



The Evolution of Fungal Metabolic Pathways

Jennifer H. Wisecaver^{1,9}, Jason C. Slot^{1,2,9*}, Antonis Rokas^{1*}

1 Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Department of Plant Pathology, The Ohio State University, Columbus, Ohio, United States of America

Abstract

Fungi contain a remarkable range of metabolic pathways, sometimes encoded by gene clusters, enabling them to digest most organic matter and synthesize an array of potent small molecules. Although metabolism is fundamental to the fungal lifestyle, we still know little about how major evolutionary processes, such as gene duplication (GD) and horizontal gene transfer (HGT), have interacted with clustered and non-clustered fungal metabolic pathways to give rise to this metabolic versatility. We examined the synteny and evolutionary history of 247,202 fungal genes encoding enzymes that catalyze 875 distinct metabolic reactions from 130 pathways in 208 diverse genomes. We found that gene clustering varied greatly with respect to metabolic category and lineage; for example, clustered genes in Saccharomycotina yeasts were overrepresented in nucleotide metabolism, whereas clustered genes in Pezizomycotina were more common in lipid and amino acid metabolism. The effects of both GD and HGT were more pronounced in clustered genes than in their non-clustered counterparts and were differentially distributed across fungal lineages; specifically, GD, which was an order of magnitude more abundant than HGT, was most frequently observed in Agaricomycetes, whereas HGT was much more prevalent in Pezizomycotina. The effect of HGT in some Pezizomycotina was particularly strong; for example, we identified 111 HGT events associated with the 15 *Aspergillus* genomes, which sharply contrasts with the 60 HGT events detected for the 48 genomes from the entire Saccharomycotina subphylum. Finally, the impact of GD within a metabolic category was typically consistent across all fungal lineages, whereas the impact of HGT was variable. These results indicate that GD is the dominant process underlying fungal metabolic diversity, whereas HGT is episodic and acts in a category- or lineage-specific manner. Both processes have a greater impact on clustered genes, suggesting that metabolic gene clusters represent hotspots for the generation of fungal metabolic diversity.

Citation: Wisecaver JH, Slot JC, Rokas A (2014) The Evolution of Fungal Metabolic Pathways. *PLoS Genet* 10(12): e1004816. doi:10.1371/journal.pgen.1004816

Editor: Jason E. Stajich, University of California-Riverside, United States of America

Received: July 9, 2014; **Accepted:** October 12, 2014; **Published:** December 4, 2014

Copyright: © 2014 Wisecaver et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University, Nashville, TN. This work was partially supported by funds provided by the National Science Foundation (<http://www.nsf.gov/>, grants IOS-1401682 to JHW, DBI-0805625 to JCS and DEB-0844968 to AR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: slot.1@osu.edu (JCS); antonis.rokas@vanderbilt.edu (AR)

9 These authors contributed equally to this work.

Introduction

As one of the primary decomposers of organic material in nature, fungal species catabolize a wide diversity of substrates [1], including cellulose and lignin, the two most abundant biopolymers on earth [2]. Fungi are also superb chemical engineers, capable of synthesizing a wide variety of metabolites, including amino acids, small peptides, pigments and other natural products with potent toxic activities, such as antibiotics and mycotoxins [3–6].

Fungal metabolites have historically been divided into primary, that is metabolites essential for growth and reproduction, and secondary, which include ecologically important metabolites not essential to cellular life [7,8]. However, this distinction is arbitrary when applied to metabolic pathways rather than their products not only because the essentiality of a given pathway is species-specific [9] but also because the pathways that generate primary and secondary metabolites are not mutually exclusive [10,11]. Perhaps more informatively, pathways can be divided into those shared by most organisms, which can be considered as belonging to general metabolism, and those specialized pathways that have

evolved in response to the specific ecologies of certain lineages and, as a result, are more narrowly taxonomically distributed.

An intriguing feature of specialized metabolic pathways in fungi is that constituent genes are often physically linked on chromosomes forming what are known as gene clusters [12,13]. Fungal metabolic gene clusters are distinct from the developmental gene clusters typically found in animal genomes, such as the Hox gene clusters; whereas animal gene clusters are composed of tandemly duplicated genes [14,15], fungal metabolic gene clusters comprise genes that are evolutionarily unrelated. Fungal metabolic gene clusters participate in diverse activities including nitrogen [16,17], carbohydrate [18], amino acid [19], and vitamin [12] metabolism as well as in xenobiotic catabolism [11,20] and the biosynthesis of secondary metabolites [e.g., 21–28].

Although this extraordinary metabolic diversity, whether in the form of clustered or non-clustered pathways, is integral to the entire spectrum of fungal ecological strategies (e.g., saprotrophic, pathogenic and symbiotic), we still know little about the evolutionary processes involved in its generation. Gene duplication (GD), a major source of gene innovation, is often implicated in the

Author Summary

Fungi are important primary decomposers of organic material as well as amazing chemical engineers, synthesizing a wide variety of natural products, some with potent toxic activities, including antibiotics and mycotoxins. In fungal genomes, the genes involved in these metabolic pathways can be physically linked on chromosomes, forming gene clusters. This extraordinary metabolic diversity is integral to the variety of ecological strategies that fungi employ, but we still know little about the evolutionary processes involved in its generation. To address this question, we analyzed 247,202 enzyme-encoding genes participating in hundreds of metabolic reactions from 208 diverse fungal genomes to examine how two major sources of gene innovation, namely gene duplication and horizontal gene transfer, have contributed to the evolution of clustered and non-clustered metabolic pathways. We discovered that gene duplication is the dominant and consistent driver of metabolic innovation across fungal lineages and metabolic categories; in contrast, horizontal gene transfer appears highly variable both across organisms and functions. The effects of both gene duplication and horizontal gene transfer were more pronounced in clustered genes than in their non-clustered counterparts suggesting that metabolic gene clusters are hotspots for the generation of fungal metabolic diversity.

evolution of fungal metabolism [e.g., 29–31], especially in the context of whole genome duplication (WGD) [32–34] and gene family expansion [35,36]. Notable examples include the GD of enzymes involved in organic decay [30], starch catabolism [37], degradation of host tissues [31,38,39] and toxin production [36]. Repeated rounds of GD, followed by divergence and differential gene loss, have also been invoked to explain the evolution of the gene clusters that generate the diverse alkaloids produced by plant symbiotic fungi [4]. A second key source of metabolic gene innovation in fungi is horizontal gene transfer (HGT) [40–44]; significant cases include the transfer of genes involved in xenobiotic catabolism [45,46], toxin production [45,47], degradation of plant cell walls [48,49], and wine fermentation [50]. More recently, HGT has been shown to be responsible for the transfer of entire metabolic gene clusters between unrelated fungi [11,51–58].

Although both GD and HGT have been extensively studied in fungal genomes, how these two major sources of gene innovation have interacted with clustered and non-clustered metabolic pathways and sculpted their evolution is largely unknown. To address this question, we analyzed 247,202 enzyme-encoding genes from 208 diverse fungal genomes whose protein products participate in hundreds of metabolic reactions. We found that both GD and HGT were more pronounced in clustered genes than in their non-clustered counterparts. On average, 90.0% of clustered metabolic genes underwent GD and 4.8% underwent HGT, whereas 88.1% and 2.9% of non-clustered metabolic genes experienced GD and HGT, respectively. Remarkably, some genera appear to have undergone a larger number of HGT events than entire subphyla. While the effect of GD was largely stable across metabolic categories, HGT varied extensively. These results suggest that GD is the dominant and stable process underlying fungal metabolic diversity, whereas HGT's impact is more pronounced in specific lineages and metabolic categories. The disproportionate effect of GD and HGT on clustered genes renders metabolic gene clusters into hotspots of metabolic innovation and diversification in fungi.

Results

Clustered genes in fungi vary extensively across lineages and metabolic categories

Analysis of 208 fungal genomes identified 247,202 Enzyme Commission (EC)-annotated metabolic genes (ECgenes for short), which encoded proteins catalyzing 875 distinct enzymatic reactions in 130 metabolic pathways (Figure 1; Table S1; Table S2). The percentage of the fungal proteome dedicated to metabolism was 15.4% in Saccharomycotina, 12.6% in Pezizomycotina and 8.9% in Agaricomycetes (Table S3; Figure S1).

