Interspecific Divergence of Transcription Networks along Lines of Genetic Variance in Drosophila: Dimensionality, Evolvability, and Constraint

Paolo Innocenti1 and Stephen F. Chenoweth*2

1Department of Ecology and Genetics, Evolutionary Biology Center, Uppsala University, Uppsala, Sweden
2School of Biological Sciences, University of Queensland, St Lucia, Queensland, Australia

*Corresponding author: E-mail: s.chenoweth@uq.edu.au.

Abstract

Change in gene expression is a major facilitator of phenotypic evolution. Understanding the evolutionary potential of gene expression requires taking into account complex systems of regulatory networks, the structure of which could potentially bias evolutionary trajectories. We analyzed the evolutionary potential and divergence of multigene expression in three well-characterized signaling pathways in Drosophila, the mitogen-activated protein kinase (MapK), the Toll, and the insulin receptor/Foxo (InR/Foxo or InR/TOR) pathways in a multivariate quantitative genetic framework. Gene expression data from a natural population of D. melanogaster were used to estimate the genetic variance–covariance matrices (G) for each network. Although most genes within each pathway exhibited significant genetic variance, the number of independent dimensions of multivariate genetic variance was fewer than the number of genes analyzed. However, for expression, the reduction in dimensionality was not as large as seen for other trait types such as morphology. We then tested whether gene expression divergence between D. melanogaster and an additional six species of the Drosophila genus was biased along the major axes of standing variation observed in D. melanogaster. In many cases, divergence was restricted to directions of phenotypic space harboring above average levels of genetic variance in D. melanogaster, indicating that genetic covariances between genes within pathways have biased interspecific divergence.

We tested whether co-expression of genes in both sexes has also biased the pattern of divergence. Including cross-sex genetic covariances increased the degree to which divergence was biased along major axes of genetic variance, suggesting that the co-expression of genes in males and females can generate further constraints on divergence across the Drosophila phylogeny. In contrast to patterns seen for morphological traits in vertebrates, transcriptional constraints do not appear to break down as divergence time between species increases, instead they persist over tens of millions of years of divergence.

Key words: G matrix, expression network, transcriptional covariance, multivariate constraint, factor analytic modeling.

Introduction

It is well known that most phenotypic traits are genetically variable and could potentially undergo evolutionary change (Lewontin 1974; Lynch and Walsh 1998). However, it is also clear that despite this evolutionary potential, traits do not diverge among species or conspecific populations in all possible directions of phenotypic space (Arnold 1992). Instead, divergence is often restricted to specific combinations of trait values, which in many cases coincide with major axes of multitrait standing variance present in contemporary populations (Schluter 1996; Merila and Bjorklund 1999; Chenoweth et al. 2010; Colautti and Barrett 2010; but see McGuigan et al. 2005; Berner et al. 2010). These results indicate that the availability and distribution of genetic variation among correlated traits commonly biases their evolutionary trajectories. Furthermore, such biases are observed between species that diverged millions of years ago, suggesting that these types of evolutionary constraints can persist for significant periods of time (Schluter 1996).

At the molecular level, phenotypic divergence can arise through changes in gene expression and/or structural changes in proteins (King and Wilson 1975; Levine and Tjian 2003). Although debate continues as to the relative importance of specific types of regulatory mutations (Carroll 2000; Hoekstra and Coyne 2007; Stern and Orgogozo 2008; Wagner and Lynch 2008), it is at least clear that changes in the regulation of gene expression are strongly associated with divergence in phenotype (Wray 2007; Stern and Orgogozo 2008). Several studies also report associations between DNA sequence divergence and expression divergence (Crawford and Powers 1989; Ogada et al. 2002; Lwniçzak et al. 2008), and while in some cases expression divergence is likely a response to divergent natural selection (Rifkin et al. 2003; Whitehead and Crawford 2006; Holloway et al. 2007; Wolf et al. 2010), it may also occur through neutral processes such as drift, and be curtailed by stabilizing selection (Rifkin et al. 2003; Lemos et al. 2005; Bedford and Harld 2009; Graze et al. 2012). Interspecific comparisons of sequence and
expression divergence have thus far been the primary tool for inferring constraints on the evolution of gene expression (Nuzhdin et al. 2004; Good et al. 2006). An aspect not considered by these comparative approaches, which consider genes one at a time, is that the expression of one gene is rarely independent of the expression of another (Li et al. 2010). Genes are co-expressed along pathways or within networks, creating varying degrees of genetic covariance between them (Ayroles et al. 2009). Although an often significant fraction of the genes in a genome exhibit genetic variance for transcript abundance (Wayne et al. 2007; Ayroles et al. 2009), genetic correlations between transcripts could nonetheless limit the genetic variance available in some directions of phenotypic space and bias divergence, as is the case for genetically correlated morphology and life history traits (Kirkpatrick 2008; Walsh and Blows 2009; Futuyma 2010). The extent to which gene expression is also subject to evolutionary constraints imposed by a lack of heritable variation for specific combinations of trait values remains poorly understood.

