Shared Selective Pressures on Fungal and Human Metabolic Pathways Lead to Divergent yet Analogous Genetic Responses

Haley R. Eidem,[†,1] Kriston L. McGary,[†,1] and Antonis Rokas*[,1]

[1]Department of Biological Sciences, Vanderbilt University

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: antonis.rokas@vanderbilt.edu.

Associate editor: Matthew Hahn

Abstract

Reduced metabolic efficiency, toxic intermediate accumulation, and deficits of molecular building blocks, which all stem from disruptions of flux through metabolic pathways, reduce organismal fitness. Although these represent shared selection pressures across organisms, the genetic signatures of the responses to them may differ. In fungi, a frequently observed signature is the physical linkage of genes from the same metabolic pathway. In contrast, human metabolic genes are rarely tightly linked; rather, they tend to show tissue-specific coexpression. We hypothesized that the physical linkage of fungal metabolic genes and the tissue-specific coexpression of human metabolic genes are divergent yet analogous responses to the range of selective pressures imposed by disruptions of flux. To test this, we examined the degree to which the human homologs of physically linked metabolic genes in fungi (fungal linked homologs or FLOs) are coexpressed across six human tissues. We found that FLOs are significantly more correlated in their expression profiles across human tissues than other metabolic genes. We obtained similar results in analyses of the same six tissues from chimps, gorillas, orangutans, and macaques. We suggest that when selective pressures remain stable across large evolutionary distances, evidence of selection in a given evolutionary lineage can become a highly reliable predictor of the signature of selection in another, even though the specific adaptive response in each lineage is markedly different.

Key words: **gene cluster, inborn error of metabolism, response to selection, selection pressure, Enzyme Commission number, primary metabolism.**

Disruptions of the flux through metabolic processes can lead to metabolic inefficiency (Tatuch et al. 1992), the accumulation of toxic metabolic intermediates (TMIs) (Rathinasabapathi et al. 1994; Jorquera and Tanguay 1997, 2001; Slot and Rokas 2010; McGary et al. 2013), and the loss of essential metabolites (Engel and Angelini 1973), which can all lead to reduced growth and survival. For example, TMIs produced as a consequence of specific genetic mutations are responsible for several human inborn errors of metabolism, such as galactosemia (Gitzelmann 1995) and tyrosinemia (Jorquera and Tanguay 1997). Similarly, mutations that result in the loss or reduction of the essential metabolite carnitine, which facilitates the transfer of fatty acids in the mitochondria, lead to muscle weakness and myopathy (Engel and Angelini 1973).

Because these disruptions of metabolic flux can have a major impact on organism fitness, they represent a ubiquitous source of selection pressures across lineages (Rathinasabapathi et al. 1994; Jorquera and Tanguay 1997, 2001; Slot and Rokas 2010; McGary et al. 2013). Nevertheless, the genetic signatures of the responses to these shared selection pressures may differ between lineages due to differences in population size, mutation rate, mating system, and a wide variety of additional factors. In fungi, the response to these selection pressures has frequently led to placing genes participating in the same pathway adjacent to each other on the chromosome and the formation of physically linked gene clusters (McGary et al. 2013). Clustering is thought to reduce the likelihood of metabolic flux disruptions by enhancing the coexpression of pathway genes (García et al. 2004; Berger et al. 2008; Gacek and Strauss 2012; Schinko et al. 2013), by maintaining favorable combinations of alleles across a pathway's genes (Hittinger et al. 2010; Lang and Botstein 2011) or by ensuring their coinheritance despite aneuploidy or other chromosomal aberrations (Walton 2000). Fungal gene clusters occur in a wide variety of both primary and secondary metabolic pathways across a wide diversity of fungi (Keller and Hohn 1997; Wong and Wolfe 2005; Osbourn 2010; Slot and Rokas 2010; Wisecaver et al. 2014).

In contrast, the human homologs of clustered fungal primary metabolic genes are typically dispersed across chromosomes. For example, the galactose and tyrosine pathways are both clustered, that is, located within close proximity on the same chromosome, in several fungal species (Peñalva 2001; Hittinger et al. 2004; Slot and Rokas 2010), whereas in humans their genes are found on separate chromosomes (Shih et al. 1984; Daude et al. 1995; Fernández-Cañón and Peñalva 1995, 1998; Stambolian et al. 1995; Fernández-Cañón et al. 1996), even though the deleterious effects of the pathways' disruptions are very strong in both lineages (Peñalva 2001; Fridovich-Keil 2006; Mumma et al. 2008). Finally, although there are no known human metabolic pathways in the form of gene clusters, our

Letter

previous analysis (McGary et al. 2013) shows that there are eight human gene pairs that are both immediately adjacent on human chromosomes and immediate metabolic neighbors in the same pathway (supplementary table S1, Supplementary Material online). Of these eight human gene pairs, only two pairs are also chromosomal neighbors in a fungal genome (McGary et al. 2013).