Examination of fungal metabolism for the presence of metabolic gene clusters revealed that 3.0% (7,409) of ECgenes belonged to 3,408 distinct gene clusters, with the average genome containing 16.7 metabolic gene clusters and 36.3 clustered ECgenes (Table S3). The percentage of clustered ECgenes was highly variable across the major lineages, being more than two-fold greater in the two Ascomycota lineages, namely Pezizomycotina (3.6% of ECgenes) and Saccharomycotina (3.7%), than in Agaricomycetes (1.6%) (Figure 1, Table S3). For example, the plant pathogen *Fusarium solani* species complex species 11 (a.k.a., *Nectria haematococca*, Sordariomycetes) had 152 clustered ECgenes (representing 6.2% of its ECgenes), the most of any genome analyzed, the yeast *Torulasporea delbrueckii* (Saccharomycotina) had 59 clustered ECgenes (7.3%), whereas the ectomycorrhizal fungus *Laccaria bicolor* (Agaricomycetes) had only 14 clustered ECgenes (1.1%).

To test whether clustering was variable across fungal metabolism, we used the Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolism hierarchy [10] to assign all ECgenes to 12 overlapping, higher-order metabolic categories (CARBOHYDRATE, ENERGY, LIPID, NUCLEOTIDE, AMINO ACID, GLYCAN, COFACTOR/VITAMIN, TERPENOID/POLYKETIDE, OTHER SECONDARY METABOLITE, XENOBIOTICS, BIOSYNTHESIS OF SECONDARY METABOLITES, and MICROBIAL METABOLISM IN DIVERSE ENVIRONMENTS). We found that the proportion of clustered ECgenes varied significantly across metabolic categories (Figure 2, Table S4). For example, clustered ECgenes from all lineages were significantly overrepresented in the KEGG categories CARBOHYDRATE and TERPENOID/POLYKETIDE and underrepresented in the GLYCAN category. In addition, the proportion of clustered ECgenes in a given category often varied significantly between lineages. For example, clustered ECgenes in the NUCLEOTIDE and XENOBIOTIC categories were only significantly overrepresented in Saccharomycotina and Agaricomycetes; clustered ECgenes in the same categories were underrepresented in Pezizomycotina (Figure 2). Similarly, clustered ECgenes in the AMINO ACID and LIPID categories were underrepresented in Saccharomycotina, whereas clustered ECgenes in these same categories were overrepresented in Pezizomycotina and Agaricomycetes (Figure 2).

GD and HGT are differentially distributed across fungal lineages

To evaluate the impact of GD and HGT on fungal metabolism, we inferred GD and HGT events by reconciling the gene tree of each ECgene to the fungal species phylogeny [59–61]. Specifically, we assigned costs to GD, HGT, gene loss, and incomplete lineage sorting (ILS) and determined the most parsimonious combination of these four events to explain the ECgene tree topology given the consensus species phylogeny. Therefore, HGT events were inferred only when an ECgene tree topology was contradictory to the species phylogeny and could not be more parsimoniously

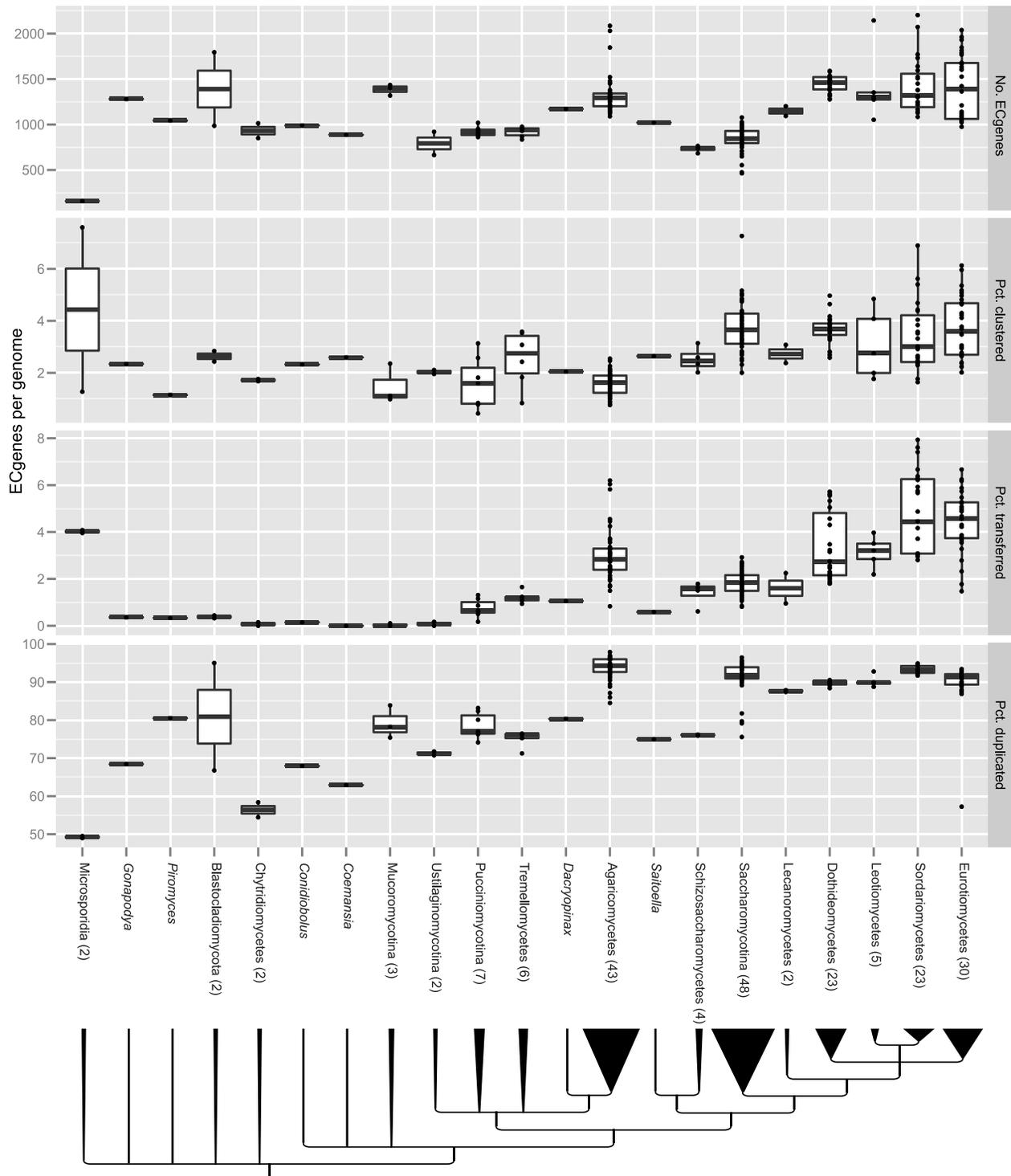


Figure 1. Variation in gene clustering, GD, and HGT across the fungal phylogeny. From top to bottom, the four box-and-whisker plots correspond to number of ECgenes per genome, percentage of clustered ECgenes per genome, percentage of horizontally transferred ECgenes per genome, and percentage of duplicated ECgenes per genome. The bottom and top of each box first and third quartiles (the 25th and 75th percentiles), respectively. The lower whisker extends from the box bottom to the lowest value within 1.5 * IQR (Inter-Quartile Range, defined as the distance between the first and third quartiles) of the first quartile. The upper whisker extends from the box top to the highest value that is within 1.5 * IQR of the third quartile. Data beyond the end of the whiskers are outliers and plotted as points. Numbers in parentheses after the lineages' names indicate numbers of genomes in each lineage; the numbers of genomes used from each lineage are also reflected by the widths of their branch triangles on the fungal species phylogeny shown at the bottom of the figure.
doi:10.1371/journal.pgen.1004816.g001