In this article, we set out to test for evolutionary constraints on expression divergence through a novel application of multivariate evolutionary quantitative genetic tools (Lande 1979). Our focus is not on the evolution of network topology, itself an active area of investigation (Wagner 1994; Azevedo et al. 2006; Fierst 2011); instead, we consider the problem of divergence in the expression level of sets of genes within known pathways. Central to our analysis is the genetic variance–covariance matrix, \( G \), a fundamental parameter in evolutionary quantitative genetic models, that summarizes all multitrait axes of genetic variance available for divergence via drift and/or natural selection (Lande 1979). In our case, \( G \) has genetic variances for the expression of each gene along its diagonal and the genetic covariances between different genes off the diagonal. When combined with the vectors of expression divergence between pairs of species, the degree to which divergence has been influenced by genetic covariances can be quantified (Schluter 1996). Applying a recently developed framework that formalizes this approach (Hansen and Houle 2008), we analyzed multiple gene expression in three well-known pathways in \( Drosophila \). We asked the following three questions: 1) What is the potential for multivariate gene expression to diverge from standing variation when compared with classic traits such as morphology; 2) Has the divergence of gene expression in these pathways followed evolutionary trajectories biased along major axes of standing variance; and 3) When genes are co-expressed in males and females, to what extent do cross-sex genetic correlations impose additional constraints on divergence?

To address these questions, we used multivariate mixed-effects models to test for genetic variance in transcript abundance for three networks (mitogen-activated protein kinase [MapK], Toll, and insulin receptor/Foxo [InR/Foxo]) in a natural population of \( D. melanogaster \) (Ayroles et al. 2009). Likelihood ratio tests indicated significant multivariate genetic variance in all three pathways [males: MapK: \( P(\chi^2_4 = 248.84) < 0.0001 \), Toll: \( P(\chi^2_2 = 381.13) < 0.0001 \), InR/Foxo: \( P(\chi^2_2 = 324.2) < 0.0001 \); females: MapK: \( P(\chi^2_2 = 212.6) < 0.0001 \), Toll: \( P(\chi^2_2 = 423.3) < 0.0001 \), InR/Foxo: \( P(\chi^2_2 = 264.5) < 0.0001 \)].

We then performed reduced-rank model comparison (Kirkpatrick and Meyer 2004) to estimate the dimensionality of each male and female genetic variance–covariance matrix, \( G \). Dimensionality varied between different pathways and sexes, but it was never less than half the number of genes. For mapK, dimensionality was lowest with five dimensions in males and only four in females out of a possible eight (table 1 and supplementary fig. S1, Supplementary Material online), indicating that for many directions of transcriptional space there is either no, or very limited genetic variance available for evolution. Surprisingly, the situation was markedly different for the toll pathway where seven dimensions were supported in both sexes. There was a strong sex difference in InR/Foxo with seven dimensions found in males but only four in females. Although we presently lack information on the dimensionality of genetic variance in expression networks, several studies have estimated the dimensionality of \( G \) for morphological and chemical phenotypes (Kirkpatrick and Loefsvold 1992; Hine and Blows 2006; Mcguigan and Blows 2007; Lienonen et al. 2011). In these studies, dimensionality is also invariably lower than the number of traits, with most estimates of \( \text{dim}(G)/k \) spanning from 0.2 to 0.5. A notable exception, and similar to our finding for the toll pathway, is wing shape in \( D. melanogaster \) where there is statistical support for the existence of all possible dimensions (Mezey and Houle 2005). Our analyses suggest that the dimensionality of whole-body transcription is higher than other traits with values of \( \text{dim}(G)/k \) always 0.5 or greater. It would be interesting to assay gene expression in specific tissue types to see