If not clustering, what is the response to these shared selection pressures in human metabolic pathways? Because, in contrast to fungi, humans have hundreds of cell types which give rise to tissues and organs, human metabolic pathway activity is not ubiquitously distributed. Rather, it is primarily confined to specialized tissues and organs, such as the liver and kidneys. Thus, the response to the selection pressure imposed by disruptions of metabolic flux in humans is the tissue-specific coexpression of pathway genes (Jordan et al. 2004, 2005; Liao and Zhang 2006).

We hypothesized that the physical linkage of primary metabolic pathway genes in fungi and the tissue-specific coexpression of their homologs in humans represent divergent yet analogous responses to a shared set of selective pressures (fig. 1). We tested this hypothesis by examining whether the homologs of physically linked fungal metabolic genes in humans, chimps, gorillas, orangutans, and macaques are more significantly correlated in their expression across six tissues than primate homologs of unlinked fungal genes over and above what is observed for genes belonging to the same pathway.

Results

We identified all pairs of human genes from the same metabolic pathway that share an intermediate compound as their product or substrate, which we will refer to as metabolic neighbors. Using data from 126 fungal genomes that contain 948 physically linked fungal gene pairs that are also metabolic neighbors (McGary et al. 2013), we divided human metabolic neighbors into two groups; those that consist of gene pairs that have physically linked fungal homologs (fungal linked homolog gene pairs or FLO gene pairs, supplementary table S2, Supplementary Material online), and those that do not (non-FLO gene pairs). The list of the 948 physically linked fungal gene pairs that are also metabolic neighbors used is provided in supplementary table S3, Supplementary Material online, and the list of fungal genomes is provided in supplementary table S4, Supplementary Material online.

The coexpression of both FLO ($n = 1{,}805$; $\bar{r} = 0.32$) and non-FLO gene pairs ($n = 10{,}200$; $\bar{r} = 0.22$) is significantly more correlated than a background comprised all possible pairs of metabolic genes ($n = 4{,}037{,}061$; $\bar{r} = 0.20$) ($P = 1.13e^{-28}$ and $P = 1.08e^{-4}$, respectively; fig. 2A). If our hypothesis that the physical linkage of pathway genes in fungi and the tissue-specific coexpression of their homologs in humans represent analogous responses to a shared set of

