Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds

Kriston L. McGary[a], Jason C. Slot[a,b], and Antonis Rokas[a,1]

[a]Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235; and [b]Department of Plant Pathology, The Ohio State University, Columbus, OH 43210

Edited* by Sean B. Carroll, University of Wisconsin-Madison, Madison, WI, and approved May 28, 2013 (received for review March 7, 2013)

Genomic analyses have proliferated without being tied to tangible phenotypes. For example, although coordination of both gene expression and genetic linkage have been offered as genetic mechanisms for the frequently observed clustering of genes participating in fungal metabolic pathways, elucidation of the phenotype(s) favored by selection, resulting in cluster formation and maintenance, has not been forthcoming. We noted that the cause of certain well-studied human metabolic disorders is the accumulation of toxic intermediate compounds (ICs), which occurs when the product of an enzyme is not used as a substrate by a downstream neighbor in the metabolic network. This raises the hypothesis that the phenotype favored by selection to drive gene clustering is the mitigation of IC toxicity. To test this, we examined 100 diverse fungal genomes for the simplest type of cluster, gene pairs that are both metabolic neighbors and chromosomal neighbors immediately adjacent to each other, which we refer to as "double neighbor gene pairs" (DNGPs). Examination of the toxicity of their corresponding ICs shows that, compared with chromosomally nonadjacent metabolic neighbors, DNGPs are enriched for ICs that have acutely toxic LD_{50} doses or reactive functional groups. Furthermore, DNGPs are significantly more likely to be divergently oriented on the chromosome; remarkably, ∼40% of these DNGPs have ICs known to be toxic. We submit that the structure of synteny in metabolic pathways of fungi is a signature of selection for protection against the accumulation of toxic metabolic intermediates.

gene cluster | gene orientation | inborn error of metabolism | secondary metabolism | specialized metabolism

There is a critical distinction between selection for a trait, which occurs at the phenotypic level, and selection of the underlying molecular genetic mechanisms that generate the trait (1). For example, some butterfly species avoid predation by mimicking the wing-pattern morphs of other butterflies that are toxic to birds (2, 3). In this case, predation avoidance selects for specific wing-pattern morphs, resulting in the selection of specific allelic combinations of genes in a supergene locus (4).

The rise of molecular biology and subsequent abundance of genomic data have shifted the focus from phenotypes to genotypes, blurring the distinction between selection for and selection of. For example, many studies indicate that selection has sculpted the order and orientation of genes in eukaryotic genomes (5–13); although these analyses convincingly argue that specific genetic mechanisms are important in fine-tuning the structure of synteny on chromosomes, they implicitly assume that it is the genetic mechanisms themselves, rather than organismal phenotypes, that selection targets.

One of the most conspicuous characteristics of fungal genomes is that the genes participating in a metabolic pathway are often physically linked, or clustered. Although both coordination of gene expression (5, 6, 11–16) and genetic linkage (7, 17) have been offered as genetic mechanisms underlying these metabolic gene clusters, unlike the butterfly example above, the phenotype(s) that drives their formation and maintenance in fungi is not known. We noted that the cause of several human metabolic disorders is the accumulation of toxic intermediate compounds (ICs). For example, galactosemia arises from disruption of the three enzymatic steps involved in galactose degradation (18). The most severe type of galactosemia results from the loss of galactose-1-phosphate uridyl transferase (encoded by *Galt* in human; *Gal7* in yeast), which leads to accumulation of galactose-1-phosphate (18), a potent competitive inhibitor of phosphoglucomutase (19), which disrupts the glycolytic pathway. Intriguingly, the genes of the *Gal* pathway are clustered in several different yeast lineages (20) (Fig. 1*A*). The human metabolic disease tyrosinemia also exists in multiple types that differ in severity, which are mimicked by the phenotypes of the corresponding gene knockouts in the model fungus *Aspergillus nidulans* (21). Tyrosinemia type I, the most severe, results from the loss of the final step in the pathway and the accumulation of the reactive, genotoxic compound fumarylacetoacetate, which increases the incidence of hepatic cancer (21). Similar to the *Gal* pathway, the genes of this metabolic pathway are also clustered in many fungal species, including *A. nidulans* (21) (Fig. 1*B*).

These examples raise the hypothesis that amelioration of the deleterious effects associated with the accumulation of toxic ICs is a metabolic phenotype under sufficient selection pressure to drive widespread physical linking of genes in fungi. To test this hypothesis, we first identified the simplest type of fungal gene cluster, gene pairs that are neighbors in the metabolic network, i.e., that share an IC and that are also immediately adjacent neighbors on the chromosome, which we refer to as double neighbor gene pairs (DNGPs). If true, we would expect to find that DNGPs handle more toxic ICs than chromosomally nonadjacent metabolic neighbors, which we will refer to as "background" (Fig. 1*C*).