**Results and Discussion**

**Network Dimensionality and Evolvability**

We used multivariate mixed effects models to test for genetic variance in transcript abundance for three networks (mitogen-activated protein kinase [MapK], Toll, and insulin receptor/Foxo [InR/Foxo]) in a natural population of \( D. melanogaster \) (Ayroles et al. 2009). Likelihood ratio tests indicated significant multivariate genetic variance in all three pathways [males: MapK: \( P(\chi^2_4 = 248.84) < 0.0001 \), Toll: \( P(\chi^2_2 = 381.13) < 0.0001 \), InR/Foxo: \( P(\chi^2_2 = 324.2) < 0.0001 \); females: MapK: \( P(\chi^2_2 = 212.6) < 0.0001 \), Toll: \( P(\chi^2_2 = 423.3) < 0.0001 \), InR/Foxo: \( P(\chi^2_2 = 264.5) < 0.0001 \)].

We then performed reduced-rank model comparison (Kirkpatrick and Meyer 2004) to estimate the dimensionality of each male and female genetic variance–covariance matrix, \( G \). Dimensionality varied between different pathways and sexes, but it was never less than half the number of genes. For MapK, dimensionality was lowest with five dimensions in males and only four in females out of a possible eight (table 1 and supplementary fig. S1, Supplementary Material online), indicating that for many directions of transcriptional space there is either no, or very limited genetic variance available for evolution. Surprisingly, the situation was markedly different for the Toll pathway where seven dimensions were supported in both sexes. There was a strong sex difference in InR/Foxo with seven dimensions found in males but only four in females. Although we presently lack information on the dimensionality of genetic variance in expression networks, several studies have estimated the dimensionality of \( G \) for morphological and chemical phenotypes (Kirkpatrick and Loefsvold 1992; Hine and Blows 2006; Mcguigan and Blows 2007; Lienonen et al. 2011). In these studies, dimensionality is also invariably lower than the number of traits, with most estimates of \( \text{dim}(G)/k \) spanning from 0.2 to 0.5. A notable exception, and similar to our finding for the Toll pathway, is wing shape in \( D. melanogaster \) where there is statistical support for the existence of all possible dimensions (Mezey and Houle 2005). Our analyses suggest that the dimensionality of whole-body transcription is higher than other traits with values of \( \text{dim}(G)/k \) always 0.5 or greater. It would be interesting to assay gene expression in specific tissue types to see
whether the apparently elevated dimensionality is a function of whole-body analysis, which could effectively pool genetic variance across the multiple tissues within which a gene is expressed.

There were differences between pathways and sexes in the way that individual genes contributed genetic variance to the matrices. For MapK, the individual gene loadings of the largest eigenvectors of $G_m$ and $G_l$ were not strongly correlated (vector correlation $\lambda_{1G_m}$ vs. $\lambda_{1G_l}$; $r = -0.31$, supplementary tables S4–S6, Supplementary Material online), suggesting that the primary axis of genetic variance is quite different in males and females for this pathway. Within each sex, there were also differences in sign between individual gene loadings, indicating some negative genetic associations between genes in this pathway. The most significant gene contributing genetic variance in males was the epidermal growth factor receptor, Egfr, which was also the largest contributor to the second eigenvector in females. The pattern was different for InR/Foxo, indicating that species have diverged in quite different directions, they are nonetheless more similar than expected by chance.

We tested whether this pattern of divergence between species was associated with the pattern of standing genetic (co)variance in D. melanogaster, applying a framework that quantifies the extent to which multitrait divergence has proceeded in directions of above or below average evolvability (i.e., high or low genetic variance) (Hansen and Houle 2008). With the exception of the Toll pathway for females, unconditional evolvabilities in the directions of realized divergence, $e_B$, were higher overall than their expected values (Males: mean $e_B$, MapK: 143% of $\bar{e}$; Toll: 139%; Foxo: 208% $\bar{e}$, fig. 1, column 1, supplementary tables S13–S21, Supplementary Material online; Females: mean $e_B$, MapK: 225% of $\bar{e}$; Toll: 83%; Foxo: 145% $\bar{e}$, column 2; (table 1), whereas other studies indicate that $n_D$ mostly falls in the interval between 1 and 2, and is never greater than 3 (Kirkpatrick 2008; Hohenlohe and Arnold 2010; Simonsen and Stinchcombe 2010). It appears that for these networks at least, genetic variance tends to be distributed slightly more evenly across the available dimensions than has so far been reported for other phenotypes (supplementary tables S4, S7, and S10, Supplementary Material online). Our intraspecific analyses in D. melanogaster suggested, both in terms of dimensionality and the distribution of genetic variance among these dimensions, that expression networks may exhibit slightly greater evolutionary potential than classic phenotypes such as morphology.
Although these results indicate an overall tendency for the distribution genetic (co)variance at the intraspecific level, summarized by the \( G \) matrix, to bias expression divergence at the interspecific level. There was some variability between the species in the degree to which divergence was biased by \( G \). For example, in females, divergence between \( D. \) melanogaster and \( D. \) yakuba for the InR/Foxo pathway and between \( D. \) melanogaster and \( D. \) virilis in MapK has occurred in regions of lower than average genetic variance.

