OXFORD UNIVERSITY PRESS | Genome Biology and Evolution

A complex history of leucine biosynthesis genes in fungi: gene fusion, fission, loss and horizontal transfer

Journal:	*Genome Biology and Evolution*
Manuscript ID	GBE-230637
Manuscript Type:	Article
Date Submitted by the Author:	27-Jun-2023
Complete List of Authors:	Anderson, Noelle ; University of California Merced, School of Natural Sciences, Quantitative and Systems Biology Graduate Group Sun, Steven; University of California San Francisco, Tetrad Graduate Program Roy, Scott; San Francisco State University, Biology
Keywords:	Fungal metabolism, Gene fusions, Genic evolution, Fungal evolution, Horizontal gene transfer, Leucine biosynthesis

A complex history of leucine biosynthesis genes in fungi: gene fusion, fission, loss and horizontal transfer

Noelle Anderson[1,2], Steven Sun[3,4], Scott William Roy[1,3],
[1]Department of Molecular and Cell Biology, University of California-Merced, Merced, CA
[2]Quantitative and Systems Biology Graduate Group, University of California-Merced, Merced, CA
[3]Department of Biology, San Francisco State University, San Francisco, CA
[4]Tetrad Graduate Program, University of California, San Francisco, California 94143

Abstract

Fungi exhibit unique metabolic capabilities and are highly dependent on their metabolic adaptations to thrive in diverse environments, making them an excellent model for studying metabolic adaptation and evolution. Here, we trace the distribution and history across fungi of a gene fusion involving two essential enzymes that work in subsequent steps in fungal leucine biosynthesis, 3-isopropylmalate dehydrogenase (IPMDH) and 2-isopropylmalate isomerase (IPMI). Through phylogenetic reconstruction, we find evidence for a complex history involving ancestral IPMDH-IPMI fusion, secondary fission and loss of the fused genes, potential cases of secondary fusion, and multiple cases of horizontal gene transfer. While genes involved in the same metabolic pathways are often physically associated in ways thought to improve metabolic efficiency, suggesting adaptive significance for the fused gene, recurrent loss or fission of the fused gene complicates this narrative. The results presented here represent a remarkably intricate history for a pair of key enzymes, highlighting the complexities not captured by current dominant models of molecular evolution.

Background

Fungi, unlike most eukaryotes, rely entirely on osmotrophic nutrient uptake and take in diverse substrates. Because of this, they are highly dependent on their metabolic capabilities, including their ability to specialize their metabolism ((Watkinson 2016; Naranjo-Ortiz and Gabaldón 2019; Naranjo-Ortiz and Gabaldón 2020)**,** Naranjo-Ortiz and Gabaldón 2020). This fact, coupled with the availability of large genomic datasets, makes fungi an excellent model for studying metabolic adaptation and evolution.

Diffusion of products between enzymes can be a limiting step in metabolic pathways, a problem that can be mitigated in part by the physical association of enzymes that act in sequential steps of the pathway, often by means of increased substrate channeling efficiency, sometimes within enzyme complexes (so-called metabolons) (Sweetlove and Fernie 2018). The physical association of functionally related genes in gene clusters that has been frequently observed in fungal genomes and may be related to the increased importance of optimizing metabolism in fungi (Wisecaver et al. 2014; Wisecaver and Rokas 2015; Slot 2017; Nützmann et al. 2018; Rokas et al. 2018; Rokas et al. 2020).

Gene fusions are a more poorly studied means than metabolic gene clusters to improve metabolism by physically associating genes. Indeed, metabolically related gene fusions have demonstrated improved enzyme kinetics over their unfused counterparts and substrate

channeling, notably with yeast mitochondrial TCA cycle enzymes in a biochemical reaction (Lindbladh et al. 1994; Elcock and Andrew McCammon 1996), though this claim has also been challenged on the basis of how kinetics are evaluated (Pettersson et al. 2000). In a study including a fusion of yeast glycerol pathway enzymes transformed in *E. coli*, multi-fold improvements in enzyme kinetics and glycerol biosynthesis were observed in vivo and in purified reaction (Meynial Salles et al. 2007). A single peptide with linked functional domains used in adjacent steps in a metabolic pathway may thus confer an advantage relative to unfused gene copies. However, in other cases no evidence has been found for increased metabolic flux in observed fusions of adjacent genes in a metabolic pathway (Castellana et al. 2014), and in some cases fusions show reduced enzyme activity compared to free enzymes (Kourtz et al. 2005).