Results

Examination of the toxicity of ICs associated with 948 DNGPs (Dataset S1, Table S1), as measured by LD_{50} dose (Dataset S1, Table S2), showed that they handled more toxic ICs than background. Interestingly, IC toxicity was significantly associated with DNGP gene orientation such that the set of divergently oriented DNGPs with toxic ICs was significantly larger than expected (Figs. 1*D* and 2*A*). In contrast, colinearly oriented DNGPs were not enriched for toxic ICs, and convergently oriented DNGPs were depleted. Further examination of DNGPs showed that cases of divergent orientation were significantly more numerous than expected by chance, whereas both convergent and colinear DNGPs were significantly fewer (Table 1). Studies have shown that divergently oriented genes are more tightly regulated (16) and tend to be coexpressed (14), so enrichment for divergent

Author contributions: K.L.M., J.C.S., and A.R. designed research; K.L.M. and J.C.S. performed research; K.L.M. and J.C.S. contributed new reagents/analytic tools; K.L.M., J.C.S., and A.R. analyzed data; and K.L.M., J.C.S., and A.R. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: antonis.rokas@vanderbilt.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1304461110/-/DCSupplemental.

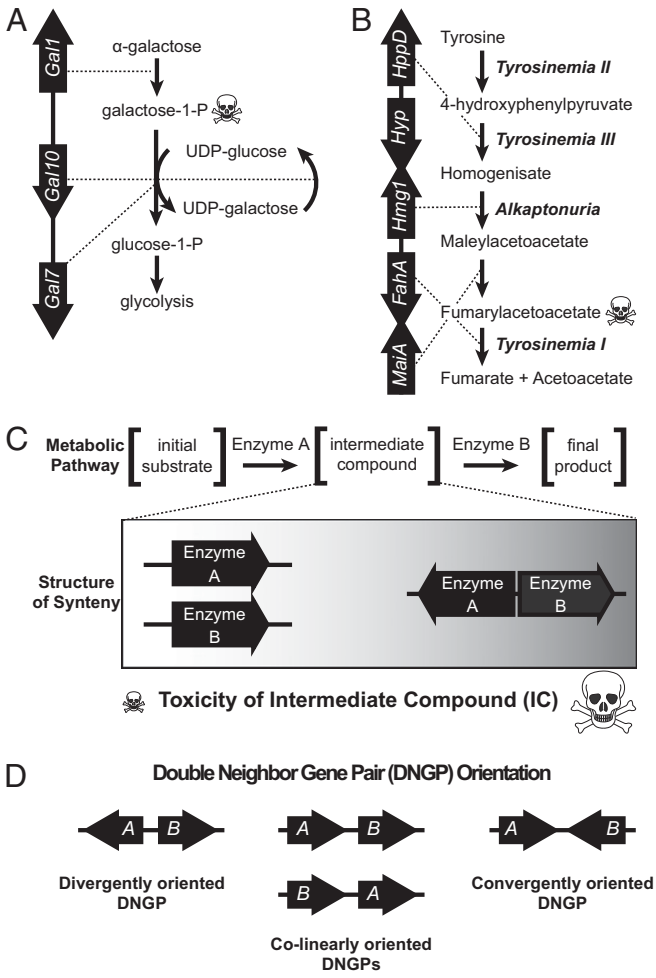

Fig. 1. Avoidance of toxic IC accumulation is a phenotype favored by selection to drive gene clustering in fungi. Several human genes that, when disrupted, lead to inborn errors of metabolism have fungal orthologs that are clustered. For example, galactosemia arises when any one of three enzymatic steps in the galactose pathway is disrupted (*A*). Type 1 galactosemia, the most severe form, leads to accumulation of galactose-1-phosphate, a toxic inhibitor of glycolysis. In many fungi, this pathway is clustered. Similarly, tyrosinemia arises when any one of three different genes in the tyrosine catabolic pathway is disrupted, with symptoms that vary in severity according to the properties of the accumulating IC (*B*). Tyrosinemia type I, the most damaging form that frequently leads to hepatic cancer, stems from loss of the final step in the pathway, leading to accumulation of fumarylacetoacetate, a mutagenizing and alkylating IC. In many fungi, the genes participating in this pathway are also clustered. Our hypothesis (*C*) is that natural selection against the accumulation of toxic ICs results in the clustering of the genes that handle them. We tested this hypothesis on the simplest type of gene cluster, gene pairs that are neighbors in the metabolic network, i.e., share an IC, and that are also immediately adjacent neighbors on the chromosome, which we refer to as DNGPs. Topologically, the two genes composing a DNGP can have four possible orientations relative to each other on the chromosome (*D*): one divergent orientation, one convergent orientation, and two different colinear orientations depending on gene order.

orientation points to a clear molecular genetic mechanism that would favor divergent DNGP orientation to mitigate the accumulation of toxic intermediate compounds.

