real relationships among physiological variables.

To compare two matrices, we vectorized their upper triangles and calculated a correlation coefficient between these vectors (See Supplementary Figure S1). This is a similarity measure related to the Mantel's test (Mantel, 1967). We also used a slight variation of the 'random skewers' method developed by Cheverud (1996). Following the published procedures, we generated vectors with 24 elements by sampling uniform random variables on the [0,1] interval and randomly making them positive or negative. We then normalized the vectors to make their length equal to 1. These are the 'random skewers' that represent numerical perturbations of the system. To compare two matrices, we took a random skewer and multiplied it by each matrix:

$$\mathbf{x} = \mathbf{G}\mathbf{s}; \quad \mathbf{y} = \mathbf{E}\mathbf{s},$$

where \mathbf{s} is the skewer vector, \mathbf{G} and \mathbf{E} are the matrices we are comparing, and \mathbf{x} and \mathbf{y} are the transformed vectors. The angle between the resulting scaled and rotated vectors represents a measure of difference between the two matrices, and can be calculated as

$$\angle(\mathbf{x}, \mathbf{y}) = \arccos\left(\frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}\right),$$

where $\mathbf{x} \cdot \mathbf{y}$ is the dot product between two vectors and $\|\mathbf{x}\|$ and $\|\mathbf{y}\|$ are their lengths. In our scheme, at each step of the Markov chain we generate a random skewer and compare samples from distributions of two matrices. Thus, we extend the original method (Cheverud, 1996) to integrate over matrix distributions as well as directions of change.

When focusing on the degree of matrix similarity, we can calculate the probability that a distribution of the element-wise correlation coefficient includes zero, since values zero or below indicate no similarity. If we are interested in establishing matrix dissimilarity, the situation is not so simple. Naively, we can compute the probability that the element-wise correlation is equal to one. Similarly, we can see whether there is a high chance that the random skewer angle is zero. However, these values are attainable only when all elements of the two matrices are identical. Since we are comparing matrix distributions, the more salient question is: what is the chance that the two matrices come from the same distribution? To test this, we compared random draws from MCMC samples of the same matrix using either the Mantel-type test or random skewers. The resulting statistics reflect values expected among identically distributed matrices, and are listed as 'expectations for matrices drawn from the same distribution' in the text. Since each matrix being compared gets its own expectation, to calculate P -values we compared the data-derived statistics with means between two matrices. Since two matrices cannot be more than identical, in this case all P -values were one-tailed.

To classify probabilistically each correlation as 'present' or 'absent' in the real data set, we treated each MCMC sample from the corresponding null correlation matrix as a distribution. We calculated empirical two-tailed probabilities that each real correlation does not belong to this null distribution. Each real correlation was then stochastically designated as present with this probability, and the sign of each picked correlation was noted (see Supplementary Figure S1 for

an illustration). We repeated this process at each step of the Markov chain and used the resulting presence/absence distributions for subsequent analyses. Because the probabilities of not belonging to a null distribution were calculated independently for each correlation, this method ignores non-independence among correlations. Unfortunately, no efficient computational method exists to take account of such non-independence (Gelman *et al*, 2004). However, since we are merely assigning presence/absence and not correlation values, we expect the error introduced to be minor. Note that while this scheme was incorporated into our Gibbs sampler, we did not use the presence/absence calls when estimating parameters of the model. These calls were only used for secondary analyses. For instance, when determining the rates of retention of correlation presence and sign, we aligned corresponding elements of two matrices (since correlation matrices are symmetric, we only used upper triangles, excluding diagonal elements) and noted the elements that were present and of the same sign. If a correlation was absent in both matrices, then it was not counted. To determine the random expectation for the number of correlation sign identities, we permuted correlation presence/absence calls, keeping the sign designations, and recalculated incidence of identity. In the text and figures, we presented the range of these randomly expected identity (or sign switch) rates across all correlations.

To compare correlation matrices among populations, we employed a slightly different procedure. The inverse-covariance matrices we discussed so far have been population-level parameters (Greenberg *et al*, 2010), analogous to parameters frequently (but not always) termed 'random effects' in quantitative genetics (Lynch and Walsh, 1998; Gelman, 2005; Greenberg *et al*, 2010). Since there are fewer lines in each population than there are physiological variables we measured, we could not model population-specific concentration matrices separately using Wishart distributions. Rather, we deterministically calculated sample correlation matrices among line means, or deviations of replicate values from corresponding block means, at each Markov chain iteration. These matrices were compared with similarly calculated overall sample correlation matrices. The presence/absence assignments were performed as detailed above.