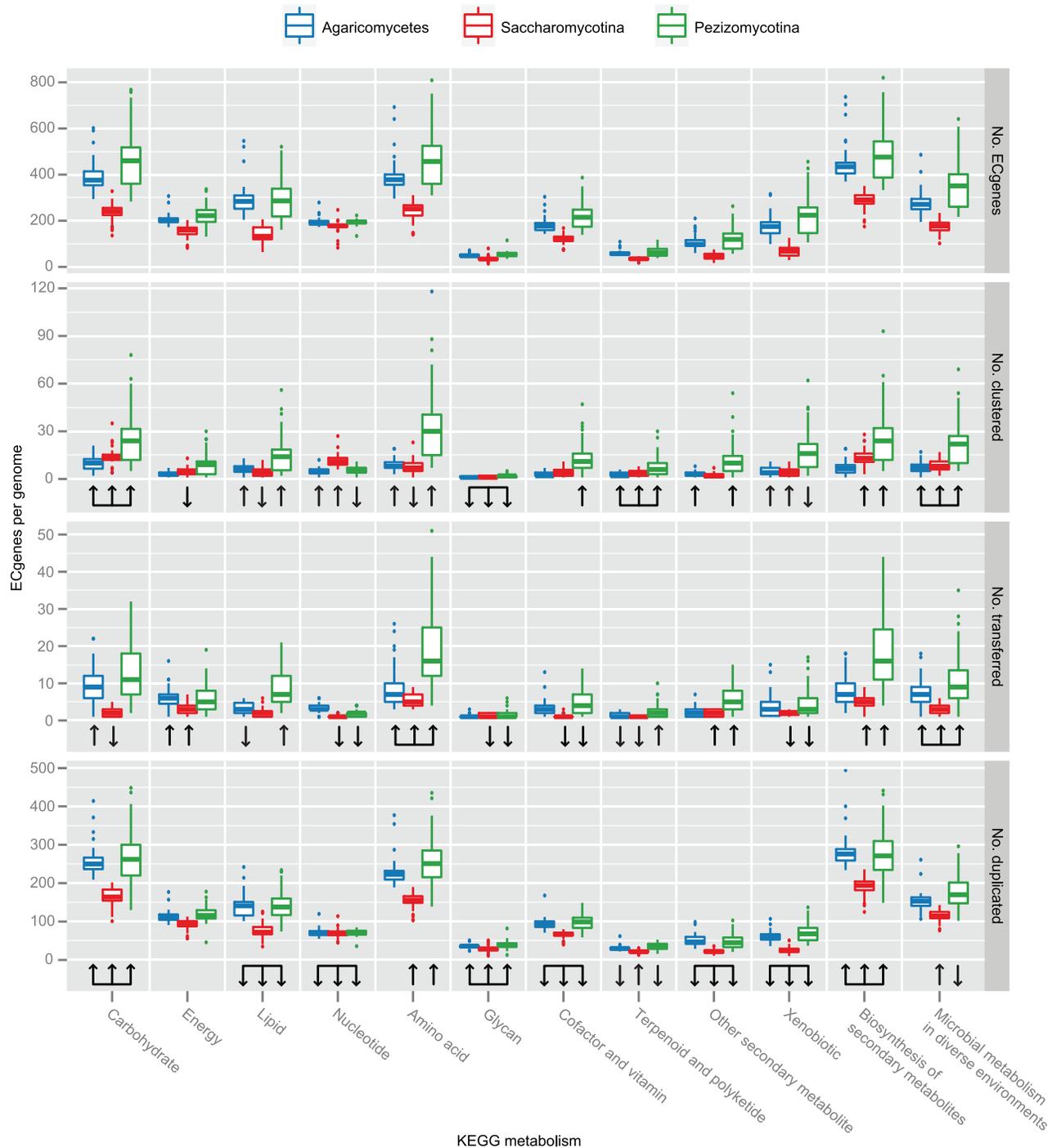


Figure 2. Over/underrepresentation of KEGG metabolic categories across three major fungal lineages. From top to bottom, the box-and-whisker plots correspond to number ECGenes per genome, number of clustered ECGenes per genome, number of transferred ECGenes per genome, and number of duplicated genes per genome. Agaricomycetes boxes are colored blue, Saccharomycotina boxes are colored red, and Pezizomycotina boxes green. Box-and-whisker convention is as described in Figure 1. Up arrows under boxes indicate overrepresentation, and down arrows indicate underrepresentation of the corresponding metabolic category in the corresponding lineage. Significance of differential representation was estimated using a two-tailed Fisher's exact test using a Benjamini & Hochberg adjusted P value ≤ 0.05 to account for multiple testing (Table S4).

doi:10.1371/journal.pgen.1004816.g002

reconciled using a combination of differential GD and gene loss. We evaluated multiple HGT costs and ultimately implemented a cost four times greater than the GD cost because it was the lowest HGT cost that recovered three published cases of HGT without

any additional (e.g., potentially spurious) cases of HGT in the corresponding ECs (Table S5).

On average, 88.7% of ECGenes per genome were inferred to have undergone one or more GD events (Table S3). This

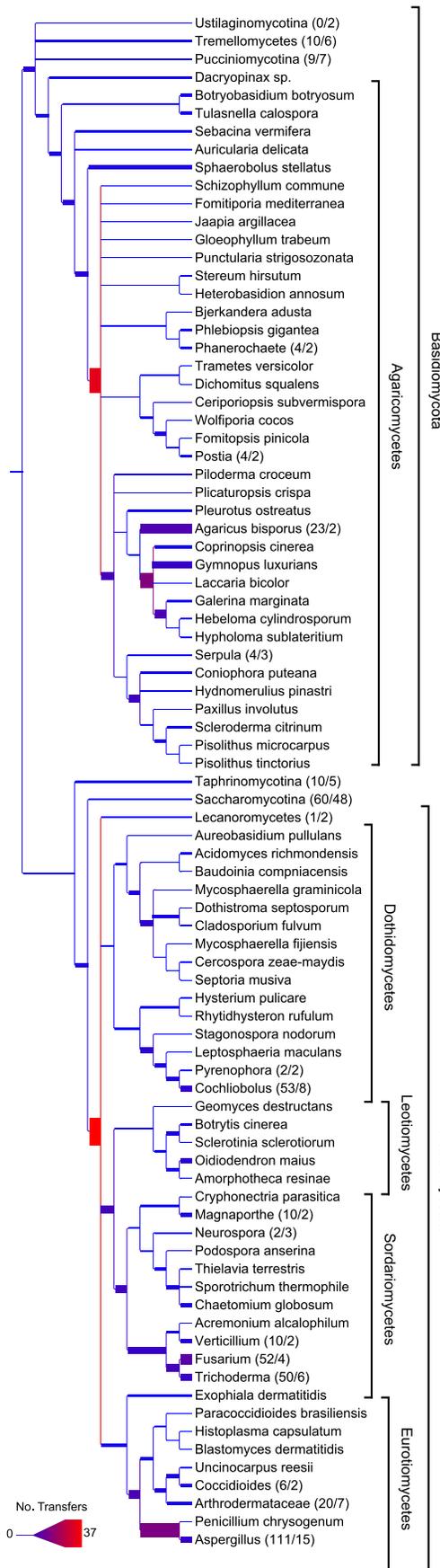


Figure 3. The episodic occurrence of HGT across the fungal species phylogeny. Numbers in parentheses indicate the number of HGT events and the number of genomes downstream of the collapsed nodes, respectively. Some clades have been collapsed for clarity; see Figure S2 for a depiction of the entire species phylogeny. The thickness and color of each branch corresponds to number of ECgenes transferred to each branch, adjusted by the number of genomes in the case of collapsed clades.
doi:10.1371/journal.pgen.1004816.g003

percentage was lower in early diverging lineages; this was the case for both taxa with typical gene densities (e.g., Chytridiomycetes) as well as for the extremely reduced microsporidians, which displayed the lowest percentages of duplicated metabolic genes (49.0% and 49.5% of ECgenes in *E. cuniculi* and *E. intestinalis*, respectively). While the low percentages of GD in microsporidians are likely explained by genome streamlining, the low percentages observed in other early diverging lineages are harder to explain, although we note that their current sparse representation in the set of sequenced fungal genomes increases the uncertainty associated with estimating GD and HGT. In contrast, 93.7% of ECgenes underwent GD in the Agaricomycetes (Figure 1), with the button mushroom, *Agaricus bisporus*, having 97.0% of its ECgenes affected by GD (704 to 722 ECgenes depending on the strain). GD percentage was also high in the Saccharomycotina (91.4%; Figure 1), including in species belonging to the *Saccharomyces sensu stricto* group, where the average increased to 95.3%, most likely as a consequence of an ancient whole genome duplication [33,62].

Our analysis also identified that on average 2.8% of ECgenes per genome had undergone one or more HGT events (Table S3), which could be traced back to 823 unique HGT events. The Pezizomycotina showed the highest percentage of HGT of all the major lineages, with an average 4.1% of ECgenes transferred per genome, and Saccharomycotina the lowest, with an average 1.8% of ECgenes transferred (Table S3; Figure 1). Remarkably, some Pezizomycotina genera showed nearly as many or more HGT events than the entire Saccharomycotina subphylum (Figure 3; Figure S2). For example, we identified 111 HGT events since the last common ancestor of the 15 *Aspergillus* species, the largest for any genus included in our analysis, but only 60 HGT events since the last common ancestor of the 48 Saccharomycotina genomes. Notwithstanding the fact that genome coverage and age are not the same across fungal genera, several other Pezizomycotina genera showed an abundance of HGT events including *Cochliobolus* (53 HGTs; 8 genomes), *Fusarium* (52 HGTs; 4 genomes), and *Trichoderma* (50 HGTs; 6 genomes). Within the Agaricomycetes, the highest concentration of HGT events was observed in the two *Agaricus bisporus* genomes (23 HGTs).