Thirty-one percent (297/948) of DNGPs are part of larger gene clusters comprising three or more genes, where it is impossible to have all genes divergently oriented. The canonical *Gal* pathway illustrates this challenge (Fig. 1*A*); in this pathway the protein products of both *Gal7* and *Gal10* are simultaneously required for the downstream detoxification step, but only one of the two genes (*Gal10*) can be divergently oriented with *Gal1*, the gene encoding for the enzyme that catalyzes the reaction that produces the toxic IC. Furthermore, two divergently oriented DNGPs adjacent to each other in a gene cluster necessarily imply a convergent gene pair between them. Given these constraints in the orientation of genes in large gene clusters, our observation that they exhibit lower-than-expected frequencies of colinear DNGPs ($P = 4.7e^{-5}$), higher frequencies of convergent DNGPs ($P = 4.7e^{-4}$), and nonsignificantly higher frequencies of divergent DNGPs ($P = 0.14$) suggests that divergently oriented DNGPs are favored in large gene clusters.

The toxicity of a given IC might not always be accurately reflected in its LD_{50} dose, and some ICs may lack LD_{50} annotation altogether. For example, betaine aldehyde, the IC of the betaine osmoprotectant pathway (Fig. 3), is not annotated as toxic by LD_{50}, but its aldehyde is a very reactive functional group, whose toxic effects have been used in gene transformant selection experiments (22). Thus, we evaluated whether the toxicity conferred by reactive functional groups on ICs associated with DNGPs was higher than that of background (Dataset S1, Table S3). We found that divergently and colinearly oriented DNGPs were enriched for toxic ICs more than twofold relative to background, whereas convergent DNGPs were depleted twofold (Fig. 2*B*). In addition, qualitative evaluation of functional group toxicity identified a subset of highly toxic reactive compounds handled by DNGPs that were predominantly divergently oriented, reflecting an enhanced selective advantage for divergent orientation and presumably transcription (Fig. 2*B*).

The enrichment of colinear DNGPs for reactive ICs may help explain the large drop (Fig. 2*A*) in relative toxicity at LD_{50} doses of 500 and 1,000 mg/kg. Because reactivity does not necessarily translate into a low-LD_{50} dose, most colinear DNGPs handling reactive ICs in our dataset lack corresponding LD_{50} data. This means that relative LD_{50} toxicity at these doses is evaluated for a set of colinear DNGPs that predominantly handles reactive ICs rather than ICs with toxicity captured by LD_{50} dose.

The combined annotation of IC toxicity using LD_{50} dose and functional group analysis showed that 38.5% (177/460, $P = 1.5 \times 10^{-6}$) of divergently oriented DNGPs and 34.6% (110/318, $P = 8.2 \times 10^{-3}$) of colinearly oriented DNGPs handle ICs known to be toxic (Fig. 2*C*). In contrast, only 25.3% (43/170, $P = 0.83$) of ICs associated with convergently oriented DNGPs are known to be toxic, which is even lower than the background rate of 28.3%. The differential enrichment of divergently and colinearly oriented DNGPs, as well as the marked difference in the strength of their association with toxic ICs (Fig. 2 *B* and *C*), suggests that selection to reduce toxic IC accumulation is likely mediated via both coordination of gene expression and genetic linkage. Although a large fraction of DNGPs is associated with toxic ICs, this does not rule out selection for other phenotypic traits that may also lead to DNGP formation; for example, selection for avoidance of futile metabolic cycles may be also mediated through DNGPs.

In addition to known toxicity, many ICs that lack LD_{50}-dose annotation and do not contain reactive functional groups may be overtly toxic only when accumulated intracellularly. For example, the most common DNGP in the biotin pathway, present in 43 genomes, links 8-amino-7-oxononanoate synthase to 8-amino-7-oxononanoate transaminase; 8-amino-7-oxononanoate synthase produces enantiomerically pure (*S*)-8-amino-7-oxononanoate, which racemizes rapidly under intracellular conditions, competitively inhibiting the next enzyme in the pathway (23) (Fig. S1). Mitigation of this somewhat surprising toxicity, not identified by LD_{50} dose or the presence of reactive functional groups, is also likely to be a phenotype targeted by selection.