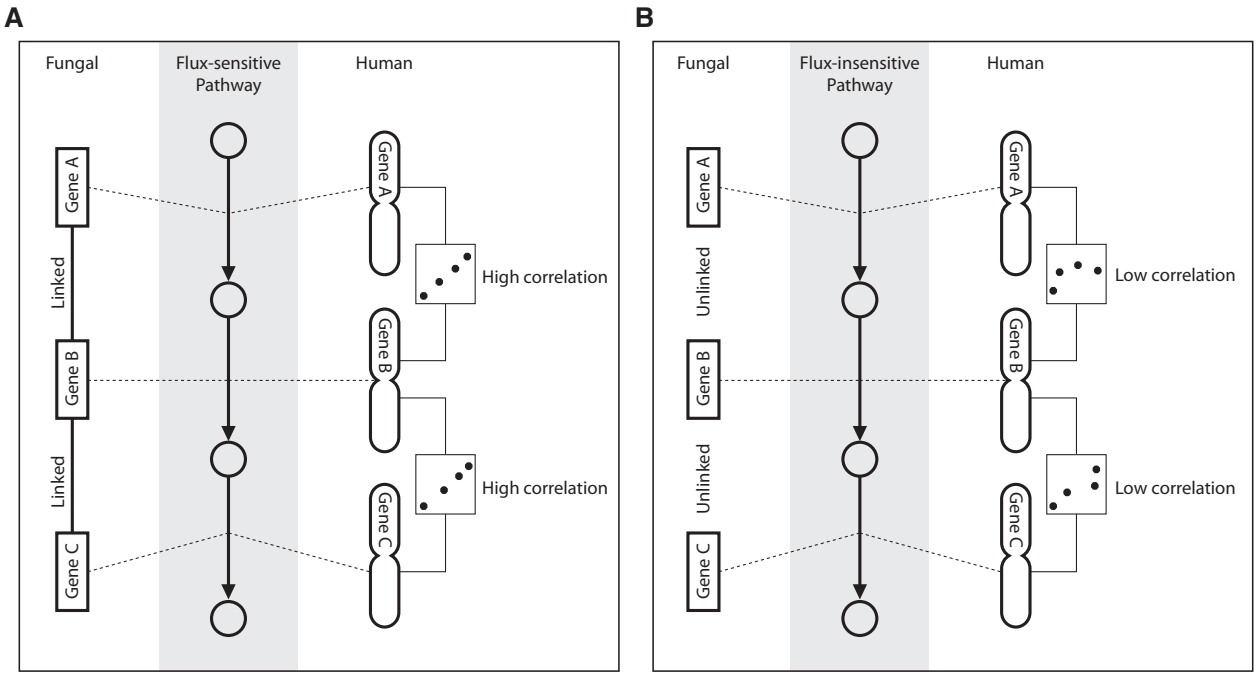

FIG. 1. An illustration of the hypothesis that the physical linkage of pathway genes in fungi and the tissue-specific coexpression of their homologs in humans represent divergent yet analogous responses to the same set of selective pressures. Metabolic pathways that are sensitive to variations in flux (*A*) face increased selection pressures to avoid disruption relative to pathways that are less sensitive to flux variation (*B*). We hypothesize that the range of selective pressures imposed by disruptions of flux has led to divergent yet analogous responses in fungal (physical linkage and the formation of gene clusters) and human (increased correlation of gene coexpression across tissues) metabolic pathways. Thus, we expect that genes in flux-sensitive metabolic pathways (*A*) have been subject to the same set of selection pressures but resulted in divergent yet analogous responses; in fungi, the signature of this response is the increased probability of gene cluster formation, whereas in humans it the increased correlation of expression of pathway genes. In contrast, we do not expect these signatures in flux-insensitive pathways (*B*).

selective pressures is correct, we would expect that FLO gene pairs show a higher correlation in their coexpression across human tissues than non-FLO gene pairs.

We found that FLO gene pair coexpression is significantly more correlated than non-FLO gene pair coexpression in human genome-wide expression data (Brawand et al. 2011) from brain, cerebellum, heart, kidney, liver, and testis ($P = 4.18e^{-18}$; fig. 2*A*). We found the same significant pattern when we analyze the same gene pairs in the same six tissues in chimps ($P = 6.14e^{-8}$), gorillas ($P = 3.16e^{-5}$), orangutans ($P = 1.06e^{-2}$), and macaques ($P = 1.64e^{-13}$) (supplementary fig. S1, Supplementary Material online).

The significance of the correlation of FLO gene pair coexpression is nicely illustrated by examining pathways associated with inborn errors of metabolism (Peñalva 2001; Fridovich-Keil 2006). For example, the enzymes galactose-1P uridylyltransferase (GALT) and galactokinase (GALK) in the galactose pathway, which handle the TMI galactose-1-phosphate associated with the human disease galactosemia, display the highest correlation in their gene expression values ($r = 0.87$). Similarly, all genes in the tyrosine catabolism pathway, whose disruption leads to three different types of tyrosinemia as well as to alkaptonuria, Garrod's first inborn error of metabolism (Garrod 1902), are highly correlated in their expression in the six human tissues examined (supplementary fig. S2, Supplementary Material online).

Interestingly, the observed distribution of the correlation of FLO gene pair coexpression across tissues is bimodal (fig. 2*A*). This might be explained by the presence of paralogs with the same enzymatic activities that are active in multiple human tissues, leading to FLO gene pair coexpression in one tissue and not in others (supplementary fig. S3, Supplementary Material online). Thus, we grouped all human metabolic genes by their Enzyme Commission (EC) numbers, which classify enzymes by their biochemical reactions, and summed their expression levels. We will refer to this sum as the cumulative EC expression. Using the same logic, we then collapsed all FLO gene pairs into FLO EC pairs.

We found that the correlation of FLO EC pair coexpression ($n = 137$; $\bar{r} = 0.52$) is significantly higher than the correlation of FLO gene pair coexpression ($P = 1.92e^{-8}$); this is illustrated by the fact that the distribution of the correlation of FLO EC pair coexpression across tissues is unimodal and substantially shifted toward higher correlation coefficients (fig. 2*B*) relative to that of FLO gene pair coexpression (fig. 2*A*). More importantly, we found that the correlation of FLO EC pair coexpression is greater than that of non-FLO EC pairs ($n = 1{,}594$; $\bar{r} = 0.32$) ($P = 4.44e^{-10}$). Furthermore, the FLO gene pairs that make up the most correlated FLO EC pairs (identified using a correlation cutoff of $r \geq 0.8$) are not always highly correlated themselves and show a bimodal distribution similar to that of all FLO gene pairs (supplementary fig. S4, Supplementary Material online).