GD and HGT rates are significantly higher for clustered genes in the Pezizomycotina

Examination of the degree to which GD and HGT have differentially impacted clustered and non-clustered metabolic genes revealed significant differences (Figure 4; Table S6). On average, 90.0% of clustered ECgenes and 88.1% of non-clustered ECgenes underwent GD ($P = 4.58 \times 10^{-4}$). Similarly, 4.8% of clustered ECgenes underwent HGT compared to 2.9% of non-clustered ECgenes ($P = 4.02 \times 10^{-12}$). Examination of the impact of GD and HGT in the three major lineages shows that only in the Pezizomycotina was the percentage of GD and HGT significantly higher for clustered ECgenes than for non-clustered ECgenes (GD: 93.3% for clustered ECgenes versus 89.5% for non-clustered, $P = 1.74 \times 10^{-11}$; HGT: 6.6% for clustered ECgenes

versus 4.0% for non-clustered, $P = 2.77 \times 10^{-10}$), suggesting that the trend is largely driven by Pezizomycotina. In fact, in both Saccharomycotina and Agaricomycetes GD was more common in non-clustered ECgenes than in clustered ECgenes ($P = 0.02$ and $P = 0.01$, respectively; Figure 4). HGT was more common in Saccharomycotina non-clustered ECgenes than in clustered ones, whereas in Agaricomycetes a higher incidence of HGT events was observed in clustered ECgenes, although neither of these associations was statistically significant ($P = 0.54$ and $P = 0.16$, respectively; Table S6).

GD is consistent across fungal metabolism; HGT acts in a category- and lineage-specific manner

To test whether GD and HGT prevalence varied across fungal metabolism, we examined the rates of the two processes in each of the 12 KEGG metabolic categories across our three major lineages. We found that the effect of GD was generally consistent across metabolic categories, with 9/12 categories showing the same pattern of under/overrepresentation of duplicated ECgenes across the three lineages (Figure 2, Table S4). Specifically, the categories CARBOHYDRATE, GLYCAN, and BIOSYNTHESIS OF SECONDARY METABOLITES were overrepresented, the categories LIPID, NUCLEOTIDE, COFACTOR/VITAMIN, OTHER SECONDARY METABOLITES,

and XENOBIOTICS were underrepresented, whereas ENERGY was not differentially represented in duplicated and non-duplicated ECgenes in all three lineages.

Unlike GD, HGT differentially affected metabolic categories in a lineage-specific fashion, with 10/12 categories differing in the pattern of under/overrepresentation of duplicated ECgenes across lineages (Figure 2, Table S4). For example, ECgenes in BIOSYNTHESIS OF SECONDARY METABOLITES were overrepresented for HGT events in Pezizomycotina and Saccharomycotina, but not in Agaricomycetes. In contrast, ECgenes were overrepresented for HGT in LIPID and TERPENOID/POLYKETIDE in Agaricomycetes but underrepresented in the Pezizomycotina. Only 2 categories, AMINO ACID and MICROBIAL METABOLISM IN DIVERSE ENVIRONMENTS, were overrepresented in transferred ECgenes across all three lineages.

Discussion

Determining the relative role of GD and HGT with clustered and non-clustered metabolic pathways is important for understanding the evolution of the fungal metabolic repertoire. Examination of the synteny and evolutionary history of 247,202 ECgenes from 875 metabolic reactions across fungal diversity showed that GD is the dominant source of metabolic gene

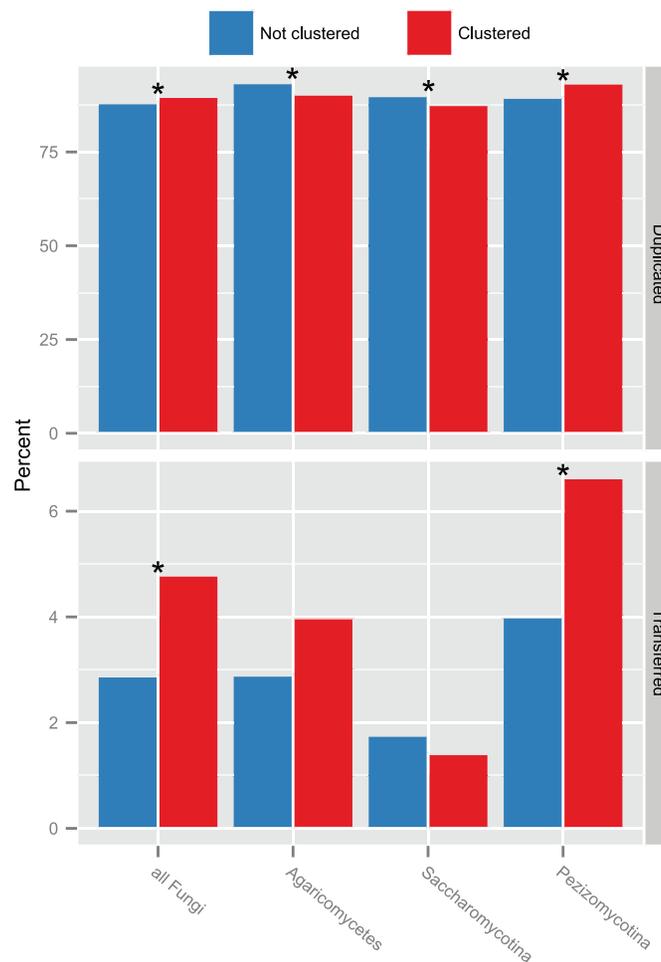


Figure 4. The association between gene innovation and gene clustering across three major fungal lineages. Percentage of non-clustered (blue bars) and clustered ECgenes (red bars) inferred to have undergone GD (top) and HGT (bottom). Asterisks (*) indicate statistically significant differences determined using a Benjamini & Hochberg adjusted P value ≤ 0.05 in a two-tailed Fisher's exact test (Table S6). doi:10.1371/journal.pgen.1004816.g004

innovation in fungi, whereas HGT is variable across metabolic categories and fungal lineages. Both GD and HGT are more pronounced in clustered genes than in their non-clustered counterparts, suggesting that metabolic gene clusters can act as hotspots for the generation of fungal metabolic innovation.

GD and HGT are sources of genetic novelty

On average 88.7% of fungal ECgenes retain the signature of one or more GD events in their ancestry compared to only 2.8% for HGT (Table S3). Even though these percentages are not directly comparable because reconciliation of ECgene histories with the species phylogeny requires that costs are assigned for every inferred GD or HGT event [60], our finding that nearly nine out of every ten metabolic genes have undergone GD suggests that this is the dominant source of gene innovation underlying fungal metabolism. These results are consistent with the hypothesis that specialized metabolic pathways evolve via GD from general metabolic precursors. Support for this hypothesis has come from phylogenetic analysis of single gene families [63,64] such as the polyketide synthases, which share a common evolutionary origin with the fatty acid synthases of general metabolism [65]. Further diversification of genes involved in specialized pathways may occur through additional duplication, functional divergence and differential loss in response to variable ecological pressures as has been proposed for polyketide, nonribosomal peptide and alkaloid biosynthesis genes [4,66–68].

Our analysis showed that certain lineages in the Pezizomycotina and Agaricomycetes have increased HGT rates. Interestingly, bacteria-to-fungi HGT events are also elevated within Pezizomycotina, particularly in *Fusarium* and *Aspergillus* genomes [43]. HGT of entire chromosomes has been reported in *Fusarium* [69,70], a genus in our analysis, which in addition to *Aspergillus*, *Cochliobolus* and *Magnaporthe*, appears not only receptive to HGT but also includes highly virulent plant and animal pathogens, ecological lifestyles associated with many known cases of HGT [11,45,47,51,69–71]. Similarly, mycoparasitism in the genus *Trichoderma* may also provide ecological opportunities for fungal-to-fungal HGT.

GD alone or in combination with HGT affected nearly every reaction in fungal metabolism (727, 95.7% of ECs that passed the phylogenomic analysis; Figure 5). The effect of both GD and HGT varied between metabolic categories, suggesting that some pathways may tolerate the introduction of new genes better than others. One possible explanation for this variation is that the metabolic networks associated with the different functional categories have different degrees of connectivity. Genes whose products make up large protein complexes or that have many interacting partners exhibit less variation in copy number [35], perhaps because unbalanced increases in gene dosage can lead to malformed protein complexes and a buildup of toxic intermediates in metabolic pathways [72–74], and might be less likely to undergo GD [75,76] as well as HGT [77]. In addition to gene dosage effects, deleterious interactions between native and horizontally acquired proteins that function as parts of multi-protein complexes, and as a consequence have distinct co-evolutionary histories, are likely also important barriers to HGT [77,78].