Given the significance of the association between DNGPs and toxic ICs, the question becomes whether the toxicity directly

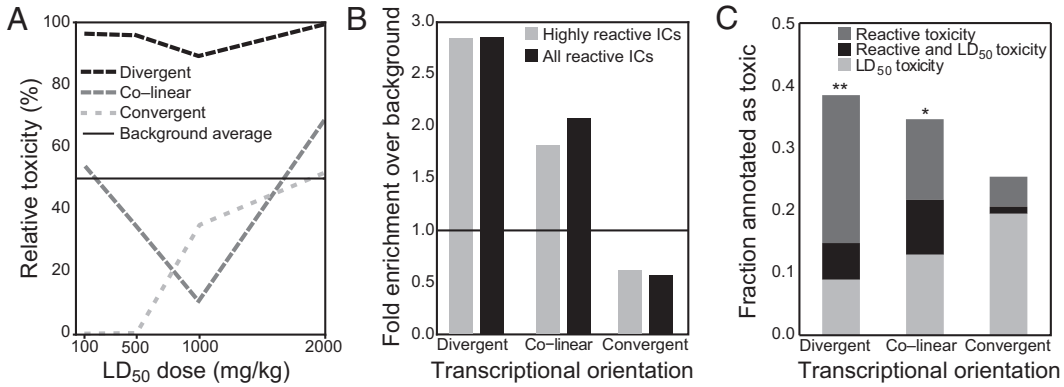

Fig. 2. DNGPs handle more toxic ICs than background. Two independent measures confirm the toxicity of ICs handled by DNGPs. (*A*) The set of divergently oriented DNGPs is more likely to have LD_{50} doses of higher toxicity than background across multiple IC LD_{50} dose cutoffs. The *x* axis shows the LD_{50} dose (mg/kg) and the *y* axis the size of divergently, colinearly, and convergently oriented DNGPs relative to the expected background. (*B*) The divergent and colinear sets of DNGPs are enriched relative to background for ICs with reactive functional groups, but the set of convergently oriented DNGPs is depleted for reactive ICs. Furthermore, divergent DNGPs are enriched for ICs with highly reactive functional groups; in contrast, colinearly oriented DNGPs are more enriched when all reactive functional groups are included. Enrichment is calculated as the fraction of ICs with reactive functional groups in each set of DNGPs divided by the fraction of all metabolic gene pairs with ICs containing reactive functional groups. (*C*) Almost 40% of the ICs handled by divergent and colinear DNGPs are known to be toxic, even with limited annotation of IC toxicity and excluding likely cases of cryptic toxicity. Thus, reduced accumulation of toxic ICs is likely to be an important phenotypic target of selection leading to DNGP formation. (Difference from background: $^{**}P = 1.5 \times 10^{-6}$; $^{*}P = 8.2 \times 10^{-3}$).

affects organism fitness. To test the phenotypic impact of toxic ICs, we assessed the rate of fungal reproduction in single-gene knockouts of all genes forming pairs of metabolic neighbors in *Saccharomyces cerevisiae*. We found that gene pairs encoding enzymes handling toxic ICs affecting the rate of growth more frequently have orthologs that are divergently oriented DNGPs in other fungi than pairs that handle nontoxic ICs (Table 2). Furthermore, in addition to galactosemia and tyrosinemia, examination of the set of fungal DNGPs identified several gene pairs whose human counterparts are involved in metabolic disease. For example, the DNGP that encodes for the enzymes isovaleryl-CoA dehydrogenase and methylcrotonyl-CoA (MCA) carboxylase, which are part of the leucine degradation pathway, is found in 41 genomes, including *A. nidulans* (24). In humans, disruption of the downstream enzyme, MCA carboxylase, leads to accumulation of 3-methylcrotonyl-CoA and 3-methylcrotonylglycine (MCG), a conjugate produced during endogenous detoxification (Fig. S2). Accumulation of excess MCA and MCG, which interferes with mitochondrial respiration in the brain (25), results in a disease characterized by several symptoms, such as failure to thrive, vomiting, metabolic acidosis, delayed mental development, and death (26). The genes encoding for the enzymes indoleamine 2,3-dioxygenase and kynureninase, which are part of the tryptophan degradation pathway and are associated with the congenital human disease hyperkynureninuria, correspond to another fungal DNGP. Disruption of kynureninase, the downstream enzyme, results in the accumulation of kynurenine and related compounds, which leads to metabolic disruption, renal disruption, hypertonia, additional neurological symptoms, and death (27). The phenotypic consequences of toxic ICs on growth and reproduction suggest that mutations that reduce their accumulation would be beneficial and likely to spread in the population.

Discussion

By examining the IC toxicity of fungal DNGPs we found that, compared with chromosomally nonadjacent metabolic neighbors, DNGPs were enriched for ICs with acutely toxic LD_{50} doses as well as for reactive functional groups. These results argue that DNGPs, and metabolic gene clusters in general, are hallmarks of selection for reduced accumulation of toxic ICs. The protective advantage offered by DNGPs could be explained either by the fact that gene colocation may safeguard against the independent loss of one gene from a pair handling an acutely toxic IC (7, 17) or, alternatively, by allowing for more precise coordination of their activity (5, 6, 11–16). In both cases, the observed phenotype would be the minimization of a cell's exposure to toxic ICs.

