Evolutionary Constraint and Adaptation in the Metabolic Network of *Drosophila*

Anthony J. Greenberg, Sarah R. Stockwell, and Andrew G. Clark

Department of Molecular Biology and Genetics, Cornell University

Organisms must carefully control their metabolism in order to survive. On the other hand, enzymes must adapt in response to evolutionary pressures on the pathways in which they are embedded. Taking advantage of the newly available whole-genome sequences of 12 *Drosophila* species, we examined how protein function and metabolic network architecture influence rates of enzyme evolution. We found that despite high overall constraint, there were significant differences in rates of amino acid substitution among functional classes of enzymes. This heterogeneity arises because proteins involved in the metabolism of foreign compounds evolve relatively rapidly, whereas enzymes that act in “core” metabolism exhibit much slower rates of amino acid replacement, suggesting strong selective constraint. Network architecture also influences enzymes’ rates of amino acid replacement. In particular, enzymes that share metabolites with many other enzymes are relatively constrained, although apparently not because they are more likely to be essential. Our analyses suggest that this pattern is driven by strong constraint of enzymes acting at branch points in metabolic pathways. We conclude that metabolic network architecture and enzyme function separately affect enzyme evolution rates.

Introduction

Metabolism lies at the core of organismal survival. It is required for fundamental and highly conserved processes, such as energy generation. Yet, changes in metabolic function accompany adaptation to novel environments (e.g., Berenbaum et al. 1996; Jones 2005). The genome sequences of 12 *Drosophila* species provide an ideal opportunity to tease apart the opposing forces of conservation and adaptability in metabolic evolution.

Enzymes do not act in isolation. They are organized into a network, where enzymes are connected by the compounds they metabolize (Jeong et al. 2000; Wagner and Fell 2001; Stelling et al. 2002; Tanaka 2005). This network consists of modules, which loosely correspond to the traditionally recognized metabolic pathways (Schuster et al. 2000, 2002; Ravasz et al. 2002; Holme et al. 2003). An enzyme’s position in the network helps determine how dramatically a change in its activity will affect flux (rate of metabolite production) through pathways (Kacser and Burns 1973; Stephanopoulos et al. 1998; Stelling et al. 2002). This influence is measured by control coefficients, which are high for enzymes with strong influence over flux (Kacser and Burns 1973). Metabolic control theory predicts that enzymes at branch points have higher control coefficients than those in linear pathways and that enzymes catalyzing irreversible reactions will have more influence on flux than reversible reaction enzymes (Kacser and Burns 1973; Heinrich and Rapoport 1974).

Changes in the activity of enzymes with low control coefficients should be nearly neutral, especially in pathways that perform nonessential functions. Thus, if purifying selection is the dominant force, genes coding for such enzymes should tolerate more amino acid-changing mutations over the course of evolution than genes coding for enzymes with high control coefficients (Wilson et al. 1977; Rausher et al. 1999). Consistent with this prediction, highly connected enzymes evolve slowly in yeast (Vitkup et al. 2006), although perhaps not in bacteria (Hahn et al. 2004). The relative importance of positive and purifying selection is still unclear, however. Recent estimates for *Drosophila* suggest that at least a third, and possibly most, amino acid differences between closely related species are the result of the action of positive selection (Smith and Eyre-Walker 2002; Sawyer et al. 2003; Shapiro et al. 2007). Any analysis of the influence of metabolic network architecture on enzyme evolution patterns must therefore discriminate between genes that show signs of adaptive evolution and those that do not.

Other aspects of network architecture may also modulate an enzyme’s influence over metabolic function. The number of connections an enzyme has to other enzymes via shared metabolites (“degree”), the number of pathways it participates in, the number of reactions it catalyzes, and the extent to which it serves as a conduit of information between modules/pathways (“betweenness”) may all affect how sensitive an organism’s physiology is to changes in a given enzyme’s activity.