Another possible explanation is that the source of the variation of GD and HGT lies in the differing functions encoded by these metabolic categories. Gene innovation is often correlated with molecular function, with informational genes such as those involved in DNA replication, transcription and translation duplicated and transferred less often than metabolic genes [35,76,78]. Within metabolism, one might expect that widely distributed pathways involved in universal metabolic functions,

such as oxidative phosphorylation and the citric acid cycle, are more likely to be functionally constrained and, as a consequence, less likely to tolerate GD or HGT of their constituent genes. In contrast, GD and HGT might be more advantageous for specialized metabolic pathways that are under strong selection in fluctuating environments [11].

33 EC reactions are associated with 332 ECgenes that are never duplicated or transferred in our analysis; 31 of these 33 reactions (93.9%) are also never clustered (Table S7a). For the majority of these ECs, the reason for the apparent lack of GD or HGT is because they are represented by only a few ECgenes in our analysis; therefore, their ECgene trees consist of few taxa with topologies in agreement with the consensus species phylogeny. For other EC reactions in this set, strong selection pressure to maintain a single, native gene copy could explain the lack of GD and HGT. Only three genes annotated with EC reaction numbers and which were never duplicated or transferred in our analysis were present in the *Saccharomyces cerevisiae* genome (YNL219C [2.4.1.259], YBR003W [2.5.1.83], and YPR184W [3.2.1.33]). When examined against the yeast phenotype and interaction data from the *Saccharomyces* Genome Database (<http://www.yeastgenome.org>), these three genes displayed a variety of phenotypes and all their null mutants were viable (Table S7b). Interestingly, overexpression of two of the ECgenes (YNL219C [2.4.1.259] and YBR003W [2.5.1.83]) resulted in reduced rate of vegetative growth in *S. cerevisiae* (Table S7b), suggesting that the acquisition of additional gene copies through GD or HGT could be disadvantageous. Furthermore, one *S. cerevisiae* ECgene, a glycosyltransferase (YNL219C [2.4.1.259]) involved in the biosynthesis of asparagine-linked glycans, has a very complex interaction network of 315 described physical and genetic interactions (Table S6a), which could serve as an additional barrier to GD and HGT.

Gene clusters are hotspots for metabolic novelty

3.0% of fungal genes examined in our study lie within gene clusters. This is likely a conservative estimate because ECgene annotation is better for general rather than specialized metabolism. Although our analysis includes many specialized pathways (Table S2), such as biotin production (KEGG map00780), nitrate assimilation (map00910) and terpenoid backbone biosynthesis (map00900), and the fraction of enzymatic reactions encoded by clustered ECgenes is extensive (441 reactions, 50.4% of ECs; Figure 5), lineage-specific genes involved in specialized metabolic pathways are less likely to be included. In addition, fungal metabolic gene clusters are often identified through the presence of one or more conserved synthesis genes (e.g., genes encoding polyketide synthase or nonribosomal peptide synthase enzymes); proper demarcation of associated genes encoding modifying enzymes (e.g., oxidases and transferases) is challenging because they often lack functional annotation and are lineage-specific, leading to underestimates of gene cluster size.

Gene clustering in fungi is positively associated with both GD and HGT, but this pattern appears to be driven by Pezizomycotina ECgenes (Figure 4). Saccharomycotina ECgenes cluster more often than the global fungal average but are less often affected by HGT, whereas Agaricomycetes display the opposite trend; they experience more HGT but less gene clustering (Figure S3). GD affects nearly all ECgenes, and this large sample size undoubtedly contributes to the statistical significance of its association with gene clustering, even though the fold increase in the percentage of GD events observed in clustered versus non-clustered ECgenes is only 1.02. In contrast, the effect of HGT on clustered genes is 1.66 fold greater than its effect on non-clustered genes.

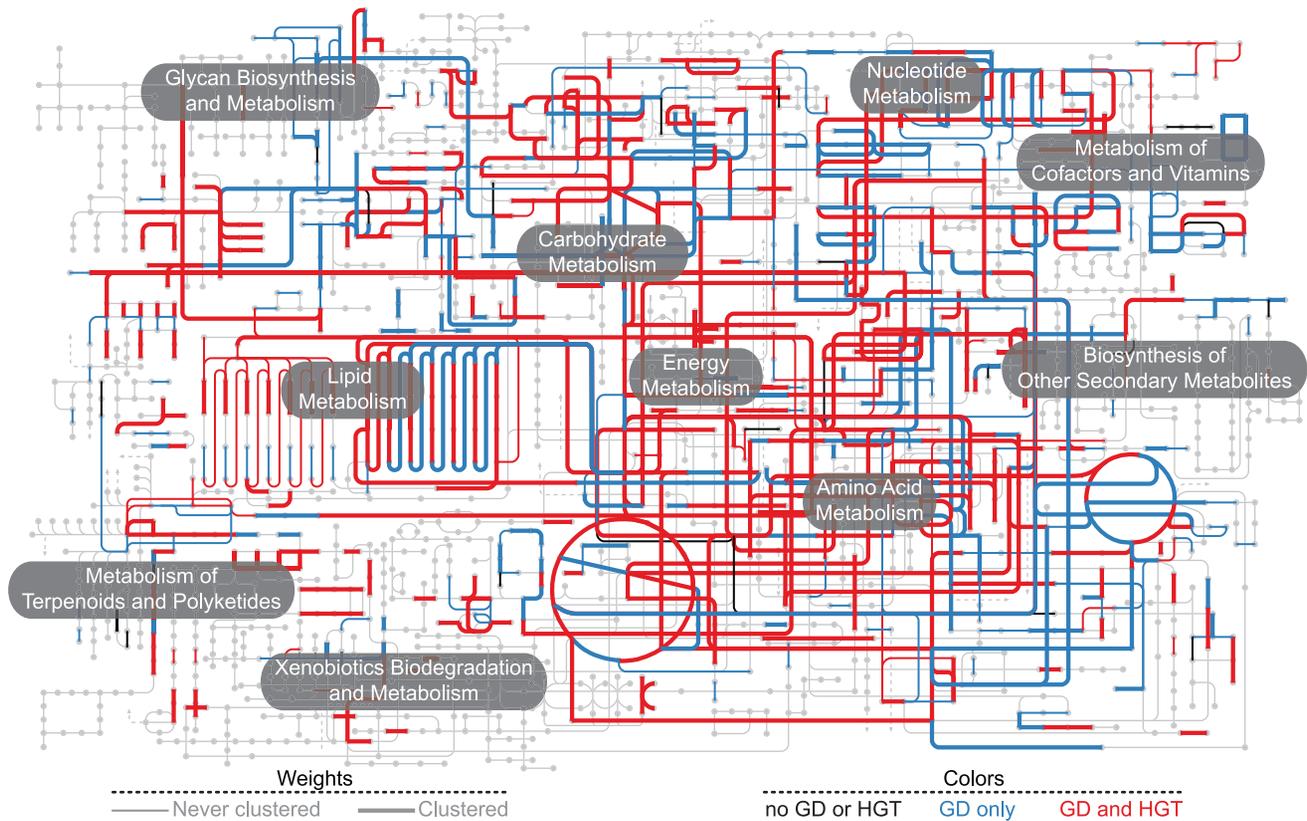


Figure 5. The fungal metabolic network of interactions between gene clustering and two major sources of gene innovation (GD and HGT). Nodes of the metabolic network correspond to KEGG compounds. Thick edges of the metabolic network correspond to EC numbers from clustered ECgenes in one or more fungal species, whereas thin edges to EC numbers whose genes show no history of gene clustering. Colored edges correspond to EC numbers whose ECgenes have undergone HGT and GD (red), GD (blue), or show no history of GD or HGT (black). Note that none of the EC numbers in our dataset were affected by HGT alone. Pathway map created using iPATH2.0 [95]. doi:10.1371/journal.pgen.1004816.g005

The uniqueness and wide distribution of fungal metabolic gene clusters has given rise to many models that attempt to explain their formation and maintenance [53,79–83]. For example, the selfish gene cluster model proposes that HGT allows gene clusters to avoid being lost by facilitating colonization of new genomes [84,85]. Although several instances of HGT of fungal gene clusters have been discovered in recent years [11,51–58], clustered pathways are also more likely to be lost than non-clustered ones [53]. The small percentage of clustered genes affected by HGT in our analysis (4.8%), albeit larger than the background percentage of transferred un-clustered genes (2.9%), suggests that selfishness is unlikely to be the predominant mechanism driving gene cluster formation and maintenance in fungi. Nevertheless, the association between metabolic gene clusters and GD/HGT suggests that gene clustering can facilitate the duplication and transfer of entire metabolic pathways. This is consistent with the view that the barriers to gene innovation acting on gene clusters may be lower than those acting on single genes because the latter undergo GD or HGT in the absence of their functional partners.