One part of fungal metabolism widely known to be associated with physically linked genes is secondary metabolism, which involves the production of unusual compounds, such as mycotoxins, not produced by the typically highly conserved biochemical pathways of primary metabolism. The set of DNGPs examined in this study was identified using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (28), a repository of standard biochemical pathways that does not focus on secondary metabolism. Nonetheless, our DNGP set does include gene pairs from certain specialized metabolic pathways (e.g., biotin production, nitrate assimilation, etc.) that, much like secondary metabolic pathways, have accessory functions and show sparse distributions across genomes. Thus, we hypothesize that selection for amelioration of IC toxicity also drives the clustering of secondary metabolic pathways. Indeed, the evolution of many mycotoxins is likely to have involved the stepwise modification of toxic compounds to modulate their toxicity; for example, sterigmatocystin, a highly toxic secondary metabolic compound, is the penultimate step in the production of aflatoxin, a potent natural carcinogen (29).

Table 1. DNGPs are enriched for divergent orientation and depleted for others

Orientation	Count, *n*	Expected	Fold difference	Probability (cumulative binomial)
Divergent	460	237	1.94	7.6e-55 ($n \geq 460$)
Colinear	318	474	0.67	1.1e-24 ($n \leq 318$)
Convergent	170	237	0.72	1.2e-07 ($n \leq 170$)