We set out to test comprehensively the effects of enzyme function and metabolic network architecture on enzyme evolution using the newly available genomic sequence from 12 *Drosophila* species (*Drosophila* 12 Genomes Consortium 2007). We found that enzymes involved in metabolizing xenobiotic (foreign) compounds evolve significantly faster than average at the amino acid level. Moreover, almost all enzymes involved in this process also participate in other pathways and significantly affect median evolution rates for those pathways. Of the network architecture parameters, only enzyme degree is significantly correlated with rates of protein evolution: highly connected enzymes are relatively constrained, regardless of whether the connections are between or within pathways. Such enzymes are not more likely to be essential, however, and have average rates of adaptive evolution. We conclude that metabolic network architecture has a measurable impact on enzyme evolution rates that is independent of the influence of enzyme function.

Methods

An expanded version of this section is available Supplementary Material online.
Genome Sequence and Evolutionary Rate Estimation

We used maximum likelihood estimates of rates of amino acid (d_3S) and silent (d_S) substitutions as well as the rate of amino acid change corrected for silent site divergence (ω = d_3S/d_S) for each gene. These estimates were calculated for six species most closely related to Drosophila melanogaster (D. melanogaster, Drosophila simulans, Drosophila sechellia, Drosophila yakuba, Drosophila erecta, and Drosophila ananassae) by Larracuente et al. (2008), based on the genome assemblies and gene models provided by the 12 genomes sequencing group (Drosophila 12 Genomes Consortium 2007). In addition to parameter estimation, Larracuente et al. (2008) performed tests of positive selection (model M8 against model M7). We classified genes as evolving under positive selection if they satisfied the q value (Storey and Tibshirani 2003) cutoff of 0.1 (i.e., the expected fraction of true positives is 90%). For statistical tests, we coded positively selected genes as “1,” and the rest as “0.” To estimate rates of gene duplication, we took advantage of the fuzzy reciprocal BLAST of all the rest as “0.” To estimate rates of gene duplication, we took advantage of the fuzzy reciprocal BLAST of all the remaining 11 genomes (the five mentioned above plus Drosophila pseudoobscura, Drosophila persimilis, Drosophila willistoni, Drosophila mojavensis, Drosophila virilis, and Drosophila grimshawi; Drosophila 12 Genomes Consortium 2007). Some of the genes had homologous clusters within species that could not be assigned an unambiguous ortholog in other species (for details, see Drosophila 12 Genomes Consortium 2007).

If an enzyme was coded by any such genes, it was marked as having evidence of duplication and was coded as “1” for statistical tests. Otherwise, it was coded as “0.” Unless otherwise indicated, we considered a gene to be duplicated if there was evidence of multiple copies in any of the 12 species.

Results

Genes Coding for Enzymes Evolve Slowly

Metabolic function is essential for survival, and many metabolic processes are highly conserved from bacteria to mammals. We thus wanted to test whether genes coding for enzymes are more constrained than other genes. To accomplish this, we compared the distributions of the number of amino acid changes per silent substitution (ω = d_3S/d_S) estimated for the whole phylogeny of six species of the D. melanogaster family, see Methods; Larracuente et al. 2008) for the two groups of genes. Low values of ω indicate that few amino acid-altering mutations have been fixed during evolution, compared with the number of silent (non-amino acid changing and thus nearly neutral) mutations fixed in the same gene. A gene with a low value of ω has been constrained from changing its amino acid sequence and thus codes for a protein that is not free to evolve.

Enzymes are indeed relatively constrained: the median ω of enzymes is 0.045, whereas it is 0.066 for nonenzymes (Wilcoxon test P = 5.7 × 10^{-24}). However, genes involved in metabolic function are also slightly more likely to be essential (17% vs. 12% for the rest; Fisher’s exact test P = 0.0157). We wanted to know whether enzymes are constrained simply because they are more essential. We divided genes into four classes: “essential,” “viable,” “no information,” and “no alleles” (for details, see Methods).
We then compared enzyme-coding to nonenzyme-coding genes within each class. We found that metabolic genes were still more constrained than other genes (fig. 1). The "enzyme" effect was highly significant (two-way ANOVA: $F = 67.0$, degree of freedom [df] = 1, $P = 3.1 \times 10^{-16}$; permutation $P \leq 0.0001$), whereas no interaction between the "enzyme" and "essentiality" terms was detectable ($F = 1.3$, df = 3, $P = 0.27$). Furthermore, the difference between enzymes and nonenzymes within each class was highly significant (Wilcoxon test $P$ values ranging from $5.4 \times 10^{-5}$ to $7.5 \times 10^{-5}$).