Previous studies have in some cases reported a decline in the strength of genetic constraint as divergence time increases (Schluter 1996). We tested whether a similar pattern could be detected for gene expression in the form of a negative association between rank divergence time from \( D. \) melanogaster and \( e_p \) and \( c_p \). The species pairs in figure 1 are arranged in order of increasing divergence time from \( D. \) melanogaster; in no case was there evidence for a significant negative correlation in the sex-specific analyses. Only for the combined-sex analysis of \( c_p \) in the Toll pathway did we observe the predicted negative association (combined-sex analysis: \( \rho = -1.0, P = 0.003 \), Spearman’s rank correlation). Overall, these results do not provide strong support for the idea that evolutionary constraints on gene expression decay, to a great degree, over long periods of evolutionary time. An interesting aspect to our findings is that the divergence times between the \( Drosophila \) species considered here are much longer than those between vertebrate taxa.

**FIG. 1.** Evolvabilities (\( e_p \), circles) and conditional evolvabilities (\( c_p \), triangles) along the vectors of divergence between \( Drosophila melanogaster \) and six other \( Drosophila \) species. The four horizontal lines represent: \( e_{\text{max}} \), maximum evolvability; \( \bar{e} \), average evolvability; \( \bar{c} \), average conditional evolvability; \( e_{\text{min}} \), minimum evolvability. The different columns correspond to analyses performed on males (column 1, M subscript), females (column 2, F subscript), and a joint analysis of males and females using \( G_{\text{int}} \) (column 3, MF subscript).
(approximately 4 My) where similar constraints were shown for morphological traits (Schluter 1996). It therefore underscores the extent to which the genetic covariance structure among genes within specific regulatory pathways may form enduring evolutionary constraints.

Either one or both of two evolutionary processes, genetic drift or natural selection, may be responsible for the observed interspecific divergence in gene expression. Because genetic drift in multiple traits is also expected to occur along major axes of genetic variance (Lande 1979), we cannot necessarily exclude it as a factor responsible for our results. However, if directional selection was distributed randomly with respect to the available axes of genetic variance, our findings are consistent with selective constraints because the response to directional selection is known to be heavily biased by the major axes of \( G \) (Lande 1979; Chenoweth et al. 2010).

Applying character process models to the same divergence data set, Bedford and Hartl (2009) found that divergence in single gene expression among these species may have been moderated by stabilizing selection. Other analyses of species pairs of Drosophila (Graze et al. 2012) also suggest a role for stabilizing selection in constraining interspecific divergence in gene expression. Conditional evolvability, \( c_B \), summarizes the available genetic variance for \( \beta \) when there is multivariate stabilizing selection on other linked traits. That is, when stabilizing selection operates not on \( \beta \) itself but on linear combinations of traits that are genetically correlated with \( \beta \). Like unconditional evolvability, \( c_B \), values were consistently higher than their expected values, \( \bar{c} \) (Males: mean \( c_B, \text{MapK}: 167\% \) of \( \bar{c} \); Toll: 121%; Foxo: 239%; Females: mean \( c_B, \text{MapK}: 190\% \) of \( \bar{c} \); Toll: 92%; Foxo: 150%) (fig. 1, columns 1 and 2) although perhaps not to the same extent. Although this is not a test of stabilizing selection per se, these results indicate that when any potential stabilizing selection acting on other genetically correlated traits is accounted for, the extent to which the distribution and availability of multivariate transcriptional genetic variance has influenced divergence between these species does not change. Therefore, stabilizing selection alone may be insufficient to account for the pattern of divergence in these pathways.