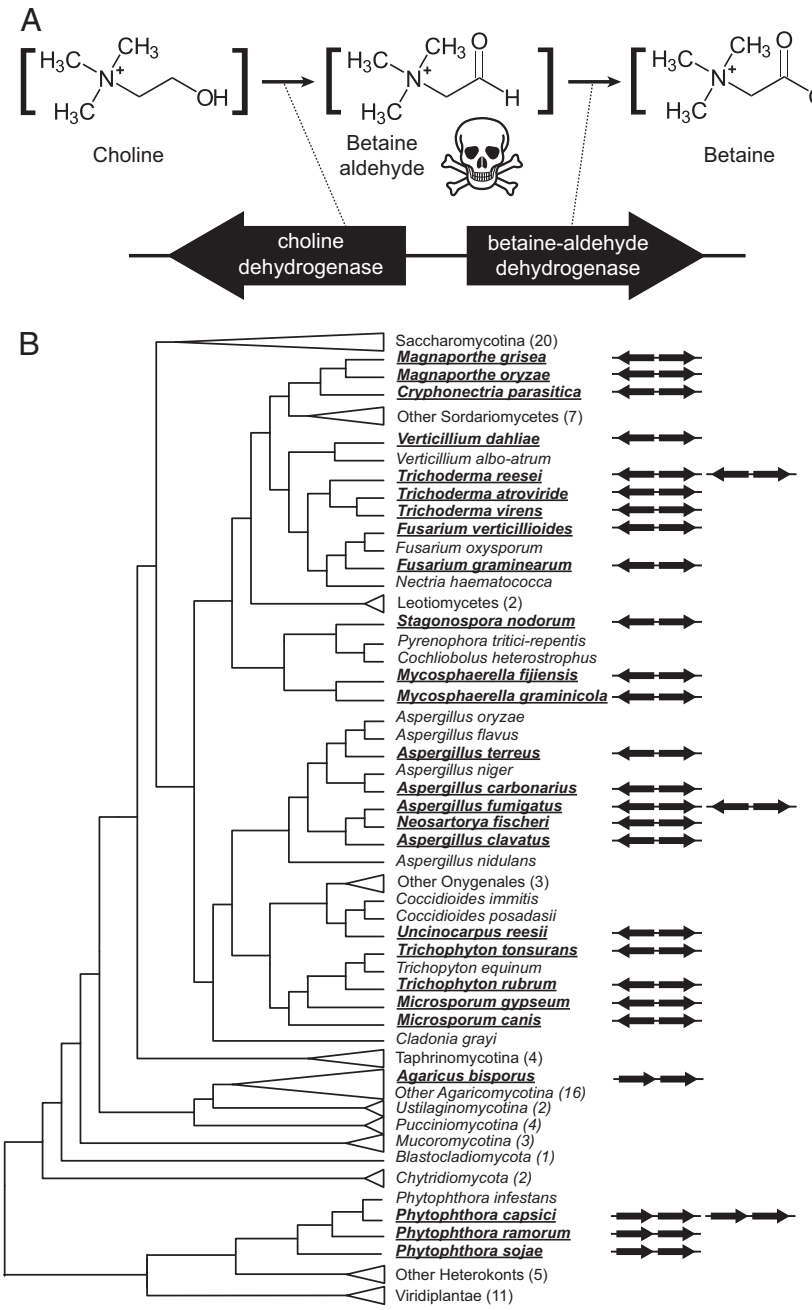

Fig. 3. The IC betaine aldehyde, produced by a DNGP, contains a highly reactive functional group. In many organisms, osmotic shock is mitigated by the formation of the osmoprotectant betaine. (*A*) This pathway, which is encoded by a DNGP, produces a reactive toxic IC, betaine aldehyde, during the conversion of choline to betaine. Betaine can reach high concentrations during osmotic stress, an indication of substantial flux that must be regulated to match environmental conditions (48). Betaine aldehyde toxicity is indicated by its use as an alternative to antibiotics for selection of transformants in plant species that lack the pathway (22). (*B*) The DNGP is divergently oriented in 23 different fungal species; 3 of these species have two DNGP copies. The *Phytophthora* oomycetes, nonfungal water molds that occupy an ecological niche similar to fungi, also have this DNGP, albeit in a colinear orientation, which suggests that their genomes have evolved in parallel to adapt to a common selection pressure imposed by intermittent high flux through this pathway.

With the distinction between selection for a phenotype versus selection of a specific molecular mechanism, we can begin to understand how inherent biochemical "pain points" in metabolic networks, such as toxic ICs, necessitate lineage (30–39) and environment-specific (20, 40) adaptations. To this end, preliminary analysis of one metazoan (*Homo sapiens*) and one plant (*Arabidopsis lyrata*) genome identified eight and four DNGPs, respectively, including two human DNGPs that match fungal DNGPs (Dataset S1, Table S4). Intriguingly, seven of the eight human DNGPs appear to be associated with hormone synthesis pathways (e.g., sex hormones) or their precursors, suggesting that toxic pain points in metazoans may reflect their increased organizational complexity and extreme reproductive sensitivity to hormonal variants. Furthermore, the paucity of human DNGPs handling ICs that have highly reactive functional groups suggests that amelioration of toxicity in metazoans may be augmented by the compartmentalization of dangerous chemistry to specific tissues, such as the liver. For example, recent work in a mouse model of tyrosinemia that contains a deficiency in

Table 2. *S. cerevisiae* gene pairs that handle growth-retarding toxic ICs form divergent DNGPs in other species

Gene pair relationship	Consistently nontoxic	Consistently toxic
Metabolic neighbor gene pairs (*S. cerevisiae*)	16	12
DNGPs (126 genomes)		
Total	3	16
Divergent	0	10
Colinear	1	3
Convergent	2	3

fumaryl-acetoacetate hydrolase (FahA), the enzyme that detoxifies the disease-causing IC, shows that selection favors liver cells that lack homogentisate dioxygenase (Hmg1), which is responsible for the production of the IC in the first place (41). Intriguingly, whereas the two genes form a divergently oriented DNGP in several fungi (Fig. 1*A*), they do not show the same organization in the human genome (21).

These results argue that amelioration of the deleterious effects of toxic ICs is a metabolic phenotype favored by selection in fungi as well as more generally across eukaryotes, resulting in DNGP formation. In turn, notwithstanding that the prevalence of DNGPs in a genome critically depends on population genetic parameters such as population size (42) and reproductive lifestyle (43), DNGPs represent signatures of selection for protection from toxic metabolic ICs.

Experimental Procedures

DNGP Identification. We annotated enzymatic functions in a local database of 100 fungal proteomes (Dataset S1, Table S5) and 26 related microbial eukaryotes by comparing each proteome with a seed database of 35 EC-annotated proteomes obtained from KEGG (28), using blastp (44, 45). Sequences that were 50–150% the length of and at least 45% conserved relative to hits were retained. Resultant homolog sets for each EC group were reduced to a more similar set using OrthoMCL (46) at an inflation rate of 2.0, retaining all Markov clusters containing seed genes. DNGPs were determined by examining all combinations of proteins that conformed to EC–EC relationships defined by KEGG for adjacent locations on chromosomes using custom perl scripts. To ensure reaction specificity, incompletely resolved EC annotations, e.g., 1.1.1.-, were excluded. Pairs of genes with overlapping EC annotation were excluded to prevent recent tandem duplications from biasing the results. DNGPs were considered part of a larger gene cluster when genes annotated with ECs in the same KEGG pathway were found on the chromosome within six genes on each side of the DNGP.

Identification of Toxic ICs. ICs between enzyme pairs were identified based on their annotation in KEGG (28) as a product or substrate of both enzymes in an EC–EC relationship pair. Water was excluded from further consideration as an IC. Oral LD_{50} data were collected from ChemIDplus (http://chem.sis.nlm.nih.gov/chemidplus/), which was downloaded January 28, 2011, and parsed using custom perl scripts. When more than one LD_{50} dose was available for an IC, the most toxic dose was used. For DNGPs with more than one IC, the IC with the most toxic dose was considered the relevant IC. Functional groups were inferred from the suffix and/or last word of IC names and were evaluated blindly by a biochemist for reactive and toxic functional groups and qualitatively assigned to one of three categories, highly reactive, moderately reactive (which, together with the highly reactive category, represents all reactive functional groups), and unknown.