Smaller $\omega$ can result from either a decrease in amino acid changes ($d_N$) or an increase in $d_S$. In this case, it is due to a decrease in $d_N$: median $d_N$ for enzymes is 0.077 versus 0.114 for nonenzymes (Wilcoxon test $P = 3.5 \times 10^{-20}$), whereas the values for $d_S$ were 1.796 and 1.777 (Wilcoxon test $P = 0.36$).

The decision to classify a gene as coding for a particular enzyme is at least partially based on its homology to enzyme-coding genes in other species. Apparent high constraint of enzymes could thus be due to ascertainment bias. To control for this, we repeated the analysis with the subset of *Drosophila* genes that have a human homolog (Blast E value $\leq 10^{-10}$), thereby eliminating from the nonenzyme group the fast-evolving genes that could inflate the group’s score. We still found that enzymes are more constrained than nonenzymes. Median $\omega$ for enzyme-encoding genes was 0.043, in contrast with 0.052 for the rest of the genes (Wilcoxon test $P = 9.9 \times 10^{-5}$; two-way ANOVA enzyme effect $F = 15.4$, df = 1, $P = 9.0 \times 10^{-5}$; permutation $P = 0.0005$). We conclude that ascertainment bias is unlikely to fully explain our observations.

**Xenobiotic-Detoxifying Enzymes Evolve Relatively Quickly**

Despite overall high levels of constraint, enzyme-encoding genes vary considerably in their ability to accommodate amino acid substitutions ($\omega$ ranges from 0.0001 to 0.2833). To test if enzyme function affects evolutionary rates, we grouped enzymes into 11 functional categories (“pathway groups”) according to the KEGG classification (Kanehisa and Goto 2000). Each category encompasses a number of pathways with similar functions. Because some pathways contain few enzymes and assignment of enzymes to functionally related pathways is prone to error, grouping pathways should increase statistical power and reduce annotation errors.

We compared median rates of amino acid change of genes in different pathway groups. As expected (fig. 2), median $\omega$ for enzymes involved in metabolism of xenobiotic compounds is the highest of all the groups. This is true for the phylogeny-wide estimate of $d_N/d_S$ as well as for individual gene classes.
species (see supplementary fig. S1, Supplementary Material online).

Because about one-third of all enzymes belong to more than one category, the distributions of \( \omega \) for each group are not independent, and normal statistical tests, such as Kruskal–Wallis, are not applicable. We therefore developed a permutation test that accounts for such nonindependence (see Methods). Using this approach, we determined that amino acid substitution rate heterogeneity among pathway groups is indeed significant (Kruskal–Wallis \( \chi^2 = 22.3 \), permutation \( P = 0.0078 \)). This is at least partly due to relatively fast evolution of xenobiotic detoxification genes (median \( \omega = 0.05 \) for the xenobiotic group vs. \( \omega = 0.04 \) overall, two-tailed permutation \( P = 0.0110 \), and these results are robust to moderate levels of random annotation error (see Methods; supplementary fig. S2, Supplementary Material online). This prompted us to further investigate the effect of this group.

Almost all the proteins in the fast-evolving xenobiotics pathway group act in other processes. How strongly are other groups’ median amino acid substitution rates influenced by genes that detoxify foreign compounds? We found that there is a strong correlation between the median \( \omega \) for a pathway group and the fraction of the category’s genes that it shares with the xenobiotics group (Spearman’s \( \rho = 0.76, P = 0.0108 \)). When we excluded the xenobiotics group from consideration but retained its genes in other pathways, we still saw significant heterogeneity in \( \omega \) among pathway groups (Kruskal–Wallis \( \chi^2 = 17.1654 \), permutation \( P = 0.0045 \)). In contrast, when we eliminated all the genes that belong to this category, we found no significant differences among groups (Kruskal–Wallis \( \chi^2 = 11.0 \), permutation \( P = 0.2167 \)). Heterogeneity in amino acid substitution rates among pathway groups is thus entirely due to the genes that detoxify foreign compounds.