**Co-expression of Genes in Separate Sexes Constrains Divergence**

We have shown how genetic covariances between genes represented within three well-known regulatory pathways have biased their evolutionary trajectories during interspecific divergence. When multiple traits are expressed in males and females, additional types of genetic covariances exist—those between males and females for the same and different shared traits. When fitness optima differ between males and females, and genetic covariances between the sexes are strong, the rate of divergence may be slowed (Lande 1980) and the direction of divergence can differ substantially from the direction of selection (Gosden et al. 2012). Thus, any sex-specific evolutionary response depends not only on the availability of genetic variance among genes within a sex, but critically, also that available between the sexes. Genetic covariance for the same and for different genes between the sexes within a network, could therefore impose additional constraints on divergence.

In all pathways we considered, expression has diverged between species in a sex-dependent manner (MANOVA — sex \( \times \) species interaction: MapK: Pillai’s Trace = 1.339, \( F_{48, 336} = 2.01 \), \( P = 2.01 \times 10^{-4} \); Toll: Pillai’s Trace = 1.685, \( F_{48, 336} = 2.73 \), \( P = 8.15 \times 10^{-8} \); Foxo: Pillai’s Trace = 1.377, \( F_{32, 160} = 2.63 \), \( P = 3.97 \times 10^{-5} \)), suggesting that levels of sexual dimorphism have changed during divergence. Our single-sex analyses have thus far excluded the potentially constraining effects of cross-sex cross-gene genetic covariances because the between-sex genetic variance—covariance matrix, \( B \), a sub-matrix of \( G_{nf} \) (eq. 8) is not included (Lande 1980; Meagher 1994; Jensen et al. 2003; Rolff et al. 2005; Steven et al. 2007). If cross-sex genetic covariances have constrained divergence, our analyses should yield different results when the intersexual genetic covariances are included. Specifically, when males and females are analyzed simultaneously in this way, we would expect an increase in the apparent degree of constraint compared with the single-sex results.

We repeated our analyses using both sexes but considering each gene in males and females as a separate trait, resulting in 16- instead of 8-gene matrices. As predicted, we observed an increase in the degree of constraint when analyzing males and females together, the unconditional evolvabilities rising relative to their average expectations (mean \( c_B, \text{MapK}: 218\% \) of \( \bar{c} \); Toll: 191%; Foxo: 269%, fig. 1, column 3, supplementary tables S19–S27, Supplementary Material online). Moreover, in several instances conditional evolvabilities rose well above their expected values. These results support the idea that co-expression in two sexes can generate additional constraints on divergence. Thus, in addition to their well-understood influence on the short-term response to selection (Lande 1980; Lewis et al. 2011; Gosden et al. 2012), the long-term focus of our analyses suggest that intersexual genetic constraints may not completely break down over time and may continually influence the evolution of male and female gene expression.

**Conclusions**

We have investigated dimensionality, evolvability, and constraints on the interspecific divergence of gene expression within three pathways across the Drosophila phylogeny. Specifically, we sought answers to the following three questions: 1) What is the potential for multivariate gene expression to diverge from standing variation when compared with classic traits such as morphology; 2) Have networks followed biased evolutionary trajectories as they have diverged between species; and 3) When genes are co-expressed in males and females, to what extent do cross-sex genetic correlations impose additional constraints on divergence? First, we showed that in D. melanogaster, the dimensions of transcriptional phenotype space available for evolutionary change are often a subset of the total phenotypic space, but correspond to more than half the total number of genes considered. This indicates a potential for evolutionary change that is in broad terms higher, and in the case of the Toll pathway
much higher, than other phenotypic traits. Second, expression has diverged among species in a nonrandom fashion that is associated with directions of higher than average genetic variance in a contemporary *D. melanogaster* population. Thus, the genetic covariance structure between genes at the intraspecific level can constrain interspecific divergence over a timescale of tens of millions of years. Finally, we found evidence that cross-sex genetic covariances between the same and different genes act as additional constraints on the divergence of gene expression.