The greater correlation of tissue coexpression for FLO EC pairs relative to FLO gene pairs points to the control of metabolic flux as one target of selection. Assuming that other factors (e.g., relative toxicity of intermediate compounds) are equal, one can hypothesize that greater selection pressure would be imposed on pathways with more total flux. To test this hypothesis, we used a model of metabolic flux in

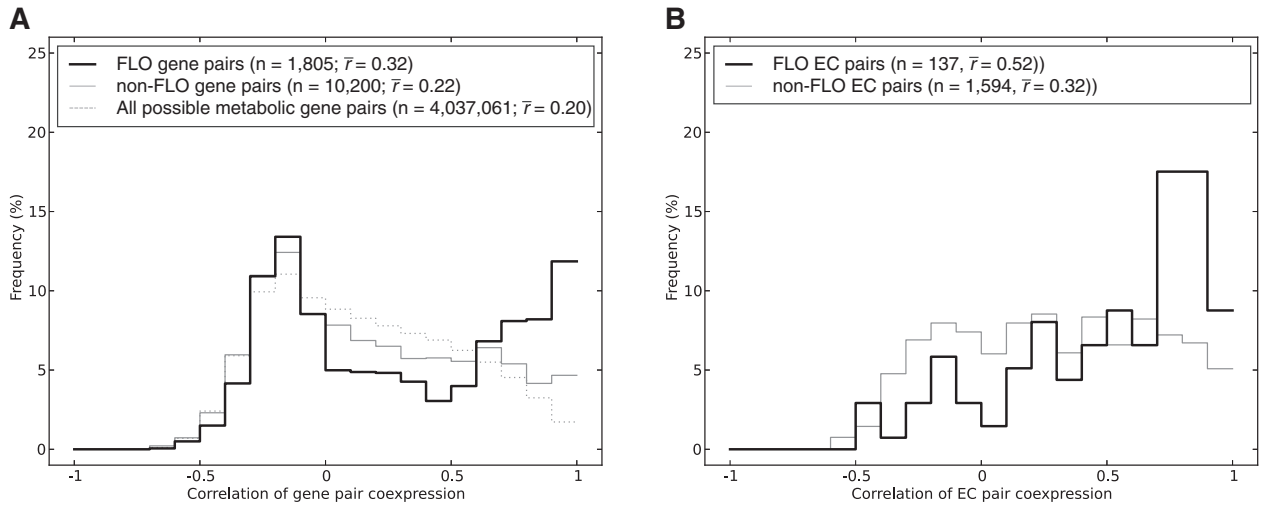

Fig. 2. The human homologs of physically linked metabolic genes in fungi (FLO gene pairs) are significantly more correlated in their expression profiles across human tissues than other metabolic genes. (*A*) FLO gene pairs are more significantly correlated in their expression across six human tissues than non-FLO gene pairs ($P = 4.18e^{-18}$) as well as than background gene pairs ($P = 1.13e^{-28}$). The average correlation of coexpression for FLO gene pairs is $\bar{r} = 0.32$ and the average correlation for non-FLO gene pairs is $\bar{r} = 0.22$; both are much higher than the. $\bar{r} = 0.20$ average for background gene pairs. Note that the distribution of the correlation of FLO gene pair coexpression is bimodal, which is likely explained by the presence of homologs with the same enzymatic activities that are active in different human tissues, leading to high correlation in one tissue and anticorrelation in others (supplementary fig. S3, Supplementary Material online). (*B*) Cumulative enzymatic coexpression of FLO EC pairs is more significantly correlated across six human tissues than coexpression of non-FLO EC pairs ($P = 4.44e^{-10}$). The average correlation of FLO EC pair coexpression is $\bar{r} = 0.52$ and that of non-FLO EC pair coexpression is $\bar{r} = 0.32$. Note that the distribution of the correlation of FLO EC pair coexpression is unimodal and shows a marked shift toward higher correlation coefficients relative to the distribution of the correlation of FLO gene pair coexpression from (*A*).

Saccharomyces cerevisiae (data from Duarte et al. 2004) to compare the average flux of FLO and non-FLO EC pairs. We found that FLO EC pairs averaged 5-fold more flux than non-FLO EC pairs (FLO, 0.04 mmol/g/h; non-FLO, 0.008 mmol/g/h; supplementary table S5, Supplementary Material online).

With the overt differences in total metabolic flux between FLO and non-FLO EC pairs, we decided to examine tissue-specific coexpression and physical linkage in light of the structure of the metabolic network. We labeled the reactions of primary metabolism as either central and peripheral using network structure (according to Ma et al. 2007). Metabolic reactions in the central part of the network, also known as the giant strong component, process metabolites that can be interconverted easily. In contrast, the peripheral component is composed of reaction involving metabolites that can only be degraded (the "in" component) and reactions involving metabolites that can only be synthesized (the "out" component). FLO EC pairs have higher average correlation of coexpression than non-FLO EC pairs regardless of whether they are part of central (FLO, $\bar{r} = 0.42$; non-FLO, $\bar{r} = 0.33$) or part of various aspects of peripheral metabolism (FLO, $\bar{r}_{in} = 0.61$ vs. non-FLO, $\bar{r}_{in} = 0.35$; FLO, $\bar{r}_{out} = 0.65$; non-FLO, $\bar{r}_{out} = 0.32$; FLO, $\bar{r}_{mixed} = 0.56$; non-FLO, $\bar{r}_{mixed} = 0.31$; supplementary table S6, Supplementary Material online). This is true even though FLO and non-FLO EC pairs are involved in central metabolism at roughly the same rate (FLO, 31% vs. non-FLO, 33%; supplementary table S6, Supplementary Material online). Thus, physical linkage in fungi is a better predictor of tissue-specific correlation of coexpression than position in the human metabolic network.