These observations suggest that interdependence among pathways may hinder optimization of metabolic function achievable through amino acid change. One way to alleviate this problem would be through gene duplication (Lynch and Force 2000). We see some evidence that this occurs. No pathway group significantly deviates from the data set average in the fraction of its enzymes coded for by genes duplicated in at least one of the 12 Drosophila genomes we analyzed (for details, see Methods; fig. 3B). Nevertheless, there is a significant positive correlation between the fraction of enzymes in a pathway group that belong to more than one category and the fraction encoded by duplicated genes (Spearman’s \( \rho = 0.673, P = 0.0268 \)). This result suggests that although some paralogs may act only in one pathway, all members of a cluster are annotated as performing each function.

Network Parameters and Constraint

The results presented so far suggest that enzyme function affects how quickly enzyme-coding genes evolve. We wished to know if enzyme evolution is also influenced by metabolic network architecture. To investigate this question comprehensively, we examined several characteristics of the metabolic network topology. In this model of the network, each node represents an enzyme, and the nodes are connected by the metabolites with which the enzymes interact (fig. 4B). This “enzyme-centered” representation ensures that each enzyme appears only once in the network, although metabolites may appear multiple times. We calculated partial correlation coefficients to estimate the associations between each pair of network parameters, while controlling for all others. Although this approach has some shortcomings (Drummond et al. 2006), it is the only method that allows us to determine relationships among all the variables (for discussion, see supplementary methods, Supplementary Material online). We used a permutation test to estimate all \( P \) values (see supplementary methods, Supplementary Material online) and corrected for multiple tests by controlling the FDR (Benjamini and Hochberg 1995) at 5%.

Reversibility of Reactions

Metabolic control theory (Kacser and Burns 1973; Heinrich and Rapoport 1974) suggests that enzymes catalyzing reversible reactions should exert little control over flux. We might therefore expect genes coding for such
enzymes to be less essential and hence more permissive of amino acid changes. In our correlation framework, this would produce a negative relationship between essentiality and reversibility and a positive one between $x$ and reversibility. We do indeed see a negative correlation between essentiality and reversibility (Spearman’s partial $q = -0.212$, permutation $P = 0.0104$; fig. 5B). However, this does not translate into a relationship between $x$ and reversibility (Spearman’s partial $q = -0.077$, permutation $P = 0.2704$).

Measures of Pleiotropy: Number of Pathways and Reactions Catalyzed

Enzymes that participate in many pathways or catalyze many reactions may be under increased constraint because any change could affect several metabolic functions. We first examined the number of pathways in which a gene’s product participates and found no correlation with $\omega$ (Spearman’s partial $\rho = 0.018$, permutation $P = 0.7840$). Xenobiotic detoxification enzymes could obscure a correlation because they are especially likely to appear in many pathways and also tend to have high values of $\omega$. To test this, we repeated our analysis without these enzymes but obtained the same result (Spearman’s partial $\rho = 0.026$, permutation $P = 0.7360$).

We then examined the effect of the number of reactions an enzyme catalyzes. We saw a slight positive correlation between the number of reactions catalyzed and $\omega$, but it was not quite significant even without a multiple test correction (Spearman’s partial $\rho = 0.126$, permutation $P = 0.0560$; fig. 5A). Our measures of pleiotropy thus do not appear to influence levels of evolutionary constraint.

Network Topology Measures: Betweenness and Degree

Engineering principles suggest that evolving modules would change internally but maintain consistent interfaces with the rest of the network so that communication between modules would be uninterrupted (Csete and Doyle 2002). If this is the case for the metabolic network, we should expect enzymes that form between-module bridges to be under increased constraint. Such nodes are considered to have high betweenness (Freeman 1977; Girvan and Newman 2002). The shortest-path betweenness (Newman and Girvan 2004) of an enzyme $i$ is defined as the number of the shortest paths between all other pairs of enzymes in the network that pass through $i$.