In our analyses, we were limited by the number of genes we could analyze simultaneously. Future applications to many more genes while desirable will require larger sample sizes than are currently available. Although sequencing costs continue to fall, targeted expression profiling experiments that trade-off gene number for the profiling of larger numbers of genotypes are likely the most practical means to achieve this goal in the short term. It will be interesting also to apply such experimental designs to genes expressed in different tissues and in different developmental contexts. Our analyses indicate that, as has become appreciated for life history and morphological traits (Kirkpatrick 2008; Walsh and Blows 2009; Futuyma 2010), and more recently for metabolite networks (Greenberg et al. 2011), multivariate quantitative genetic tools can facilitate a deeper understanding of the evolutionary constraints acting on naturally high-dimensional phenotypes such as gene expression.

**Materials and Methods**

**Genetic Pathways**

We chose genes expressed within three well-characterized signaling pathway classes for our analysis: the MapK, the Toll, and the insulin receptor/Foxo (InR/Foxo, or InR/TOR) pathways. MapK is a signal-transduction pathway conserved from yeast to mammals (Widmann et al. 1999). In *Drosophila*, it is known to play different roles. For example, in pre-adult life-stages, it regulates dorsal closure, a morphogenetic process occurring during embryogenesis, and eye development (Sawamoto and Okano 1996). In the adult, it is involved in the response to stress (Martin-Blanco 2000) and in the late-stage attenuation of the immune function (Widmann et al. 1999).

The Toll pathway was initially characterized in *Drosophila* for its role in dorsoventral patterning in the embryo (Morisato and Anderson 1994), whereas its central role in the immune response to fungi and bacteria was later identified (Anderson 2000). This pathway consists of three major components: the microbial recognition proteins, the intracellular signaling cascade, which promote the transcription of the last component, the antimicrobial peptides, which act as effectors (Leulier and Lemaitre 2008; Valanne et al. 2011).

The InR/Foxo is a conserved pathway, which regulates organ growth and cell proliferation in the larvae (Puig et al. 2003), and plays a role in determining lifespan (Kapahi et al. 2004) and metabolic levels dependent on nutritional condition (Britton et al. 2002).

In our data sets, we were able to retrieve 10 genes for the MapK pathway (information from the KEGG pathways database, Kanehisa and Goto 2000), 13 genes for the Toll pathway (from Valanne et al. 2011), and 16 genes for the InR/Foxo pathway (from Nuzhdin et al. 2009; Shen and Tower 2010). Details of the genes initially considered for analysis are given in figure 2.

**Estimating Network Dimensionality and Evolvability**

We have used the comprehensive data set of Ayroles et al. (2009), which is based on 40 wild-derived inbred *D. melanogaster* lines sampled from a single population in Raleigh (North Carolina, USA) and is therefore suitable for population level inferences. Raw data were downloaded from ArrayExpress (www.ebi.ac.uk/arrayexpress, accession number E-MEXP-1594) and preprocessed using the vsn algorithm for normalization (Huber et al. 2002). The vsn approach to normalization removes the dependency between mean and variance that can often be associated with microarray data. Following normalization, we summarized the data by taking the mean expression of all probes within each gene for further analysis (Bedford and Hartl 2009). In all the analyses, we worked on mean standardized expression values. Our evolutionary analyses assume a predominantly linear relationship between multiple expression traits. We inspected the reliability of this assumption visually and found general agreement for analyses applied both within and across the sexes.

We took a quantitative genetic approach to investigate evolvability at the transcriptional level in each of these pathways. For a general introduction to G-matrix-based approaches, see McGuigan (2006). Here, we refer to evolvability as the availability of genetic variance (Houle 1992), which dictates a network’s ability to respond to selection from standing variation rather than the way in which new mutations produce variability in network topology (Hansen 2006). First, we estimated the rank (number of significant dimensions or nonzero eigenvalues) of the genetic variance–covariance matrix, G, for each network using factor analytic modeling (Meyer and Kirkpatrick 2005) implemented in the package WOMBAT (Meyer 2007), which fits multivariate mixed effects models via restricted maximum likelihood. Factor analysis has been shown to be a useful tool in recovering groups of co-expressed genes in microarray data (Coffman et al. 2005). Our application of it here allowed us to isolate and test the number of factors at the genetic (line) level within a likelihood framework. Briefly, we fitted full rank models first and then dropped one dimension per run testing for a significant drop in Likelihood where $D = -2\ln(L_{\text{full}}/L_{\text{reduced}})$ and is distributed as a $\chi^2$ with k degrees of freedom where $k$ is number of parameters that differ between the full and reduced models. For each pathway, we estimated the rank of eight-gene within-sex variance–covariance matrices, $G_n$ and $G_r$.