Another important process in metabolic evolution in fungi is lateral gene transfer (Richards 2011; Richards and Talbot 2013; Naranjo-Ortiz and Gabaldón 2020). While much more frequent among prokaryotes, this mode of evolution has now been observed across all three domains of life, often even involving HGT from one domain of life to another (Zhaxybayeva and Doolittle 2011; Gabaldón 2020; Cote-L'Heureux et al. 2022). Horizontally transferred genes can have many positive effects on the genome, bringing in novel functions, replacing ancestral copies with equivalent functions, and even replacing ancestral copies with slightly (or moreso) improved copies (Soucy et al. 2015). Genes that have been horizontally transferred in fungi are overrepresented in primary and secondary metabolic functions and gene clusters have been shown to be frequently horizontally transferred (Richards et al. 2011; Fitzpatrick 2012; Wisecaver et al. 2014; Wisecaver and Rokas 2015).Thus, fungi provide an abundance of opportunities for understanding the role of metabolically-related physically associated genes that may improve metabolic efficiency, and how HGT can be a path to propagating these metabolic improvements between divergent species.

Here, we report the complex evolution of fungal 3-isopropylmalate dehydrogenases (IPMDH, EC 1.1.1.85) and 2-isopropylmalate isomerases (IPMI, EC 4.2.1.33), two enzymes sequential in leucine biosynthesis and with homologous variants adjacent in the TCA cycle (Fig. 1). In leucine biosynthesis, IPMI catalyzes the isomerization of 2-isopropylmalate into 3-isopropylmalate, while IPMDH decarboxylates the 3-isopropylmalate into α-ketoisocaproate. This fusion was previously identified in fungi and phylogenetically analyzed by Leonard and Richards, but too little fungal data was available at that time to reveal the more complex story we illustrate here (Leonard and Richards 2012). Consistent with roles for gene fusion and HGT in improving existing metabolic pathways, we find evidence for fused genes that encode both enzymes as separate domains, and for HGT of these fused constructs. At the same time, we also find evidence for recurrent fission and loss of these fused copies, complicating the simple story of directional evolution for increased metabolic efficiency. These results demonstrate the complex evolutionary histories within eukaryotes of even the most essential and well-understood genes.

Figure 1. IPMDH-IPMI gene fusion domain structure. Length ranges show 10-90th percentiles across fused genes observed across fungi.

Methods

Stimulated by a report of a gene fusion involving an aconitase gene in fission yeast (Jung et al. 2015), we used the 'Clusters' tool on the JGI's Mycocosm website to manually search for unknown additional aconitase fusions (Grigoriev et al. 2011; Grigoriev et al. 2014). Multiple fusions of IPMDH-IPMI were found and in each case, the IPMDH domain (PF00180/IPR024084) lay upstream of the IPMI domain (PF00330/IPR001030) with an intervening linker region mode of 95 residues (Fig. 1). Additional IPMDH-IPMI fusions not found in JGI's Mycocosm Clusters were identified using NCBI's Conserved Domain Architecture Retrieval Tool (Geer et al. 2002). IPMDH and IPMI were then searched separately against all available fungal proteomes from the 1000 Fungal Genomes dataset in December 2018 (Grigoriev et al. 2014), and Genbank's non-redundant protein database using BlastP 2.7.1 (Altschul et al. 1990; Camacho et al. 2009) and highly-significant results compiled (e.g., e-value < 10^{-20}). To compile the unfused homologs, 50 unfused versions of IPMDH and IPMI were randomly (but broadly taxonomically) chosen from JGI's database and pblast searches were performed against all the same two databases. Additional sampling of taxa within groups underrepresented in our initial blast search was done to ensure adequate sampling. Full non- genera or bootstrap value-collapsed IPMDH and IPMI gene trees with taxa NCBI accessions or JGI protein IDs are available in supplemental materials.

Domain boundaries were identified HMMscan using HMMER version 3.1b2 (http://hmmer.org/) sequences were then split into the respective IPMDH and IPMI domains before aligning individually using MAFFT with alignment strategy FFT-NS-2 (Katoh et al. 2002). Model testing was performed in IQ-TREE version 1.6.10 (Nguyen et al. 2015) and for both alignments, the LG model with 6 rate categories was chosen. Gene trees were constructed using IQ-TREE and ultrafast bootstrapping approximation with 10,000 pseudoreplicates was performed (Minh et al. 2013).