Fig. 4.—Illustration of the network representations we used in this study. (A) A connectivity pattern showing both enzymes and metabolites. (B) The enzyme-centered network derived from (A). To analyze the enzymes’ topological parameters and relate them to $\omega$, we redraw the network so that nodes represent enzymes. An edge between enzymes indicates that they share a metabolite in (A). Each enzyme appears only once. Note that high betweenness does not necessarily imply high degree or vice versa. Enzyme E4’s role as the sole conduit between the branching parts of the network gives it high betweenness, but its degree is only 2. (C) The metabolite-centered network derived from (A). To measure the mean metabolite degree of each enzyme, we transform the network in (A) so that nodes represent metabolites. Each metabolite appears only once in this version of the network, and its degree is the number of other metabolites from which it is only one reaction away. (D and E) Enzymes with the same mean metabolite degree but different distributions of connections.
High-betweenness enzymes play an important role in the flow of biomass through the network because they often act as "bottlenecks" (Girvan and Newman 2002; Wunderlich and Mirny 2006; Liu et al. 2007; Yu et al. 2007).

Enzymes that connect pathways do have relatively high betweenness (mean betweenness for pathway-connecting enzymes is 1027.1, whereas for enzymes with no between-pathway connections, it is 290.2; Wilcoxon test $P < 1.2 	imes 10^{-10}$), supporting the claims that the traditionally recognized pathways appear to correspond to modules in the metabolic network (Schuster et al. 2000, 2002; Ravasz et al. 2002; Holme et al. 2003). However, we see no correlation between $\omega$ and betweenness (Spearman’s partial $\rho = 0.028$, permutation $P = 0.6974$; fig. 5A).

Another indication of an enzyme’s importance in the network topology is its degree, defined as the number of other enzymes that share an edge with it. In the enzyme-centered network used here, where edges represent metabolites, an enzyme’s degree is the number of other enzymes to which it is connected through common metabolites (fig. 4B). Note that this is not necessarily equal to the number of reactions catalyzed (see supplementary methods, Supplementary Material online). We found a significant negative correlation between $\omega$ and enzyme degree (fig. 5A) after correcting for multiple tests (Spearman’s partial $\rho = -0.240$, permutation $P = 0.0020$). This result appears to be consistent across multiple Drosophila species and is robust to alternate statistical methods (see supplementary fig. S1 and methods, Supplementary Material online), and the correlation persists if we exclude xenobiotic-detoxifying enzymes from the data set (Spearman’s partial $\rho = -0.212$, permutation $P = 0.0070$).

A potential confounding factor in interpreting this result is gene expression. Highly expressed genes are
relatively constrained, and relationships between \( \omega \) and various variables often disappear once expression levels are accounted for (Drummond et al. 2006 and Larracuente et al. 2008). As a measure of gene expression, we used a principal component that includes whole-fly mRNA levels from FlyAtlas (Chintapalli et al. 2007) and a measure of codon bias (frequency of preferred codons; for details, see Larracuente et al. 2008). The association between degree and \( \omega \) remained unchanged (Spearman’s partial \( p = -0.210 \), permutation \( P = 0.0100 \)).

The relationship between \( \omega \) and degree is due to a deficit of amino acid changes in high-degree enzymes; the correlation of \( d_N \) with degree is significant (Spearman’s partial \( p = -0.235 \), permutation \( P = 0.0034 \), whereas that for \( d_S \) is not \( (\rho = -0.041 \), permutation \( P = 0.5542 \)). This result also strongly indicates that selection on silent sites does not compromise the estimates of \( \omega \) enough to affect our results.

In protein–protein interaction networks, highly connected proteins are relatively constrained only if the connections are inside modules (Fraser 2005). If the same pattern holds for metabolic networks, we would expect excess constraint only for enzymes with many connections within their own pathways. We therefore partitioned the connections into within-pathway and between-pathway links and repeated the partial correlation analysis to assess their independent association with \( \omega \). We found that both kinds of connections constrain enzyme evolution: Spearman’s partial \( \rho = -0.182 \) (permutation \( P = 0.0090 \)) for connections within pathways and \( \rho = -0.184 \) (permutation \( P = 0.0100 \)) for links between them.