Multiple testing

The focus of Bayesian statistics is parameter estimation rather than hypothesis testing. Therefore, because there are no 'tests', multiple testing is not considered as a problem. For example, rather than declare some individual correlations as significant by using some threshold, we sample from their distributions or make presence/absence calls probabilistically as described above. Nevertheless, we do assign posterior probabilities in some cases to convey our belief about the deviation of our distributions from some biologically meaningful value. In such cases, we compare our model estimates with an analog of a null hypothesis that assumes a fixed, usually biologically less interesting, scenario. For example, when we analyze degree distributions for each enzyme and physiological variable, we subtract the values for a randomly rewired network that should have a uniform degree distribution. Posterior probabilities of each node having more or fewer connections than expected for the rewired network is then more analogous to a false discovery rate than the frequentist P -value. Likewise, when we calculate probabilities of correlation conservation or sign switch between two matrices, we also provide 'null' distributions of P -values. The expected number of real P -values that fall outside the null range is analogous to the true positive rate.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Amar Kelkar for technical assistance. Comments from the Section editor, Dr AL Hufton, and four anonymous reviewers helped to improve the manuscript. This work was supported by NIH grant

DK074136 to AGC and LGH. AJG was supported by an NSF Plant Genome Research Program Grant #1026555 during manuscript preparation.

Author contributions: AGC and LGH conceived project; SRH, AGC, AJG, and LGH designed the experiments; SRH and AJG conducted the experiments; AJG analyzed the data; AJG, LGH and AGC wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Al-Anzi B, Zinn K (2010) Colorimetric measurement of triglycerides cannot provide an accurate measure of stored fat content in *Drosophila*. *PLoS ONE* **5**: e12353
- Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, Schubert LA, Birditt B, Shay T, Goren A, Zhang X, Smith Z, Deering R, McDonald RC, Cabili M, Bernstein BE et al (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326**: 257–263
- Arnold SJ (1992) Constraints on phenotypic evolution. *Am Nat* **140**: S85–S107
- Ayroles JF, Carbone MA, Stone EA, Jordan KW, Lyman RF, Magwire MM, Rollmann SM, Duncan LH, Lawrence F, Anholt RRH, Mackay TFC (2009) Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet* **41**: 299–307
- Barabási AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5**: 101–113
- Berenbaum MR, Favret C, Schuler MA (1996) On defining 'key innovations' in an adaptive radiation: cytochrome P450s and Papilionidae. *Am Nat* **148**: S139–S155
- Bijlsma R (1980) Polymorphism at the *G6pd* and *6Pgd* loci in *Drosophila melanogaster*. IV. Genetic factors modifying enzyme activity. *Biochem Genet* **18**: 699–715
- Box GEP, Tiao GC (1973) *Bayesian Inference in Statistical Analysis*. New York: Wiley Classics
- Capaldi AP, Kaplan T, Liu Y, Habib N, Regev A, Friedman N, O'Shea EK (2008) Structure and function of a transcriptional network activated by the MAPK hog1. *Nat Genet* **40**: 1300–1306
- Chefurka W (1965) Intermediary metabolism of carbohydrates in insects. In *The Physiology of Insecta*, Rockstein M (ed), Vol. 2, pp 581–667. New York: Academic Press
- Cheverud JM (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
- Clark AG (1989) Causes and consequences of variation in energy storage in *Drosophila melanogaster*. *Genetics* **123**: 131–144
- Clark AG, Keith LE (1989) Rapid enzyme kinetic assays of individual *Drosophila* and comparisons of field-caught *D. melanogaster* and *D. simulans*. *Biochem Genet* **27**: 263–277
- Djawdan M, Rose MR, Bradley TJ (1997) Does selection for stress resistance lower metabolic rate? *Ecology* **78**: 828–837
- Eanes WF (1999) Analysis of selection on enzyme polymorphisms. *Annu Rev Ecol Syst* **30**: 301–326
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868
- Fell DA (2005) Metabolic control analysis. In *Systems Biology*, Alberghina L, Westerhoff HV (eds), Vol. 13, pp 69–80 of *Top Curr Genet*. Berlin, Heidelberg: Springer
- Fisher RA (1930) *The Genetical Theory of Natural Selection*. Oxford, UK: Clarendon
- Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* **40**: 35–41
- Gelman A (2005) Analysis of variance: why it is more important than ever. *Ann Stat* **33**: 1–31
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian Data Analysis*, 2nd edn, London: CRC Press
- Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* **7**: 457–472
- Greenberg AJ, Hackett SR, Harshman LG, Clark AG (2010) A hierarchical Bayesian model for a novel sparse partial diallel crossing design. *Genetics* **185**: 361–373
- Haldane JBS (1949) Suggestions as to quantitative measurement of rates of evolution. *Evolution* **3**: 51–56
- Hansen TF, Houle D (2008) Measuring and comparing evolvability and constraint in multivariate characters. *J Evol Biol* **21**: 1201–1219
- Hayes JP (2001) Mass-specific and whole-animal metabolism are not the same concept. *Physiol Biochem Zool* **74**: 147–150
- Hinman LM, Blass JP (1981) An NADH-linked spectrophotometric assay for pyruvate dehydrogenase complex in crude tissue homogenates. *J Biol Chem* **256**: 6583–6586
- Houle D (1992) Comparing evolvability and variability of quantitative traits. *Genetics* **130**: 195–204
- Hughes MB, Lucchesi JC (1977) Genetic rescue of a lethal 'null' activity allele of 6-phosphogluconate dehydrogenase in *Drosophila melanogaster*. *Science* **196**: 1114–1115
- Kacser H, Burns JA (1973) The control of flux. *Symp Soc Exp Biol* **27**: 65–104
- Kitano H (2002) Systems biology: a brief overview. *Science* **295**: 1662–1664
- Kitano H (2004) Biological robustness. *Nat Rev Genet* **5**: 826–837
- Lande R (1980a) The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**: 203–215
- Lande R (1980b) Sexual dimorphism, sexual selection, and adaptation in polygenic characters. *Evolution* **34**: 292–305
- Lande R, Arnold SJ (1983) The measurement of selection on correlated characters. *Evolution* **37**: 1210–1226
- Laurie-Ahlberg CC, Maroni G, Bewley GC, Lucchesi J, Weir BS (1980) Quantitative genetic variation of enzyme activities in natural populations of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **77**: 1073–1077
- Logsdon BA, Mezey J (2010) Gene expression network reconstruction by convex feature selection when incorporating genetic perturbations. *PLoS Comput Biol* **6**: e1001014
- Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the folin phenol reagent. *J Biol Chem* **193**: 265–275
- Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* **27**: 209–220
- McGuigan K (2006) Studying phenotypic evolution using multivariate quantitative genetics. *Mol Ecol* **15**: 883–896
- Misarović MD (1968) Systems theory and biology—view of a theoretician. In: Misarović MD (ed), *Systems Theory and Biology*. New York: Springer-Verlag, pp 59–87
- Nepokroeff CM, Lakshmanan MR, Porter JW, Lowenstein JM (1975) Fatty acid synthase from rat liver. *Methods Enzymol* **35**: 37–44
- Odell PL, Feiveson AH (1966) A numerical procedure to generate a sample covariance matrix. *J Am Stat Assoc* **61**: 199–203
- Packard GC, Boardman TJ (1988) The misuse of ratios, indices, and percentages in ecophysiological research. *Physiol Zool* **61**: 1–9
- Pearson K (1896–1897) Mathematical contributions to the theory of evolution. On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc R Soc Lond* **60**: 489–498
- Pearson K (1903) Mathematical contributions to the theory of evolution. XI. On the influence of natural selection on the variability and correlation of organs. *Phil Trans R Soc Lond A* **200**: 1–66