Horizontal transfer events were identified by comparing the gene trees to the species tree. If the species with the identified fusions cluster while the rest of the phylogeny remains largely congruent with the species tree topology, the genes are likely to be derived from a common fusion event with multiple rounds of HGT, particularly when the taxonomic distribution of the fusion is punctate.

For searching the Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP) dataset (Keeling et al. 2014), the fused protein sequence JGI protein ID 441756 from the

Entophlyctis helioformis JEL805 v1.0 genome was used as a BlastP query against all translated transcriptomes of the MMETSP project.

Results

Kingdom-wide search for IPMDH-IPMI fusions in fungi

In our manual search of JGI's Mycocosm, we identified genes containing IPMDH and IPMI (two enzymes in the isocitrate/isopropylmalate dehydrogenase and aconitase families, respectively) fused together in various Pezizomycotina species. Intrigued, we searched for the fusion across all available fungal protein sequences in Genbank and JGI's Mycocosm database. To our surprise, we found that fusion genes were present in nearly all major groups. However, despite this broad distribution, the overall phylogenetic distribution is punctate, suggesting a complex history. Fused genes are absent from entire groups (e.g. Basidiomycota, Saccharomycotina, Monoblepharidomycota, Microsporidia and Rozellomycota), and are found in only some species within other groups (e.g., Pezizomycotina, Taphrinomycotina, Mucoromycotina, Blastocladiomycota, Chytridiomycota, Aphelidiomycota) (See Fig. 2).

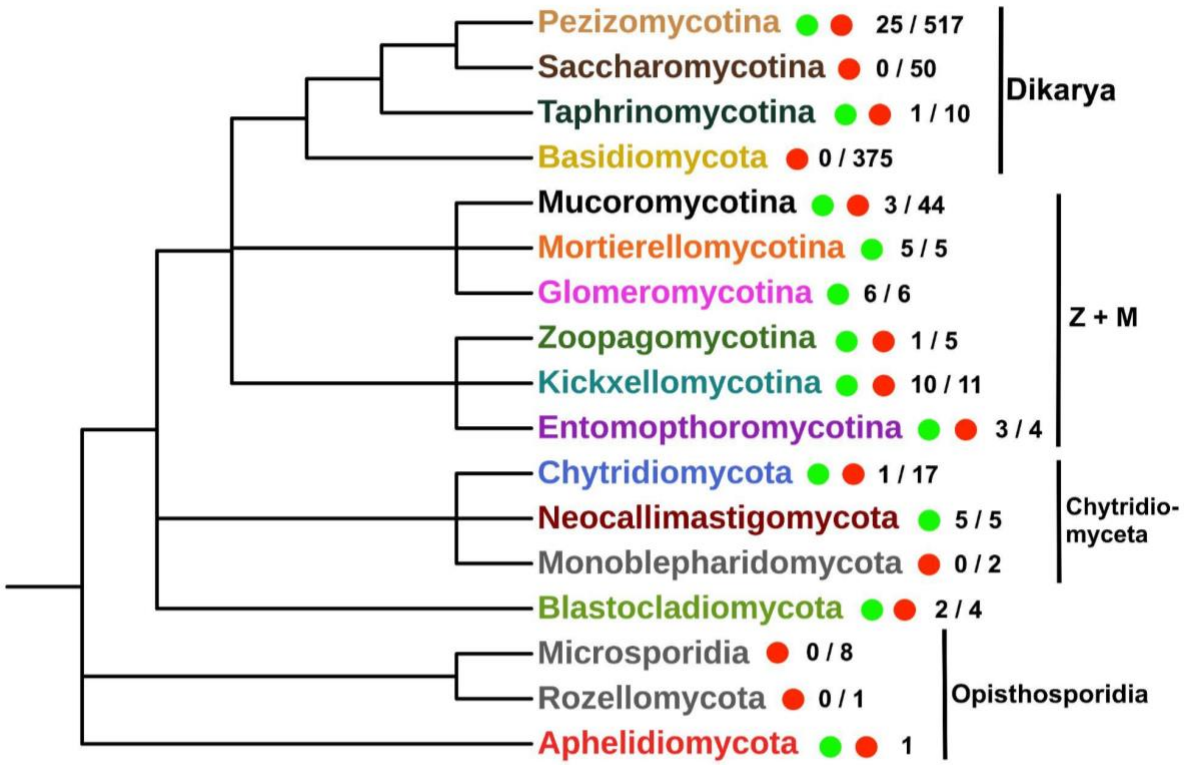

Figure 2. Cladogram indicating the phylogenetic distribution of the fusion gene across all major fungal groups, with topology based on (Li et al. 2021). Groups in which all assessed species either contain the fusion gene (green circle) or lack the fusion gene (red circle), or in which diversity within the group is observed (green and red circles together). Values by each taxonomic group represent the number of species with confirmed fusions found within that group (left), out of the number in the search database from JGI's Mycocosm at the time of

download, plus any fusions in species found only through NCBI (right). The single Aphelidiomycota fusion was located via NCBI.