The multivariate random effects model for the estimation of $G$ was as follows:

$$ Y = \mu + \text{line} + \epsilon, $$

(1)
where $Y$ is a matrix of transcript expression values, $\mu$ is the vector of trait means, $\epsilon$ is the error term. The variance–covariance matrix associated with the line effect was taken as an estimate of $G$. In all runs of WOMBAT, we generated starting values using the “–itsum” option with (co)variance component estimates from an exhaustive set of bivariate models.

The multivariate models we applied in the dimensionality tests are sensitive to statistical power and require the estimation of a large number of parameters that increases quadratically as the number of traits, $k$, increases. Estimation at genetic and residual levels in model 1 requires $k(k+1)/2$ (co)variance parameters. As we had 80 observations per sex in the data set, we reduced the number of genes analyzed, from those initially available for analysis, to a consistent set of $k = 8$ per network. This way, we had consistency in the statistical power applied to each network and also avoided the issue of analyzing too many traits for the number of available data points. For the genes that could potentially have been analyzed (i.e., those that were available on the $D. melanogaster$ array, and had orthologs represented on the arrays of an additional six species), we performed screens for univariate genetic variance fitting a univariate version of equation (1) and included 8 genes exhibiting significant genetic variance ($P < 0.05$) in either sex for further analysis. Results are given in supplementary table S1, Supplementary Material online.

Once we had estimated the rank of each $G$ matrix we estimated three additional summary statistics after Kirkpatrick (2008). The effective dimensionality $n_D$ was calculated as follows:

$$n_D = \frac{\text{Tr}(G)}{\lambda_1},$$  

(2)

where $\text{Tr}(G)$ is the sum of all eigenvalues of $G$ or its trace and $\lambda_1$ is the largest eigenvalue. When $n_D$ equals one, there is only one dimension and if $n_D$ equals the number of traits in the analysis the $G$ matrix is spherical, with equal amounts of genetic variance in all directions. We also calculated the maximum evolvability of $G$, which equals the genetic coefficient of variation for largest eigenvalue:

$$e_{\text{max}} = \sqrt{\lambda_1},$$  

(3)

as well as the total genetic variance, $\nu_T$, which is simply the trace of $G$.

Network Divergence along Lines of Genetic Variance

To explore how genetic constraints may have biased the direction of network divergence between species, we used male and female adult expression data from seven species of $Drosophila$ using the data from Zhang et al. (2007). The seven species are $D. melanogaster$, $D. simulans$, $D. yakuba$, $D. ananassae$, $D. pseudoobscura$, $D. virilis$, and $D. mojavensis$. Three species diverged to approximately 10–13 Ma ($D. melanogaster$, $D. simulans$, and $D. yakuba$), whereas the entire group last shared a common ancestor 45–60 Ma (Zhang et al. 2007).
vsn-preprocessed data were obtained from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/), last accessed April 11, 2013, accession number GSE66440). Orthologs of the selected genes were identified using “Flybase melanogaster gene ortholog report” (updated on January 2011). For consistency with the D. melanogaster data set, data were summarized by taking the mean of multiple probes mapping to the same gene. For the InR/Foxo pathway, only five species were retained: D. virilis and D. mojavensis were excluded because orthologs could not be retrieved for these species.

We used a theoretical framework for analyzing evolutionary constraints on multiple phenotypes developed by Hansen and Houle (2008). Briefly, this approach describes multivariate evolutionary divergence between two species in a way that accounts for the available genetic variance described by the genetic variance–covariance matrix, $G$, in a focal species, in our case D. melanogaster. The approach has several key advantages over the classic method of Schluter (1996) of comparing the angle between the axis of phenotypic divergence, $z$, and the leading eigenvector of $G$, $g_{\text{max}}$. In cases where phenotypic divergence has occurred in a direction poorly oriented with $g_{\text{max}}$, but is nonetheless well aligned with other relatively large axes of genetic variance, the Schluter’s approach will underestimate the importance of genetic constraints (Blows and Higgie 2003; Mcguigan et al. 2005; Hansen and Houle 2008; Hansen and Voje 2011). As the approach by Hansen and Houle simultaneously accounts for all axes of genetic variance contained within $G$, it does not suffer from this underestimation issue.