Why are highly connected enzymes relatively constrained? It is not because of their central position in the network, because betweenness is a measure of centrality (Freeman 1977) and we find that it has no measurable effect on constraint. Perhaps, high-degree enzymes interact with a few metabolites that are each involved in many reactions (branch point metabolites), providing many neighbors in the enzyme-centered network. Alternatively, a high-degree enzyme may interact with many metabolites that participate in a few reactions each. To distinguish between these topologies, consider a network representation where nodes are metabolites instead of enzymes (and edges represent enzymes/reactions) (fig. 4C). Here, nodes with many edges (high “metabolite degree”) are branch point metabolites, and the enzymes on those edges are branch point enzymes. We can calculate the degree of each metabolite in this “metabolite-centered” network and average the degrees of all substrates and products for a given enzyme. We call this measure “mean metabolite degree,” and it should be high for branch point enzymes.

If we substitute mean metabolite degree for enzyme degree in our partial correlation analysis, the relationship with \( \omega \) remains essentially the same (Spearman’s partial \( \rho = -0.224 \), permutation \( P = 0.0020 \), fig. 5A). This pattern persists if products or substrates are considered separately, and if the skewness (illustrated in fig. 4D and E) of the metabolite degree distribution for each enzyme is accounted for (see “Network statistics” in supplementary methods, Supplementary Material online; examples of skewed topologies for enzymes with high mean metabolite degree can be found in supplementary fig. S5A and B, Supplementary Material online). Enzyme degree and mean metabolite degree are highly correlated (Spearman’s partial \( \rho = 0.749 \), permutation \( P \leq 0.0002 \), although only one protein is among the top five enzymes for both measures (supplementary figs. S4 and S5, Supplementary Material online). After including both the enzymes’ number of connections and their mean metabolite degree in the analysis, we see no independent relationship between \( d_S / d_S \) and either variable (Spearman’s partial \( \rho = -0.108 \), permutation \( P = 0.1376 \) between \( \omega \) and degree; Spearman’s partial \( \rho = -0.066 \), permutation \( P = 0.3584 \) between \( \omega \) and mean metabolite degree), suggesting that \( \omega \) is correlated with a parameter that is equally well described by either measure. To test this, we constructed principal components (Jolliffe 1986) from mean metabolite degree and enzyme degree and used them in our partial correlation analysis. As expected, we find that the PC that includes positive contributions from both parameters is negatively correlated with \( d_S / d_S \) (Spearman’s partial \( \rho = -0.164 \), permutation \( P = 0.0356 \)), whereas the PC that represents a contrast between mean metabolite degree and enzyme degree is not (Spearman’s partial \( \rho = 0.030 \), permutation \( P = 0.6800 \)). We conclude that high-degree enzymes are those interacting with high-degree metabolites and thus are the branch point enzymes predicted to have high control coefficients (Kacser and Burns 1973).

Relatively high constraint of enzymes with many connections should reflect their importance to network function. Therefore, one might expect that such enzymes are more likely than average to be encoded by essential genes. This does not appear to be the case (fig 5B). Although essentiality is negatively correlated with reversibility, there is no significant relationship with either enzyme degree (Spearman’s partial \( \rho = 0.129 \), permutation \( P = 0.1198 \) or mean metabolite degree (Spearman’s partial \( \rho = -0.066 \), permutation \( P = 0.4322 \); the same results are obtained if each measure is analyzed separately). This result could be explained if genes coding for highly connected enzymes were preferentially duplicated, and duplicated genes were in turn less likely to be essential due to functional redundancy. Neither proposition appears to be true for this data set. Enzyme-coding genes with duplicates in D. melanogaster (for details, see Methods) are, if anything, slightly more likely to be essential than nonduplicated genes (50.0% of duplicated genes are essential vs. 47.9% of nonduplicated genes, Fisher’s exact test \( P = 1 \)). Moreover, highly connected enzymes are less likely to be duplicated than average (fig 5C; although the results shown are for duplications in any of the species, they are virtually identical if only D. melanogaster duplications are considered).

Enzymes under Positive Selection

To identify enzymes that have evolved under positive selection in the Drosophila melanogaster group, we used the likelihood ratio test of the codeml model M8 versus M7 (see Methods; Yang 1997; Larracuente et al. 2008). Because only a small fraction of all genes examined shows signs of positive selection, we used the relatively liberal
We identified 35 metabolic genes that appear to be evolving adaptively (supplementary table S1, Supplementary Material online). As expected, the list includes genes coding for glutathione-S-transferase and cytochrome P450, which are known to evolve rapidly (Low et al. 2007; Drosophila 12 Genomes Consortium 2007). Interestingly, we also found a number of enzymes implicated in female germ line development.