Finally, although our data suggest that the physical linkage of fungal metabolic genes and the tissue-specific coexpression of human metabolic genes are divergent yet analogous responses to the range of selective pressures imposed by disruptions of flux, it is possible that the correlation of FLO gene pair coexpression across tissues in humans simply mirrors fungal gene coexpression. As gene coexpression data from multiple fungal tissues are not available, we used condition-specific correlation of coexpression data from *S. cerevisiae* (data from Waern and Snyder 2013; correlation calculated as in Soria et al. 2014). We found that FLO gene pair coexpression across multiple yeast growth conditions is only weakly correlated ($\bar{r} = 0.065$) and minimally different to the coexpression correlation of non-FLO gene pairs ($\bar{r} = 0.044$) (supplementary table S7, Supplementary Material online). This weak correlation contrasts with the much higher correlation of human FLO gene pair coexpression across tissues ($\bar{r} = 0.32$; fig. 2). Furthermore, the correlation of FLO gene pair coexpression across multiple growth conditions in yeast is not correlated with the amount of flux inferred (from Duarte et al. 2004) to be shared by each gene pair for both FLO and non-FLO gene pairs (supplementary table S7, Supplementary Material online).

Discussion

The major result of this study is that human metabolic neighbors whose fungal homologs are physically linked (i.e., FLOs) are significantly more likely to be coexpressed than human metabolic neighbors without linked fungal homologs (i.e., non-FLOs; fig. 2). This result supports the hypothesis that the physical linkage of pathway genes in fungi and the tissue-specific coexpression of their homologs in humans represent divergent yet analogous responses to a shared set of selective pressures. It should be noted that these responses (physical linkage in fungi, tissue-specific coexpression in primates) are over and above the responses to the more general set of selection pressures that have favored the coactivity of genes in a given pathway.

Even though previous studies have argued that eukaryotic gene order is not random such that genes with similar expression patterns tend to be linked (Hurst et al. 2004), our study shows that disruption of metabolic flux in humans and primates frequently leads to tissue-specific coexpression of primary metabolic genes in the absence of tight physical linkage. Consider, for example, homogentisate dioxygenase (Hmg1) and fumaryl-acetoacetate hydrolase (FahA). Hmg1 is the producer of the toxic intermediate responsible for the human disease tyrosinemia type I and FahA is the enzyme that detoxifies the Hmg1-produced intermediate. Although the two genes are physically linked in several fungi, in humans they show a very high degree of tissue-specific coexpression even though they are located on different chromosomes (supplementary fig. S2, Supplementary Material online).

Although disruptions of flux represent a set of selection pressures shared between fungi and primates, the specific responses favored in the two lineages differ. This is not surprising, because any response to any selection pressure critically depends on a wide variety of parameters such as population size, reproductive lifestyle, mutational opportunity, and genome architecture (Lynch 2007). Examination of other lineages supports this notion; for example, the response to selection in certain plant metabolic pathways results in, like fungi, the physical linkage of pathway genes (Takos and Rook 2012; Nützmann and Osbourn 2014). In contrast, in bacterial lineages the same pressures manifest themselves most obviously in the formation of operons (Notebaart et al. 2008), whose structure and function capture both physical linkage and tight coexpression of pathway genes.

Whatever the specific response favored in a given lineage may be, disruptions of metabolic flux are expected to lead to significant increases in the correlated activities of the enzymes in a given pathway. This correlation of enzyme activity may be achieved in a number of ways, for example, through subcellular colocalization (Srere 2000; Ovádi et al. 2004; Durek and Walther 2008; Pérez-Bercoff et al. 2011) or through tissue-specific coexpression (fig. 2*A* and *B*).

Our results support the hypothesis that the physical linkage of pathway genes in fungi and the tissue-specific coexpression of their homologs in humans represent divergent yet analogous responses to a shared set of selective pressures. One implication of these results is that the observed response to selection in fungal metabolic pathways becomes predictive of the response of their human and primate homologs, and vice versa. More generally, when selective pressures remain

constant across large evolutionary distances, evidence of selection in a given evolutionary lineage can become a highly reliable predictor of the signature of selection in another, allowing for potentially surprising predictions.