- Plummer M, Best N, Cowles K, Vines K (2009) *coda: Output analysis and diagnostics for MCMC*
- Poissant J, Wilson AJ, Coltman DW (2010) Sex-specific genetic variance and the evolution of sexual dimorphism: a systematic review of cross-sex genetic correlations. *Evolution* **64**: 97–107
- R Development Core Team (2008) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing
- Rockman MV (2008) Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature* **456**: 738–744
- Sacktor B (1970) Regulation of intermediary metabolism, with special reference to the control mechanisms in insect flight muscle. *Adv Insect Physiol* **7**: 267–347
- Salway JG (2003) *Metabolism at a Glance*. Osney Mead, Oxford: Blackwell Science
- Schadt EE (2005) Exploiting naturally occurring DNA variation and molecular profiling data to dissect disease and drug response traits. *Curr Opin Biotechnol* **16**: 647–654
- Schluter D (1996) Adaptive radiation along genetic lines of least resistance. *Evolution* **50**: 1766–1774
- Simpson EH (1951) The interpretation of interaction in contingency tables. *J R Stat Soc Ser B-Stat Methodol* **13**: 238–241
- Sprague GF, Tatum LA (1942) General vs. specific combining ability in single crosses of corn. *J Am Soc Agron* **34**: 923–932
- Stam LF, Laurie-Ahlberg CC (1982) A semi-automated procedure for the assay of 23 enzymes from *Drosophila melanogaster*. *Insect Biochem* **12**: 537–544
- Turelli M (1988) Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**: 1342–1347
- van Dyk DA, Meng XL (2001) The art of data augmentation. *J Comput Graph Stat* **10**: 1–50
- Wagner A (2001) How to reconstruct a large genetic network from n gene perturbations in fewer than n^2 easy steps. *Bioinformatics* **17**: 1183–1197
- Watt WB (1977) Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of *Colias* butterflies: biochemical and population aspects. *Genetics* **87**: 177–194
- Wilton AN, Laurie-Ahlberg CC, Emigh TH, Curtsinger JW (1982) Naturally occurring enzyme activity variation in *Drosophila melanogaster*. II. Relationships among enzymes. *Genetics* **102**: 207–221
- Yang R, Berger JO (1994) Estimation of a covariance matrix using the reference prior. *Ann Stat* **22**: 1195–1211



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.