Background Estimation. Using the same annotation pipeline used to identify DNGPs, we sampled 1,000 size-matched random sets of metabolic neighbors from the same genomic background without regard to their chromosomal location. Not all ICs have LD_{50} data, so the background was sampled using the same annotation limits as those for the DNGPs; i.e., metabolic neighbors were excluded from sampling if they did not have an IC with a reported LD_{50}. For each set of DNGPs with the same orientation, the toxicity of their ICs was evaluated as a percentage rank by counting ICs in each set that are more toxic than the specified dose cutoff and ranking them relative to the count of toxic ICs belonging to the background sets. Enrichment for ICs with toxic/reactive functional groups was calculated by drawing 1,000 random sets of the same size as each DNGP set from all metabolic neighbors to estimate the background rate of toxicity. Enrichment of toxic ICs was calculated by dividing the number of observed toxic ICs by the expected number of toxic ICs from the background. For this functional group analysis, all of the ICs of all metabolic pairs were considered annotated as high, moderate, or unknown toxicity. For the analysis using combined annotation, we calculated the cumulative binomial probability of toxic IC enrichment of each orientation using the background rate of toxicity observed for all possible metabolic neighbors in the genomic background (0.283).

Orientation Probability Calculation. We calculated the cumulative binomial probability for the distribution of DNGP orientation using a model of random gene orientation on a chromosome where the probability that a pair of adjacent genes have a given orientation is the following: 50% colinear, 25% convergent, and 25% divergent. A survey of multiple fungal genomes confirmed that the proportions of these orientations are generally correct.

Inference of IC Toxicity from Yeast Growth. To infer toxicity of intracellular ICs in vivo, we used growth rates of *S. cerevisiae* knockout strains grown in rich and minimal media (downloaded July 6, 2011 from http://chemogenomics.stanford.edu/supplements/01yfh/files/orfgenedata.txt) (47). We parsed KEGG annotation to identify ICs that are annotated as the product of a single enzyme and the substrate of a single enzyme, each catalyzing irreversible reactions. The knockout of either enzyme blocks production of the final product, but knocking out the downstream gene presumably causes the target IC to accumulate; therefore, we infer that an IC is toxic when the knockout strain of the downstream gene grows more slowly than the knockout strain of the upstream gene; otherwise, we consider the IC nontoxic (Fig. S3). Essential genes were considered to have a zero growth rate. To control for differences in metabolic flux through pathways in different environments, we filtered for ICs that were consistently inferred to be either toxic or nontoxic based on growth profiles of the knockout strains in both rich and minimal media. We then identified matching DNGPs with the same EC pair annotation from other genomes.

ACKNOWLEDGMENTS. We thank Brian Bachmann for estimating the reactivity and toxicity of functional groups. 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 National Science Foundation Grants DBI-0805625 (to J.C.S.) and DEB-0844968 (to A.R.).

1. Sober E (1984) *The Nature of Selection: Evolutionary Theory in Philosophical Focus* (MIT Press, Cambridge, MA).
2. Brown KS, Benson WW (1974) Adaptive polymorphism associated with multiple Müllerian mimicry in Heliconius numata (Lepid. Nymph.). *Biotropica* 6(4):205–228.
3. Joron M, Wynne IR, Lamas G, Mallet J (1999) Variable selection and the coexistence of multiple mimetic forms of the butterfly Heliconius numata. *Evol Ecol* 13:721–754.
4. Joron M, et al. (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477(7363):203–206.
5. Batada NN, Hurst LD (2007) Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet* 39(8):945–949.
6. Singer GAC, Lloyd AT, Huminiecki LB, Wolfe KH (2005) Clusters of co-expressed genes in mammalian genomes are conserved by natural selection. *Mol Biol Evol* 22(3):767–775.
7. Pál C, Hurst LD (2003) Evidence for co-evolution of gene order and recombination rate. *Nat Genet* 33(3):392–395.
8. Hurst LD, Pál C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5(4):299–310.
9. Woo YH, Li W-H (2011) Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes. *Proc Natl Acad Sci USA* 108(8):3306–3311.
10. Li Y-Y, et al. (2006) Systematic analysis of head-to-head gene organization: Evolutionary conservation and potential biological relevance. *PLOS Comput Biol* 2(7):e74.