Materials and Methods

Identifying FLOs in humans

We used the KEGG database (Ogata et al. 1999) to identify human metabolic genes, their enzymatic reactions, and their metabolic substrates and products. The enzyme relationships were parsed from KEGG's KGML format. Rather than using reaction (EC) pairs that co-occur in arbitrarily defined pathways, we only used immediate metabolic neighbors, which are pairs of genes that directly handle a shared metabolic intermediate compound with no intervening reaction. We then categorized the reaction (EC) pairs as FLOs or non-FLOs. Such pairs of metabolic neighbors were identified as FLOs when the homologs for both human genes were found to be physically linked (i.e., immediately adjacent on a chromosome) in one or more fungal genomes, based on our previously published collection of metabolic gene clusters from fungi (McGary et al. 2013) (supplementary table S3, Supplementary Material online). Briefly, homology of fungal genes used a core set of fungal genomes annotated with EC numbers in KEGG and extended the EC annotations to unannotated genomes using Basic Local Alignment Search Tool (BLAST) (McGary et al. 2013). We were only interested in gene pairs and EC pairs that catalyze neighboring enzymatic reactions, not identical reactions. Therefore, we specifically excluded fungal gene pairs that had overlapping EC annotations, that is, genes in each pair could not have the same enzymatic activity or be similar enough by BLAST to receive the same annotation, which effectively removes tandem duplicates (McGary et al. 2013). Homology between fungal genes and human genes was inferred based on shared EC number annotation in KEGG; this functionally inferred homology typically indicates that the homologs are at least related at the level of a gene family. EC annotation was extended from human genes to primate genes according to previously published primate orthology assignments (Brawand et al. 2011). Metabolic gene pairs without linked homologs in any fungal genome were considered non-FLOs. The background is all possible combinations of pairs of metabolic genes.

It should be noted that in fungal genomes, gene clusters are observed in both primary (e.g., galactose assimilation, allantoin utilization, tyrosine catabolism) and secondary metabolism (e.g., penicillin biosynthesis). Although primary metabolic pathways, and as a consequence primary metabolic gene clusters, are well represented in the KEGG database, this is not always the case for secondary metabolic pathways and gene clusters. Thus, the metabolic gene pairs that are linked on fungal chromosomes (FLOs) in our data set are generally part of primary metabolism (supplementary table S2, Supplementary Material online), even though some pairs (e.g., those in steroid biosynthesis or porphyrin metabolism) might be in the gray zone between classic primary metabolism and classic secondary metabolism.

Estimating the Correlation of Coexpression for FLO, Non-FLO, and Background Gene Pairs

We calculated the tissue-specific correlation of coexpression for FLO, non-FLO, and background gene pairs, using the normalized gene expression data (measured using RNA-seq and expressed as reads per kilobase per million mapped reads, RPKM) from multiple samples from six human, chimp, gorilla, orangutan, and macaque tissues (brain, cerebellum, heart, kidney, liver, and testes) reported by Brawand et al. (2011). We calculated the degree of correlation between genes in a given pair using Pearson's correlation coefficient as implemented in Python's Scipy module (Oliphant 2007). The statistical significance of the differences in the distributions of correlation coefficients for FLO, non-FLO, and background gene pairs was calculated using a Mann–Whitney *U* test (Wilcoxon 1945; Mann and Whitney 1947) as implemented in Python's Scipy module (Oliphant 2007).

Estimating the Correlation of Cumulative Enzymatic Expression for FLO, Non-FLO, and Background EC Pairs

To infer cumulative enzymatic expression for FLO, non-FLO, and background gene pairs, we grouped all human metabolic genes by their EC numbers as annotated in KEGG (Ogata et al. 1999). For each tissue sample, we inferred the cumulative expression for each EC number by summing the RPKM expression of all genes associated with that EC number, which are typically members of the same gene family. Next, we collapsed all FLO gene pairs into FLO EC pairs. Similarly to our analysis of the correlation of gene coexpression, the degree of correlation of EC pair coexpression was calculated using Pearson's correlation coefficient as implemented in Python's Scipy module (Oliphant 2007). The statistical significance of the differences in the distributions of correlation coefficients for FLO, non-FLO, and background EC pairs was calculated using a Mann–Whitney *U* test (Wilcoxon 1945; Mann and Whitney 1947) as implemented in Python's Scipy module (Oliphant 2007).

Because gene expression can be highly correlated in tandem duplicates, we specifically checked the 2,020 human FLO gene pairs (reported in supplementary table S2, Supplementary Material online) and found that the genes from only 83 pairs were colocated on the same chromosomal arm. Manual evaluation of these 83 gene pairs revealed only one EC pair that may be affected by tandem duplication (1.14.13.30 and 1.14.14.1, involving members of the cytochrome p450 family).

Calculating the Relationships between Physical Linkage, Metabolic Flux, Network Structure, and TissueSpecific Coexpression

A model of metabolic flux for *S. cerevisiae* in minimal media (Duarte et al. 2004) was downloaded from http://systemsbiology.ucsd.edu/sites/default/files/Attachments/Images/InSilicoOrganisms/yeast/Sc_iND750_flux1.xml (last accessed December 11, 2014) and parsed for flux values

inferred by the model for each EC number. Shared flux through each EC pair was calculated to be the minimum flux through either of the two reactions. All yeast EC pairs were divided into FLO and non-FLO pairs as described above for human EC pairs. We report all EC pairs with inferred flux values in supplementary table S5, Supplementary Material online.