Materials and Methods

Enzyme annotation

A custom enzyme classification pipeline assigned EC numbers to protein-coding genes from the genomes of 208 fungi and 9 stramenopiles (five oomycetes and four algal relatives), which were

included in this analysis because of published reports of HGT between oomycetes and fungi [44]. Each gene was queried against a database of KEGG orthology (KO)-annotated proteins from 53 KEGG Organisms (Table S8) using UBLAST (<http://drive5.com/ublast>) with an accel setting of 0.7 and minimum identity cutoff of 0.3. A KO term was assigned to the query for UBLAST hits with greater than 80% sequence identity and no more than 10% difference in length. In cases where highly similar matches were not recovered, KO terms were assigned to query sequences with respect to the UBLAST hits showing the lowest e-values; all UBLAST hits that followed the first e-value increase of 10^{-50} or greater were excluded. EC numbers were assigned according to KO term (http://www.genome.jp/kegg-bin/get_htext?ko00001.keg).

Detection of fungal metabolic gene clusters

Fungal proteomes were screened for metabolic gene clusters as described [81]. Briefly, two ECgenes were considered clustered if they were separated by no more than 6 intervening genes according to published annotation and their EC numbers were nearest neighbors in one or more KEGG pathways. Gene clusters were inferred by joining overlapping metabolic gene pair ranges that were separated by no more than 6 intervening genes; the cutoff of 6 intervening genes was determined empirically with reference to previous analyses of both primary [52,53] and secondary [54] metabolism clusters.

Phylogenetic reconstruction and gene tree-species phylogeny reconciliation

We constructed a draft fungal species phylogeny using protein sequences of the widely used DNA-directed RNA polymerase II subunit RPB2 marker, which were aligned with MAFFT using the E-INS-i strategy [86]. The resulting alignment was trimmed with TRIMAL using the automated1 strategy [87], and the topology was inferred using maximum likelihood (ML) as implemented in RAXML version 7.2.8 [88] using a PROTGAMMALGF substitution model and rapid bootstrapping (100 replications). Branches with bootstrap support less than 50 were collapsed using the Consense module in the PHYLIP program [89]. The final bifurcating and consensus (multifurcating) species phylogenies (File S1) were constructed by making targeted corrections to the RPB2 topology based on published literature (Table S9).

ECgene trees were constructed using a custom phylogenomic pipeline (Figure S4). Guide trees were first constructed for each ECgene family with MAFFT using the scores of pairwise global alignments [86] and rooted with the NOTUNG rooting optimization algorithm using event parsimony. This distance-based guide tree and the consensus species phylogeny were used to delineate groups of homologs by aiming to maximize taxonomic diversity while minimizing the number of paralogs in each gene tree. The ECgene sequences from each one of these groups of homologs were then extracted in FASTA format for phylogenomic analysis. FASTA files of ECgenes with less than 4 or more than 1000 sequences were excluded. Sequences were aligned in MAFFT using the auto strategy selection [86]. Alignments were trimmed in TRIMAL using the automated1 trimming strategy [87], and trimmed alignments shorter than 150 amino acid residues were discarded. Phylogenetic trees were constructed using FASTTREE [90] with a WAG+CAT amino acid model of substitution, 1000 resamples, four rounds of minimum-evolution subtree-prune-regraft moves (-spr 4), and the more exhaustive ML nearest-neighbor interchange option enabled (-mlacc 2 -slownni).

Gene tree-species phylogeny reconciliation was performed in NOTUNG using its duplication, transfer, loss and ILS aware parsimony-based algorithm [59–61,91]. Ambiguity in the fungal species phylogeny and low branch support in ECgene trees were handled through a multi-step approach. First, ECgene tree branches with less than 0.90 SH-like local support were collapsed using TREECOLLAPSERCL v4 (<http://emmahodcroft.com/TreeCollapseCL.html>). This collapsed ECgene tree was rooted and its polytomies resolved against the bifurcating species phylogeny. This resolved ECgene tree was then reconciled to the multifurcating, consensus species phylogeny using a duplication cost of 1.5, loss cost of 1 and ILS cost of 0. Transfer costs of 2, 4, 6, 8, 10 and 12 as well as the option to prune taxa not present in the gene tree from the species phylogeny were evaluated. A transfer cost of 6 with the prune option enabled best recovered published cases of HGT between fungi (Table S5). Percent GD and HGT were expressed over the 152,835 fungal ECgenes that passed this reconciliation pipeline. Because a single ancestral HGT event could be recorded in multiple ECgene trees, we defined unique HGT events as all cases where ECgenes assigned to the same EC number were inferred to have undergone HGT to/from the same recipient/donor nodes in the species phylogeny.

Statistical analyses

Fisher's exact tests were performed using the R function `fisher.test` with a two-sided alternative hypothesis [92]. *P* values were adjusted for multiple comparisons using the R function `p.adjust` with the Benjamini & Hochberg (BH) method [93]. Box-and-whisker plots were created using the R plotting system `ggplot2` [94].

Supporting Information

Figure S1 Variation in gene clustering, HGT, and GD across fungal lineages, expanded version. From top to bottom, the four box-and-whisker plots correspond to number of ECgenes per genome, percentage of clustered ECgenes per genome, percentage of horizontally transferred ECgenes per genome, and percentage of duplicated ECgenes per genome. Box-and-whisker convention is as described in Figure 1. Numbers in parentheses after the lineages' names indicate numbers of genomes in each lineage; the numbers of genomes used from each lineage are also reflected by the widths of their branch triangles on the fungal species phylogeny shown at the bottom of the figure.

(PDF)

Figure S2 HGT across fungal species phylogeny, expanded version. Numbers above branches indicate number of HGT events predicted to have occurred onto each branch. The thickness and color of each branch corresponds to number of ECgenes transferred to each branch.

(PDF)

Figure S3 Incidence of gene clustering, GD and HGT mapped onto the global metabolism networks of Pezizomycotina, Saccharomycotina and Agaricomycetes. Nodes of the metabolic network correspond to KEGG compounds. Thick edges of the metabolic network correspond to EC numbers from clustered ECgenes in one or more fungal species, whereas thin edges to EC numbers whose genes show no history of gene clustering. Colored edges correspond to EC numbers whose ECgenes have undergone HGT and GD (red), GD only (blue), HGT only (green), or show no history of GD or HGT (black). Pathway maps created using iPATH2.0 [95].

(PDF)

Figure S4 Phylogenomics pipeline. A schematic diagram showing the functional components and data flow of the phylogenomics pipeline and gene tree-species phylogeny reconciliation.

(PDF)

Table S1 List of genomes used.

(XLSX)

Table S2 List of KEGG categories and pathways used.

(XLSX)

Table S3 Average gene clustering, GD and HGT per genome.

(XLSX)

Table S4 Fisher's exact tests for over/underrepresentation of KEGG metabolic categories in ECgene subsets.

(XLSX)

Table S5 Number of inferred HGT events in different iterations of the pipeline vs published literature.

(XLSX)

Table S6 Fisher's exact tests for association between sources of gene innovation (i.e., GD or HGT) and gene clustering.

(XLSX)

Table S7 Analysis of yeast phenotype and interaction data. a) List of EC reactions associated with genes that are never duplicated or transferred in the NOTUNG analysis with corresponding gene name in *S. cerevisiae* and number of protein interactions where available. b) Phenotype data from the *Saccharomyces* Genome Database.

(XLSX)

Table S8 List of KEGG organisms used for ECgene annotation.