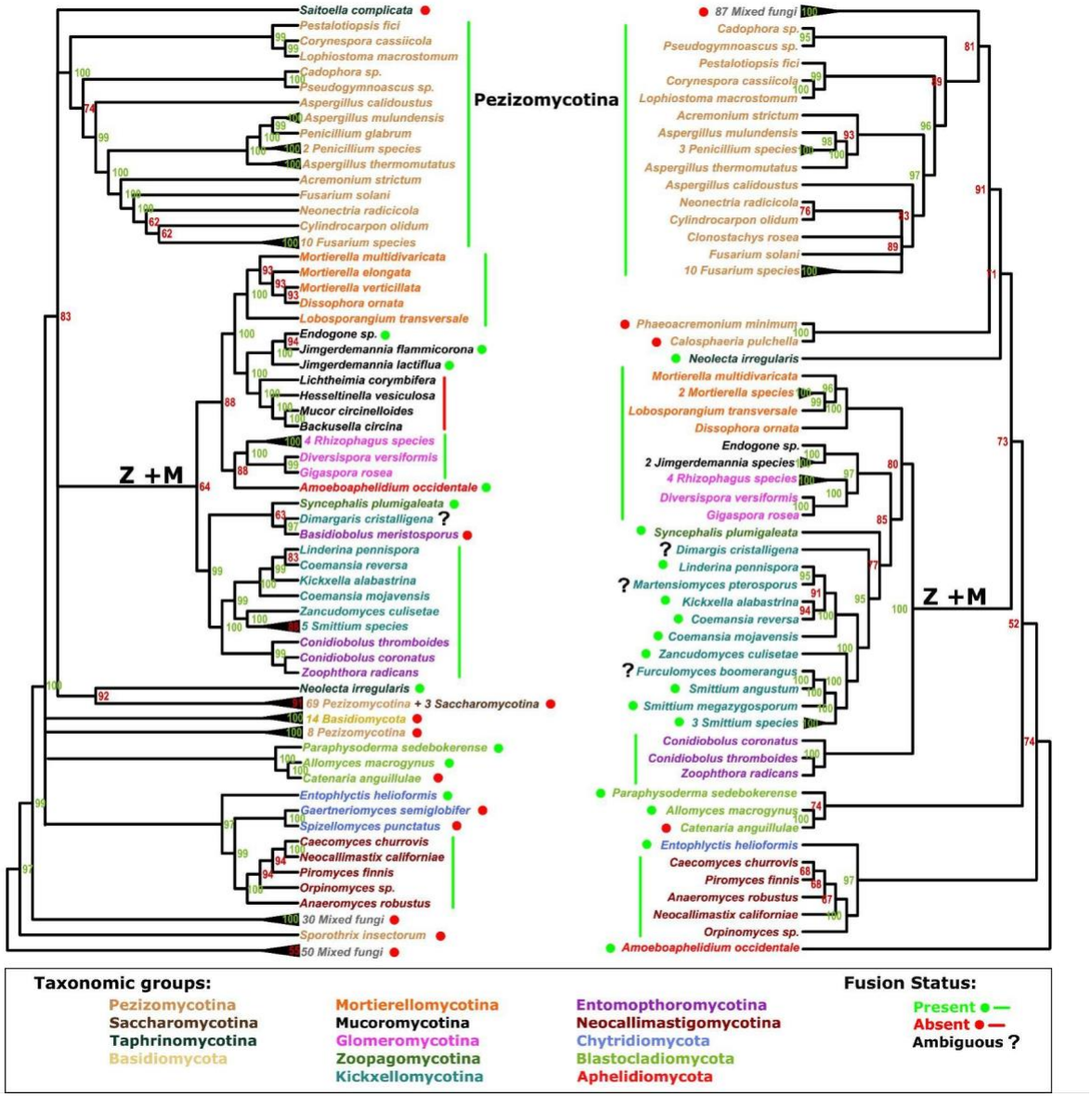

Figure 3. IPMDH (left) and IPMI (right) summary gene trees. Species with IPMDH-IPMI fusions are indicated with green circles and bars and those without with red circles and bars. Ambiguous cases (mostly likely missannotations, see text) are shown with question marks. Species are colored by subphyla, and congeners have been collapsed. Only nodes with > 50 ultrafast bootstrapping support (from 10,000 pseudoreplicates) are shown as resolved and BS values below 95 are shown in red (See Supplemental Figures S1 and S2 for full uncollapsed trees). The clade of fusions containing Zoopagomycota and Mucoromycota is marked as 'Z+M'.

Phylogenetic analysis of fused genes and unfused homologs

We used phylogenetic analysis to better understand the history of the IPMDH-IPMI fusion genes and their irregular distribution across fungi. We first separated fused genes into their two constituent domains, and then conducted extensive recursive BlastP searches of the NR database on NCBI and the JGI fungal genomes database to identify as many homologs as possible, both fused and unfused, from across available fungal genomes. We also performed searches across eukaryotes and prokaryotes more generally, and found a single fused gene outside of fungi in the Aphelidiomycota phylum (in the species *Amoeboaphelidium occidentale,* KAI3662018.1), which represents Opisthosporidia, a sister group to fungi containing Aphelidiomycota, Rozellomycota, and Microsporidia.

We separately performed phylogenetic reconstruction of gene trees, one each for the IPMDH- and IPMI domains, along with homologs from across fungal species and the aphelid *Amoeboaphelidium occidentale* (Fig. 2). Through much of the trees, there is a high degree of similarity between the two trees (described in detail below), as expected for multi-domain genes with a single shared history. However, other parts of the obtained trees revealed complex histories. For both IPMDH and IPMI trees, most fused domains fell within two major clades, both of which consisted exclusively or almost exclusively of fused domains. We now discuss these two major clades in detail.

A diverse clade of fused genes suggests the presence of a fused IPMDH-IPMI gene in early fungal ancestors