In each pathway, we estimated the pairwise divergence vector, $\beta$, between D. melanogaster and each of the remaining six species using estimates of transcript means for males and females. In our case where pairs of species are considered, $\beta$ is equivalent to the first eigenvector of the variance–covariance matrix among species mean vectors. Note also that in the two species case, this is identical to Schluter’s (1996) $z$. To complement our sex-specific and sex-combined approaches to the estimation of $G$ and its dimensionality, we estimated divergence vectors for males and females separately, $\beta_m$, $\beta_f$, as well as combined $\beta_{mf}$. Two key metrics were then calculated from estimates of $G$ and $\beta$. First, the unconditional evolvability, $e_\beta$, describes the length of the vector of divergence within the space of $G$ scaled by the length of $\beta$:

$$e_\beta = \frac{\beta' G \beta}{|\beta|^2}.$$  

where $'$ denotes matrix transposition and $|\beta|$ is the norm of $\beta$.

Although $e_\beta$ describes the response to directional selection, the conditional evolvability, $c_\beta$, describes the evolvability when there is multivariate stabilizing selection on other trait combinations that are genetically correlated with $\beta$:

$$c_\beta = (\beta' G^{-1} \beta)^{-1}.$$  

For each evolvability metric, we also estimated their expected values, under the null expectation that network divergence has occurred at random among species where $k$ equals the number of genes analyzed.

$$\tilde{e} = \sum_{i=1}^{k} \frac{\lambda_i}{k}$$  

$$\tilde{c} \approx H(\lambda) \left( 1 + \frac{2I[\lambda]}{k + 2} \right)$$

where $H[\lambda]$ is the harmonic mean $H[\lambda] = \sum_\lambda \frac{1}{\lambda}$ and $I[\lambda]$ denotes the mean standardized variance $I[\lambda] = \frac{\text{var}(x)}{\sum_\lambda x^2}$ (Hansen and Houle 2008). Example R-code and a worked example are supplied for our evolvability analyses in supplementary files S2–S4, Supplementary Material online.

Sex-Specific Constraints on Network Divergence

Finally, we wanted to examine the extent to which co-expression in the two sexes had formed additional constraints on divergence. For these analyses, we estimated the within-plus-between-sex $G$ matrix, $G_{mf}$ (Lande 1980) for each pathway:

$$G_{mf} = \begin{pmatrix} G_m & B \\ B' & G_f \end{pmatrix}.$$  

In our case, $G_{mf}$ represents a matrix that is twice the size of $G_m$ or $G_f$ with 16 traits instead of 8, because the same gene is coded as a separate “trait” depending upon whether it is expressed in males or females. $G_{mf}$ contains Lande’s $B$ matrix of cross-sex and cross-sex-cross trait genetic covariances (Lande 1980), which are known to form microevolutionary constraints on the evolution of sexual dimorphism (Lande 1980; Gosden et al. 2012). A two-trait example of $G_{mf}$ highlighting individual (co)variance components is as follows.

$$G_{mf} = \begin{pmatrix} \sigma_{m_1,m_1} & \sigma_{m_1,m_2} & \sigma_{f_1,m_1} & \sigma_{f_1,m_2} \\ \sigma_{m_2,m_1} & \sigma_{m_2,m_2} & \sigma_{f_2,m_1} & \sigma_{f_2,m_2} \\ \sigma_{m_1,f_1} & \sigma_{m_2,f_1} & \sigma_{f_1,f_1} & \sigma_{f_2,f_1} \\ \sigma_{m_1,f_2} & \sigma_{m_2,f_2} & \sigma_{f_1,f_2} & \sigma_{f_2,f_2} \end{pmatrix}.$$  

We then applied the framework of Hansen and Houle using these larger matrices. Analyzing the data in this way provided us with a contrast between analyses that either included or excluded cross-sex genetic covariances.

Supplementary Material

Supplementary tables S1–S27 and figure S1 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank S.L. Allen, A.G. Clark, and T.P. Gosden for comments and discussion. They also thank L. McIntyre and three anonymous reviewers for thoughtful comments that improved the manuscript. Funding for this work was provided by an Australian Research Fellowship awarded to S.F.C. from the Australian Research Council. This study is
based on published data sets; the authors acknowledge the authors of the original papers for collecting, and making publicly available, such data.

References