ECs were labeled as either central and peripheral (according to Ma et al. 2007) using the file "inline-supplementary-material-8.xls," last accessed December 16, 2014 from http://msb.embopress.org/highwire/filestream/23931/field_highwire_a_enclosures/0/supplementary-material.inline-supplementary-material-8.xls?download=true. Central metabolism is labeled GSC, peripheral metabolism is subdivided into two categories, IN and OUT. However, the provided annotation often labeled one or both ECs in a pair as being part of both central and peripheral metabolism. For EC pairs with any mixed annotation, they are reported as MIXED in supplementary table S6, Supplementary Material online.

We used condition-specific gene expression data from *S. cerevisiae* across 18 conditions (Waern and Snyder 2013). We calculated correlation of gene coexpression according to the methods in Soria et al. (2014). We report each gene pair by their EC annotation along with their coexpression correlation and the flux for the associated EC pair (inferred from model as reported above) in supplementary table S7, Supplementary Material online.

Supplementary Material

Supplementary tables S1–S7 and figures S1–S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University. This work was supported by the National Science Foundation (DEB-0844968) and the March of Dimes.

References

Berger H, Basheer A, Böck S, Reyes-Dominguez Y, Dalik T, Altmann F, Strauss J. 2008. Dissecting individual steps of nitrogen transcription factor cooperation in the *Aspergillus nidulans* nitrate cluster. *Mol Microbiol.* 69:1385–1398.

Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478: 343–348.

Daude N, Gallaher TK, Zeschnigk M, Starzinskipowitz A, Petry KG, Haworth IS, Reichardt JKV. 1995. Molecular-cloning, characterization, and mapping of a full-length cDNA encoding human UDP-galactose 4′-epimerase. *Biochem Mol Med.* 56:1–7.

Duarte N, Herrgård M, Palsson B. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 14:1298–1309.

Durek P, Walther D. 2008. The integrated analysis of metabolic and protein interaction networks reveals novel molecular organizing principles. *BMC Syst Biol.* 2:100.

Engel AG, Angelini C. 1973. Carnitine deficiency of human skeletal muscle with associated lipid storage myopathy: a new syndrome. *Science* 179:899–902.

Fernández-Cañón JM, Granadino B, De Bernabé DB, Renedo M, Fernández-Ruiz E, Peñalva MA, De Córdoba SR. 1996. The molecular basis of alkaptonuria. *Nat Genet.* 14:19–24.

Fernández-Cañón JM, Peñalva MA. 1995. Fungal metabolic model for human type I hereditary tyrosinaemia. *Proc Natl Acad Sci U S A.* 92: 9132–9136.

Fernández-Cañón JM, Peñalva MA. 1998. Characterization of a fungal maleylacetoacetate isomerase gene and identification of its human homologue. *J Biol Chem.* 273:329–337.

Fridovich-Keil JL. 2006. Galactosemia: the good, the bad, and the unknown. *J Cell Physiol.* 209:701–705.

Gacek A, Strauss J. 2012. The chromatin code of fungal secondary metabolite gene clusters. *Appl Microbiol Biotechnol.* 95:1389–1404.

García I, Gonzalez R, Gómez D, Scazzocchio C. 2004. Chromatin rearrangements in the prnD-prnB bidirectional promoter: dependence on transcription factors. *Eukaryot Cell.* 3:144–156.

Garrod A. 1902. The incidence of alkaptonuria: a study in chemical individuality. *Lancet* 160:1616–1620.

Gitzelmann R. 1995. Galactose-1-phosphate in the pathophysiology of galactosemia. *Eur J Pediatr.* 154:S45–S49.

Hittinger CT, Gonçalves P, Sampaio JP, Dover J, Johnston M, Rokas A. 2010. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464:54–58.

Hittinger CT, Rokas A, Carroll S. 2004. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci U S A.* 101:14144–14149.

Hurst LD, Pál C, Lercher MJ. 2004. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet.* 5:299–310.

Jordan IK, Mariño-Ramírez L, Koonin EV. 2005. Evolutionary significance of gene expression divergence. *Gene* 345:119–126.

Jordan IK, Mariño-Ramírez L, Wolf YI, Koonin EV. 2004. Conservation and coevolution in the scale-free human gene coexpression network. *Mol Biol Evol.* 21:2058–2070.

Jorquera R, Tanguay RM. 1997. The mutagenicity of the tyrosine metabolite, fumarylacetoacetate, is enhanced by glutathione depletion. *Biochem Biophys Res Commun.* 232:42–48.