11. Kensche PR, Oti M, Dutilh BE, Huynen MA (2008) Conservation of divergent transcription in fungi. *Trends Genet* 24(5):207–211.
12. Hurst LD, Williams EJ, Pál C (2002) Natural selection promotes the conservation of linkage of co-expressed genes. *Trends Genet* 18(12):604–606.
13. Sugino RP, Innan H (2012) Natural selection on gene order in the genome reorganization process after whole-genome duplication of yeast. *Mol Biol Evol* 29(1): 71–79.
14. Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91(3):243–248.
15. Poyatos JF, Hurst LD (2007) The determinants of gene order conservation in yeasts. *Genome Biol* 8(11):R233.
16. Wang G-Z, Lercher MJ, Hurst LD (2011) Transcriptional coupling of neighboring genes and gene expression noise: Evidence that gene orientation and noncoding transcripts are modulators of noise. *Genome Biol Evol* 3:320–331.
17. 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.
18. Leslie ND (2003) Insights into the pathogenesis of galactosemia. *Annu Rev Nutr* 23:59–80.
19. de Jongh WA, et al. (2008) The roles of galactitol, galactose-1-phosphate, and phosphoglucomutase in galactose-induced toxicity in Saccharomyces cerevisiae. *Biotechnol Bioeng* 101(2):317–326.
20. Slot JC, Rokas A (2010) Multiple GAL pathway gene clusters evolved independently and by different mechanisms in fungi. *Proc Natl Acad Sci USA* 107(22):10136–10141.
21. Peñalva MÁ (2001) A fungal perspective on human inborn errors of metabolism: alkaptonuria and beyond. *Fungal Genet Biol* 34(1):1–10.
22. 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(2):155–162.
23. Mann S, Colliandre L, Labesse G, Ploux O (2009) Inhibition of 7,8-diaminopelargonic acid aminotransferase from Mycobacterium tuberculosis by chiral and achiral anologs of its substrate: Biological implications. *Biochimie* 91(7):826–834.
24. Rodríguez JM, Ruíz-Sala P, Ugarte M, Peñalva MÁ (2004) Fungal metabolic model for 3-methylcrotonyl-CoA carboxylase deficiency. *J Biol Chem* 279(6):4578–4587.
25. Moura AP, et al. (2012) 3-Methylcrotonylglycine disrupts mitochondrial energy homeostasis and inhibits synaptic Na(+),K (+)-ATPase activity in brain of young rats. *Cell Mol Neurobiol* 32(2):297–307.
26. Arnold GL, et al. (2008) A Delphi-based consensus clinical practice protocol for the diagnosis and management of 3-methylcrotonyl CoA carboxylase deficiency. *Mol Genet Metab* 93(4):363–370.
27. Cheminal R, Echenne B, Bellet H, Duran M (1996) Congenital non-progressive encephalopathy and deafness with intermittent episodes of coma and hyperkynureninuria. *J Inherit Metab Dis* 19(1):25–30.
28. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* 32D277–D280.
29. Brown DW, et al. (1996) Twenty-five coregulated transcripts define a sterigmatocystin gene cluster in Aspergillus nidulans. *Proc Natl Acad Sci USA* 93(4):1418–1422.
30. Field B, Osbourn A (2012) Order in the playground: Formation of plant gene clusters in dynamic chromosomal regions. *Mobile Genet Elements* 2(1):46–50.
31. Kliebenstein DJ, Osbourn A (2012) Making new molecules: Evolution of pathways for novel metabolites in plants. *Curr Opin Plant Biol* 15(4):415–423.
32. Chu HY, Wegel E, Osbourn A (2011) From hormones to secondary metabolism: The emergence of metabolic gene clusters in plants. *Plant J* 66(1):66–79.
33. Osbourn A (2010) Secondary metabolic gene clusters: Evolutionary toolkits for chemical innovation. *Trends Genet* 26(10):449–457.
34. Weng J-K, Philippe RN, Noel JP (2012) The rise of chemodiversity in plants. *Science* 336(6089):1667–1670.
35. Takos AM, Rook F (2012) Why biosynthetic genes for chemical defense compounds cluster. *Trends Plant Sci* 17(7):383–388.
36. Winzer T, et al. (2012) A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* 336(6089):1704–1708.
37. Field B, et al. (2011) Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci USA* 108(38):16116–16121.
38. Field B, Osbourn AE (2008) Metabolic diversification: Independent assembly of operon-like gene clusters in different plants. *Science* 320(5875):543–547.
39. Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ (2009) CYP76M7 is an ent-cassadiene C11α-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 21(10):3315–3325.
40. Hittinger CT, Rokas A, Carroll SB (2004) Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci USA* 101(39): 14144–14149.
41. Duncan AW, et al. (2012) Aneuploidy as a mechanism for stress-induced liver adaptation. *J Clin Invest* 122(9):3307–3315.
42. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer Associates, Sunderland, MA).
43. Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. *Proc Natl Acad Sci USA* 105(12):4957–4962.
44. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389–3402.
45. Altschul SF, et al. (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS J* 272(20):5101–5109.
46. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res* 13(9):2178–2189.
47. Deutschbauer AM, et al. (2005) Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* 169(4):1915–1925.
48. Chen THH, Murata N (2011) Glycinebetaine protects plants against abiotic stress: Mechanisms and biotechnological applications. *Plant Cell Environ* 34(1):1–20.