(XLSX)

Table S9 Curation to species phylogeny with references. (DOCX)

File S1 Species phylogeny in newick format. Tree 1: RAXML best tree of RBP2. Tree 2: Consense majority rule phylogeny. Tree 3: Curated consensus species phylogeny. Tree 4: Curated bifurcating species phylogeny. (DOCX)

References

- Wainwright M (1988) Metabolic diversity of fungi in relation to growth and mineral cycling in soil - a review. *Trans Br Mycol Soc* 90: 159–170.
- Bouws H, Wattenberg A, Zorn H (2008) Fungal secretomes-nature's toolbox for white biotechnology. *Appl Microbiol Biotechnol* 80: 381–388. doi:10.1007/s00253-008-1572-5.
- Hoffmeister D, Keller N (2007) Natural products of filamentous fungi: enzymes, genes, and their regulation. *Nat Prod Rep* 24: 393–416. doi:10.1039/b603084j.
- Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, et al. (2013) Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. *PLoS Genet* 9: e1003323. doi:10.1371/journal.pgen.1003323.s012.
- Dufossé L, Fouillaud M, Caro Y, Mapari SA, Sutthiwong N (2014) Filamentous fungi are large-scale producers of pigments and colorants for the food industry. *Curr Opin Biotechnol* 26C: 56–61. doi:10.1016/j.copbio.2013.09.007.
- Kohlhaw GB (2003) Leucine biosynthesis in fungi: entering metabolism through the back door. *Microbiol Mol Biol Rev* 67: 1–. doi:10.1128/MMBR.67.1.1-15.2003.
- Demain AL, Fang A (2000) The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol* 69: 1–39.
- Keller N, Turner G, Bennett J (2005) Fungal secondary metabolism-from biochemistry to genomics. *Nat Rev Microbiol* 3: 937–947. doi:10.1038/nrmicro1286.
- Koonin EV (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol* 1: 127–136. doi:10.1038/nrmicro751.
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, et al. (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36: D480–D484. doi:10.1093/nar/gkm882.
- Greene GH, McGary KL, Rokas A, Slot JC (2014) Ecology drives the distribution of specialized tyrosine metabolism modules in fungi. *Genome Biol Evol* 6: 121–132. doi:10.1093/gbe/evt208.
- Hall C, Dietrich FS (2007) The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics* 177: 2293–2307. doi:10.1534/genetics.107.074963.
- Keller N, Hohn T (1997) Metabolic pathway gene clusters in filamentous fungi. *Fungal Genet Biol* 21: 17–29.
- Holland PWH (2013) Evolution of homeobox genes. *Wiley Interdiscip Rev Dev Biol* 2: 31–45. doi:10.1002/wdev.78.
- Irimia M, Maeso I, Garcia-Fernández J (2008) Convergent evolution of clustering of Iroquois homeobox genes across metazoans. *Mol Biol Evol* 25: 1521–1525. doi:10.1093/molbev/msn109.
- Jargeat P, Rekanalt D, Verner M, Gay G, Debaud J, et al. (2003) Characterisation and expression analysis of a nitrate transporter and nitrite reductase genes, two members of a gene cluster for nitrate assimilation from the symbiotic basidiomycete *Hebeloma cylindrosporum*. *Current Genetics* 43: 199–205. doi:10.1007/s00294-003-0387-2.
- Wong S, Wolfe KH (2005) Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet* 37: 777–782. doi:10.1038/ng1584.
- Hittinger CT, Rokas A, Carroll SB (2004) Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci U S A* 101: 14144–14149. doi:10.1073/pnas.0404319101.
- Hull EP, Green PM, Arst HN, Scazzocchio C (1989) Cloning and physical characterization of the L-proline catabolism gene cluster of *Aspergillus nidulans*. *Mol Microbiol* 3: 553–559.
- Bobrowicz P, Wysocki R, Owsianik G, Goffeau A, Ulaszewski S (1997) Isolation of three contiguous genes, ACR1, ACR2 and ACR3, involved in resistance to arsenic compounds in the yeast *Saccharomyces cerevisiae*. *Yeast* 13: 819–828.
- Subazini TK, Kumar GR (2011) Characterization of Lovastatin biosynthetic cluster proteins in *Aspergillus terreus* strain ATCC 20542. *Bioinformatics* 6: 250–254.
- Bushley KE, Raja R, Jaiswal P, Cumbie JS, Nonogaki M, et al. (2013) The genome of *Tolyposcladium inflatum*: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster. *PLoS Genet* 9: e1003496. doi:10.1371/journal.pgen.1003496.
- Gardiner DM, Cozijnsen AJ, Wilson LM, Pedras MSC, Howlett BJ (2004) The sirodesmin biosynthetic gene cluster of the plant pathogenic fungus *Leptosphaeria maculans*. *Mol Microbiol* 53: 1307–1318. doi:10.1111/j.1365-2958.2004.04215.x.
- Yu J, Chang PK, Ehrlich KC, Cary JW, Bhatnagar D, et al. (2004) Clustered pathway genes in aflatoxin biosynthesis. *Appl Environ Microbiol* 70: 1253. doi:10.1128/AEM.70.3.1253-1262.2004.
- Tudzynski P, Höltzer K, Correia T, Arntz C, Grammel N, et al. (1999) Evidence for an ergot alkaloid gene cluster in *Claviceps purpurea*. *Mol Gen Genet* 261: 133–141.
- Ahn J-H, Cheng Y-Q, Walton JD (2002) An extended physical map of the TOX2 locus of *Cochliobolus carbonum* required for biosynthesis of HC-toxin. *Fungal Genet Biol* 35: 31–38. doi:10.1006/fgbi.2001.1305.
- Brown DW, McCormick SP, Alexander NJ, Proctor RH, Desjardins AE (2001) A genetic and biochemical approach to study trichothecene diversity in *Fusarium sporotrichioides* and *Fusarium graminearum*. *Fungal Genet Biol* 32: 121–133. doi:10.1006/fgbi.2001.1256.
- Smith DJ, Burnhap MK, Bull JH, Hodgson JE, Ward JM, et al. (1990) Beta-lactam antibiotic biosynthetic genes have been conserved in clusters in prokaryotes and eukaryotes. *Embo J* 9: 741–747.
- Hittinger CT, Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449: 677–U1. doi:10.1038/nature06151.
- Floudas D, Binder M, Riley R, Barry K, Blanchette RA, et al. (2012) The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science* 336: 1715–1719. doi:10.1126/science.1221748.
- Powell AJ, Conant GC, Brown DE, Carbone I, Dean RA (2008) Altered patterns of gene duplication and differential gene gain and loss in fungal pathogens. *BMC Genomics* 9: 147. doi:10.1186/1471-2164-9-147.
- Ma L-J, Ibrahim AS, Skory C, Grabherr MG, Burger G, et al. (2009) Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet* 5: e1000549. doi:10.1371/journal.pgen.1000549.
- Kellis M, Birren BW, Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428: 617–624. doi:10.1038/nature02424.
- Wolfe K (2004) Evolutionary genomics: Yeasts accelerate beyond BLAST. *Curr Biol* 14: R392–R394. doi:10.1016/j.cub.2004.05.015.
- Wapinski I, Pfeffer A, Friedman N, Regev A (2007) Natural history and evolutionary principles of gene duplication in fungi. *Nature* 449: 54–61. doi:10.1038/nature06107.
- Cornell MJ, Alam I, Soanes DM, Wong HM, Hedeler C, et al. (2007) Comparative genome analysis across a kingdom of eukaryotic organisms: Specialization and diversification in the Fungi. *Genome Res* 17: 1809–1822. doi:10.1101/gr.6531807.
- Hunter AJ, Jin B, Kelly JM (2011) Independent duplications of alpha-amylase in different strains of *Aspergillus oryzae*. *Fungal Genet Biol* 48: 438–444. doi:10.1016/j.fgb.2011.01.006.
- Xu J, Saunders CW, Hu P, Grant RA, Boekhout T, et al. (2007) Dandruff-associated *Malassezia* genomes reveal convergent and divergent virulence traits shared with plant and human fungal pathogens. *Proc Natl Acad Sci U S A* 104: 18730–18735. doi:10.1073/pnas.0706756104.
- Joneson S, Stajich JE, Shiu S-H, Rosenblum EB (2011) Genomic transition to pathogenicity in chytrid fungi. *PLoS Pathog* 7: e1002338. doi:10.1371/journal.ppat.1002338.
- League GP, Slot JC, Rokas A (2012) The *ASP3* locus in *Saccharomyces cerevisiae* originated by horizontal gene transfer from *Wickerhamomyces*. *FEMS Yeast Research* 12: 859–863. doi:10.1111/j.1567-1364.2012.00828.x.
- Hall C, Brachat S, Dietrich FS (2005) Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryotic Cell* 4: 1102–1115. doi:10.1128/EC.4.6.1102-1115.2005.
- Richards TA, Soanes DM, Foster PG, Leonard G, Thomson CR, et al. (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. *Plant Cell* 21: 1897–1911. doi:10.1105/tpc.109.065805.
- Marcel-Houben M, Gabaldon T (2010) Acquisition of prokaryotic genes by fungal genomes. *Trends Genet* 26: 5–8. doi:10.1016/j.tig.2009.11.007.
- Richards TA, Dacks JB, Jenkinson JM, Thomson CR, Talbot NJ (2006) Evolution of filamentous plant pathogens: gene exchange across eukaryotic kingdoms. *Curr Biol* 16: 1857–1864. doi:10.1016/j.cub.2006.07.052.
- Gardiner DM, McDonald MC, Covarelli L, Solomon PS, Rusu AG, et al. (2012) Comparative pathogenomics reveals horizontally acquired novel virulence genes

Acknowledgments

We thank Haley Eidem, Abigail Lind, Kris McGary, and Patricia Soria for helpful discussions.

Author Contributions

Conceived and designed the experiments: JHW JCS AR. Performed the experiments: JHW JCS. Analyzed the data: JHW JCS AR. Contributed reagents/materials/analysis tools: JHW JCS. Wrote the paper: JHW JCS AR.