Jorquera R, Tanguay RM. 2001. Fumarylacetoacetate, the metabolite accumulating in hereditary tyrosinemia, activates the ERK pathway and induces mitotic abnormalities and genomic instability. *Hum Mol Genet.* 10:1741–1752.

Keller NP, Hohn TM. 1997. Metabolic pathway gene clusters in filamentous fungi. *Fungal Genet Biol.* 21:17–29.

Lang GI, Botstein D. 2011. A test of the coordinated expression hypothesis for the origin and maintenance of the GAL cluster in yeast. *PLoS ONE* 6(9):e25290.

Liao B-Y, Zhang J.. 2006. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol Biol Evol.* 23:1119–1128.

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.

Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I. 2007. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol.* 3:135.

Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat.* 18: 50–60.

McGary KL, Slot JC, Rokas A. 2013. Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds. *Proc Natl Acad Sci U S A.* 110: 11481–11486.

Mumma JO, Chhay JS, Ross KL, Eaton JS, Newell-Litwa KA, Fridovich-Keil JL. 2008. Distinct roles of galactose-1P in galactose-mediated growth arrest of yeast deficient in galactose-1P uridylyltransferase (GALT) and UDP-galactose 4′-epimerase (GALE). *Mol Genet Metab.* 93:160–171.

Notebaart RA, Teusink B, Siezen RJ, Papp B. 2008. Co-regulation of metabolic genes is better explained by flux coupling than by network distance. *PLoS Comput Biol.* 4:e26.
Nützmann H-W, Osbourn A. 2014. Gene clustering in plant specialized metabolism. *Curr Opin Biotechnol.* 26:91–99.
Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 27: 29–34.
Oliphant TE. 2007. Python for scientific computing. *Comput Sci Eng.* 9: 10–20.
Osbourn A. 2010. Secondary metabolic gene clusters: evolutionary toolkits for chemical innovation. *Trends Genet.* 26:449–457.
Ovádi J, Orosz F, Hollán S. 2004. Functional aspects of cellular microcompartmentation in the development of neurodegeneration: mutation induced aberrant protein-protein associations. *Mol Cell Biochem.* 256–257:83–93.
Peñalva MÁ. 2001. A fungal perspective on human inborn errors of metabolism: alkaptonuria and beyond. *Fungal Genet Biol.* 34: 1–10.
Pérez-Bercoff.Å, McLysaght A, Conant GC. 2011. Patterns of indirect protein interactions suggest a spatial organization to metabolism. *Mol Biosyst.* 7:3056–3064.
Rathinasabapathi B, McCue KF, Gage DA, Hanson AD. 1994. Metabolic engineering of glycine betaine synthesis: plant betaine aldehyde dehydrogenases lacking typical transit peptides are targeted to tobacco chloroplasts where they confer betaine aldehyde resistance. *Planta* 193:155–162.
Schinko T, Gallmetzer A, Amillis S, Strauss J. 2013. Pseudo-constitutivity of nitrate-responsive genes in nitrate reductase mutants. *Fungal Genet Biol.* 54:34–41.
Shih LY, Suslak L, Rosin I, Searle BM, Desposito F. 1984. Gene dosage studies supporting localization of the structural gene for galactose-1-phosphate uridyl transferase (GALT) to band p13 of chromosome 9. *Am J Med Genet.* 19:539–543.
Slot JC, Rokas A. 2010. Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc Natl Acad Sci U S A.* 107:10136–10141.
Soria PS, McGary KL, Rokas A. 2014. Functional divergence for every paralog. *Mol Biol Evol.* 31:984–992.
Srere PA. 2000. Macromolecular interactions: tracing the roots. *Trends Biochem Sci.* 25:150–153.
Stambolian D, Ai Y, Sidjanin D, Nesburn K, Sathe G, Rosenberg M, Bergsma DJ. 1995. Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts. *Nat Genet.* 10: 307–312.
Takos AM, Rook F. 2012. Why biosynthetic genes for chemical defense compounds cluster. *Trends Plant Sci.* 17:383–388.
Tatuch Y, Christodoulou J, Feigenbaum A, Clarke JT, Wherret J, Smith C, Rudd N, Petrova-Benedict R, Robinson BH. 1992. Heteroplasmic mtDNA mutation (T—-G) at 8993 can cause Leigh disease when the percentage of abnormal mtDNA is high. *Am J Hum Genet.* 50: 852–858.
Waern K, Snyder M. 2013. Extensive transcript diversity and novel upstream open reading frame regulation in yeast. *G3* 3:343–352.
Walton JD. 2000. Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet Biol.* 30:167–171.
Wilcoxon F. 1945. Individual comparisons by ranking methods. *Biometr Bull.* 1:80–83.
Wisecaver JH, Slot JC, Rokas A. 2014. The evolution of fungal metabolic pathways. *PLoS Genet.* 10:e1004816.
Wong S, Wolfe KH. 2005. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet.* 37:777–782.