Overall, despite their relative constraint, enzymes are only slightly (and not significantly) less likely to evolve under positive selection than nonenzymes (7.8% and 10.5%, respectively, Fisher’s exact test $P = 0.0787$). All pathway groups include similar fractions of adaptively evolving genes ($\chi^2 = 7.16$, permutation $P = 0.5726$). For example, despite its relatively high rate of amino acid substitution, the xenobiotic detoxification pathway group only ranks fourth in the fraction of genes under positive selection, behind groups involved in metabolism of nucleotides, cofactors and vitamins, and carbohydrates. Because several pathway groups contain enzymes that also detoxify foreign compounds, the uniformity in fractions of adaptively evolving genes could arise because the same positively selected genes contribute to several groups. To test this possibility, we eliminated xenobiotic-detoxifying genes from all other pathways. After this manipulation, we still saw no significant heterogeneity in the fractions of adaptively evolving genes among pathway groups ($\chi^2 = 5.8$, permutation $P = 0.7856$). Furthermore, the correlation between the fraction of genes a group has in common with the xenobiotics group and the fraction under positive selection is not significant (Spearman’s $\rho = 0.28$, $P = 0.4339$).

The probability of positive selection does not correlate significantly with any of the network parameters. In particular, although high-degree enzymes exhibit relatively low amino acid substitution rates, they are not less likely to be under positive selection (Spearman’s partial $\rho = -0.092$, permutation $P = 0.0810$ for degree; Spearman’s partial $\rho = 0.013$, permutation $P = 0.8116$ for mean metabolite degree). Moreover, enzymes encoded by adaptively evolving genes have no fewer connections, on average, than other enzymes (e.g., mean degree: 7.59 and 7.56 connections, respectively, Wilcoxon test $P = 0.86$).

Discussion

In this study, we investigated the effects of network architecture on rates of amino acid substitution, duplication, and adaptation. We found that despite overall high conservation, enzymes vary appreciably in their evolutionary rates. This variation is shaped both by enzymes’ functional properties (xenobiotic-detoxifying enzymes evolve relatively rapidly) and by their position in the network (highly connected enzymes are more constrained). Although the variables we examined explain only a small fraction of the variance in protein evolution rates, the results show that even with the noisy data at our disposal, we can detect the effects of metabolic network architecture on enzymes’ rates of amino acid change.

We find that a large proportion of enzymes (36%) act in more than one pathway group. This raises the possibility that modifying one pathway during evolution could lead to changes in other metabolic functions, limiting the precision with which a species can adapt to the array of challenges facing it. Indeed, all the heterogeneity in amino acid substitution rates among pathway groups is due to the presence of genes that are also responsible for detoxification of foreign compounds. This interference could, in principle, be resolved through gene duplication with subsequent subfunctionalization (Lynch and Force 2000). Indeed, we see some evidence that this is happening: the proportion of a pathway group’s enzymes that it shares with other groups correlates strongly with the proportion of its genes that have undergone duplication in at least one of the 12 Drosophila species.

It is unclear how effectively subfunctionalization of duplicate genes frees pathways from interfering with each others’ evolution in our data set. When an enzyme-coding gene has been duplicated and the copies have acquired new functions, the duplicates still code for the same enzyme, and each may be mistakenly annotated as belonging to several pathways. For example, in our data set, each cytochrome P450 is listed as belonging to four pathway groups, although some of these genes may perform specialized tasks in development (Warren et al. 2002). Such misannotation would hamper our ability to detect subfunctionalization. Although random annotation errors do not appear to influence our conclusions (supplementary fig. S2, Supplementary Material online), only experimental analyses can eliminate more systematic biases.