- in fungi infecting cereal hosts. *PLoS Pathog* 8: e1002952. doi:10.1371/journal.ppat.1002952.
46. Tiburcio RA, Lacerda Costa GG, Carazzolle MF, Costa Mondego JM, Schuster SC, et al. (2010) Genes acquired by horizontal transfer are potentially involved in the evolution of phytopathogenicity in *Moniliophthora perniciosa* and *Moniliophthora roreri*, two of the major pathogens of cacao. *J Mol Evol* 70: 85–97. doi:10.1007/s00239-009-9311-9.
 47. Friesen TL, Stukenbrock EH, Liu Z, Meinhardt S, Ling H, et al. (2006) Emergence of a new disease as a result of interspecific virulence gene transfer. *Nat Genet* 38: 953–956. doi:10.1038/ng1839.
 48. Sun B-F, Xiao J-H, He S, Liu L, Murphy RW, et al. (2013) Multiple interkingdom horizontal gene transfers in *Pyrenophora* and closely related species and their contributions to phytopathogenic lifestyles. *PLoS ONE* 8: e60029. doi:10.1371/journal.pone.0060029.
 49. Garcia-Vallve S, Romeu A, Palau J (2000) Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol Biol Evol* 17: 352–361.
 50. Novo M, Bigey F, Beyne E, Galeote V, Gavory F, et al. (2009) Eukaryote-to-eukaryote gene transfer events revealed by the genome sequenced by the wine yeast *Saccharomyces cerevisiae* EC1118. *Proc Natl Acad Sci U S A* 106: 16333–16338. doi:10.1073/pnas.0904673106.
 51. Khaldi N, Collemare J, Lebrun M (2008) Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol* 9: R18.
 52. Slot JC, Hibbett DS (2007) Horizontal transfer of a nitrate assimilation gene cluster and ecological transitions in fungi: a phylogenetic study. *PLoS ONE* 2: e1097. doi:10.1371/journal.pone.0001097.
 53. 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. doi:10.1073/pnas.0914418107.
 54. Slot JC, Rokas A (2011) Horizontal transfer of a large and highly toxic secondary metabolite gene cluster between fungi. *Curr Biol* 21: 134–139. doi:10.1016/j.cub.2010.12.020.
 55. Campbell MA, Rokas A, Slot JC (2012) Horizontal transfer and death of a fungal secondary metabolite gene cluster. *Genome Biol Evol* 4: 289–293. doi:10.1093/gbe/evs011.
 56. Campbell MA, Staats M, van Kan JAL, Rokas A, Slot JC (2013) Repeated loss of an anciently horizontally transferred gene cluster in *Botrytis*. *Mycologia* 105: 1126–1134. doi:10.3852/12-390.
 57. Patron NJ, Waller RF, Cozijnsen AJ, Straney DC, Gardiner DM, et al. (2007) Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. *BMC Evol Biol* 7: 174. doi:10.1186/1471-2148-7-174.
 58. Khaldi N, Wolfe KH (2011) Evolutionary origins of the fumonisin secondary metabolite gene cluster in *Fusarium verticillioides* and *Aspergillus niger*. *Int J Evol Biol* 2011: 423821–423827. doi:10.4061/2011/423821.
 59. Durand D, Halldórsson BV, Vernot B (2006) A hybrid micro-macroevolutionary approach to gene tree reconstruction. *J Comput Biol* 13: 320–335. doi:10.1089/cmb.2006.13.320.
 60. Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, et al. (2012) Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics* 28: I409–I415. doi:10.1093/bioinformatics/bts386.
 61. Vernot B, Stolzer M, Goldman A, Durand D (2007) Reconciliation with non-binary species trees. *Comput Syst Bioinformatics Conf* 6: 441–452.
 62. Wolfe KH, Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387: 708–713. doi:10.1038/42711.
 63. Vining LC (1992) Secondary metabolism, inventive evolution and biochemical diversity—a review. *Gene* 115: 135–140.
 64. Trapp SC, Croteau RB (2001) Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics* 158: 811–832.
 65. Hopwood DA (1997) Genetic contributions to understanding polyketide synthases. *Chemical reviews* 97: 2465–2498. doi:10.1021/cr960034i.
 66. Kroken S, Glass N, Taylor J, Yoder O, Turgeon B (2003) Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc Natl Acad Sci U S A* 100: 15670–15675. doi:10.1073/pnas.2532165100.
 67. Bushley KE, Turgeon BG (2010) Phylogenomics reveals subfamilies of fungal nonribosomal peptide synthetases and their evolutionary relationships. *BMC Evol Biol* 10: 26. doi:10.1186/1471-2148-10-26.
 68. Condon BJ, Leng Y, Wu D, Bushley KE, Ohm RA, et al. (2013) Comparative genome structure, secondary metabolite, and effector coding capacity across *Cochliobolus* pathogens. *PLoS Genet* 9: e1003233. doi:10.1371/journal.pgen.1003233.
 69. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, et al. (2010) Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature* 464: 367–373. doi:10.1038/nature08850.
 70. Coleman JJ, Rounsley SD, Rodriguez-Carres M, Kuo A, Wasmann CC, et al. (2009) The genome of *Nectria haematococca*: contribution of supernumerary chromosomes to gene expansion. *PLoS Genet* 5: e1000618. doi:10.1371/journal.pgen.1000618.
 71. de Jonge R, van Esse HP, Maruthachalam K, Bolton MD, Santhanam P, et al. (2012) Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc Natl Acad Sci U S A* 109: 5110–5115. doi:10.1073/pnas.1119623109.
 72. Liang H, Plazonic KR, Chen J, Li W-H, Fernández A (2008) Protein underwrapping causes dosage sensitivity and decreases gene duplicability. *PLoS Genet* 4: e11. doi:10.1371/journal.pgen.0040011.
 73. Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P (2007) Genome-Wide Experimental Determination of Barriers to Horizontal Gene Transfer. *Science* 318: 1449–1452.
 74. Papp B, Pal C, Hurst LD (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* 424: 194–197. doi:10.1038/nature01771.
 75. Li L, Huang Y, Xia X, Sun Z (2006) Preferential duplication in the sparse part of yeast protein interaction network. *Mol Biol Evol* 23: 2467–2473. doi:10.1093/molbev/msl121.
 76. Prachumwat A, Li W-H (2006) Protein function, connectivity, and duplicability in yeast. *Mol Biol Evol* 23: 30–39. doi:10.1093/molbev/msi249.
 77. Cohen O, Gophna U, Pupko T (2011) The complexity hypothesis revisited: connectivity rather than function constitutes a barrier to horizontal gene transfer. *Mol Biol Evol* 28: 1481–1489. doi:10.1093/molbev/msq333.
 78. Jain R, Rivera MC, Lake JA (1999) Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* 96: 3801–3806.
 79. Hurst LD, Williams E, Pal C (2002) Natural selection promotes the conservation of linkage of co-expressed genes. *Trends Genet* 18: 604–606.
 80. Takos AM, Rook F (2012) Why biosynthetic genes for chemical defense compounds cluster. *Trends Plant Sci* 17: 383–388. doi:10.1016/j.tplants.2012.04.004.
 81. 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. doi:10.1073/pnas.1304461110.
 82. Hittinger CT, Gonçalves P, Sampaio JP, Dover J, Johnston M, et al. (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* 464: 54–58. doi:10.1038/nature08791.
 83. 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: e25290. doi:10.1371/journal.pone.0025290.
 84. Walton JD (2000) Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genet Biol* 30: 167–171. doi:10.1006/fgbi.2000.1224.
 85. Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143: 1843–1860.
 86. Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33: 511–518. doi:10.1093/nar/gki198.
 87. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973. doi:10.1093/bioinformatics/btp348.
 88. Stamatakis A (2006) RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. doi:10.1093/bioinformatics/btl446.
 89. Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) version 3.6. Available: <http://evolution.genetics.washington.edu/phylip.html>.
 90. Price MN, Dehal PS, Arkin AP (2010) Fasttree 2 - approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5: e9490. doi:10.1371/journal.pone.0009490.
 91. Chen K, Durand D, Farach-Colton M (2000) NOTUNG: A program for dating gene duplications and optimizing gene family trees. *J Comput Biol* 7: 429–447. doi:10.1089/106652700750050871.
 92. R Code Team (2014) R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available: <http://www.R-project.org/>.
 93. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc, Series B* 57: 289–300.
 94. Wickham H (2009) ggplot2: elegant graphics for data analysis. New York: Springer.
 95. Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P (2011) iPath2.0: interactive pathway explorer. *Nucleic Acids Res* 39: W412–W415. doi:10.1093/nar/gkr313.