Whatever their influence on the other pathways, it is clear that xenobiotic detoxification genes evolve faster than those in any other group. Are novel foreign compounds driving adaptive amino acid changes and causing this pattern? With the caveat that our methods may not have adequate power to detect positive selection, the answer appears to be, “No.” First, at 0.054, median $d_S/d_S$ for this group is well below 1 (neutrality); second, the proportion of adaptively evolving genes in the xenobiotics group is not significantly above the data set average; third, removing the adaptively evolving genes from consideration does not change the rank of the xenobiotics group (median $\omega$ actually slightly increases to 0.055). Instead, it appears that genes involved in detoxifying foreign compounds are under reduced constraint. This could be because they are less critical to survival than the core metabolism is. However, it is also possible that because most organisms face a wide variety of toxic compounds in nature, amino acid variants that are deleterious in one context can be neutral or beneficial in another (Berenbaum et al. 1996). Purifying selection would be less effective in such a scenario, leading to elevated rates of amino acid substitution.

Purifying selection seems to act more strongly on genes coding for highly connected enzymes, which is consistent with previous observations in yeast (Vitkup et al. 2006). Investigating this further, we find that the correlation is actually driven by a parameter that is best described by the intersection of the degree of an enzyme and the mean degree of its products and substrates. We propose that enzymes with high values of this parameter act at branch
points. Metabolic control theory predicts that such enzymes should have relatively high control over flux in metabolic pathways (Kacser and Burns 1973), so changing their amino acid sequence may cause large phenotypic effects. Our results do indeed indicate that branch point enzymes are more constrained than average. This is consistent with earlier analyses of patterns of polymorphism at enzyme loci (Eanes 1999; Flowers et al. 2007), which found evidence of purifying selection primarily in branch point enzymes. However, we find that such enzymes are not more likely to be coded for by essential genes. Metabolic control theory also predicts high control coefficients for enzymes catalyzing irreversible reactions. In our data, such enzymes are more likely to be encoded by essential genes, although this does not appear to translate into stronger evolutionary constraint. This distinction between essentiality and evolutionary constraint is intriguing, but given the noisy data, it may be due to an ascertainment or statistical artifact.

Despite its impact on purifying selection, enzyme degree does not appear to affect rates of adaptive evolution (although, because we have estimates of degree for only 18 adaptively evolving genes, we may not have the power to detect its effect) nor does it influence the rate of gene duplication. High mean metabolite degree does seem to dampen the rate of gene duplication, although the effect is only marginally significant.

Studies of protein interaction networks also found that proteins with high degree are relatively constrained in amino acid sequence evolution (Fraser et al. 2002; Hahn et al. 2004; Fraser 2005; Hahn and Kern 2005). However, the underlying reason for this association is probably quite different (Fraser et al. 2002; Vitkup et al. 2006; Yu et al. 2007): High-degree proteins devote a larger proportion of their amino acid sequence to protein-binding sites, which are often under constraint.

Our finding that betweenness is uncorrelated with constraint was somewhat surprising, given that it is a significant predictor of the probability that an enzyme is present in a large number of distantly related species (Guimerà and Nunes Amaral 2005; Liu et al. 2007). Perhaps, amino acid substitution rates among closely related species and gene retention across a deep phylogeny are governed by different forces. We also find no correlation between gene essentiality and betweenness (Spearman’s partial ρ = −0.065, permutation P = 0.4498). This is consistent with recent results from yeast (see supplementary table S2 [Supplementary Material online] of Yu et al. [2007]).

In several protein–protein interaction networks, the opposite patterns appear: betweenness is negatively correlated with amino acid substitution rates (Hahn and Kern 2005) and positively associated with essentiality (Yu et al. 2007). Thus, despite some striking similarities in structure and its effect on evolution among different networks, details of their construction are likely to play an important role and should not be overlooked (Doyle et al. 2005).

Taking advantage of whole-genome sequencing and published metabolic network information, we dissected the influences of network architecture and individual protein function on rates of genic evolution. As the data become more refined through experimental analyses, this approach will increase in power and specificity.

Supplementary Material

Supplementary methods, figures S1–S5, and tables S1 and S2 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank A. Larracuente and T. Sackton for the PAML analysis results and K. Montooth and L. Harshman for helpful discussion. The associate editor and two anonymous reviewers provided helpful comments that improved the manuscript. This research was supported by a National Institutes of Health grant to A.G.C. and a National Science Foundation Graduate Research Fellowship to S.R.S.

Literature Cited


Douglas Crawford, Associate Editor

Accepted August 16, 2008