For both of the IPMDH and IPMI gene trees, the larger of the two clades contained species from two major fungal phyla (Zoopagomycota and Mucoromycota) with representatives from all subphyla within (Mortierellomycotina, Mucoromycotina, Glomeromycotina, Zoopagomycotina, Kickxellomycotina, Entomopthoromycotina) containing the fused gene. Thus we refer to this clade as Z+M. Note that while Zoopagomycota and Mucoromycota were formerly classified as a clade under the group Zygomycetes based on morphology and reproduction, later evidence supported Zygomycete paraphyly where either group forms a clade with Dikarya, but still more recent event supports the initial hypothesis Z+M monophyly – thus the true relationships are best regarded as unresolved (Li et al. 2021).

For both IPMDH and IPMI, within the Z+M clade, we found that the gene tree topology of the clades (see simplified tree summaries in Fig. 3) has clear similarities to known organismal relationships between the represented fungal groups (Fig. 2). This is particularly the case for IPMDH, which recovers the Mucoromycota clade (Glomeromycotina, Mortierellomycotina, and Mucoromycotina) with moderately high support (88bs), as well as a well-supported clade containing species within Zoopagomycota (Zoopagomycotina, Kickxellomycotina, and Entomophthoromycotina) species, as well as the speculated sister relationship between these two large clades (64bs) (Fig. 3, left). In the IPMI gene tree (Fig. 3, right), we reconstruct a highly supported (100bs) clade containing Mucoromycota and Zoopagomycota and while species from each subphyla group with high support and both Mucoromycotina and Glomeromycotina (97bs) and Kickxellomycotina and Zoopagomycotina (77bs) respectively form sister clades, relationships between other subphyla are less stable, with both Entomopthoromycotina and

Mortierellomycotina species constituting a clade, but grouping outside of their expect subphyla-level relationships. Given the species-like IPMDH grouping and the modest IPMI bootstrap support (77-85bs), we speculate that these portions of the tree also represent vertical inheritance but with phylogenetic reconstruction errors in the IPMI tree reconstruction.

The only clear exceptions to the Z+M broad species-like grouping are observed in the IPMDH gene tree. First, the aphelid *Amoeboaphelidium occidentale* falls within the Z+M clade as sister to a species-tree like group of Glomeromycotina. Second, within the IPMDH Z+M clade, in contrast to the clearly species-like relationships between most species, there exists an additional three-gene clade including genes from three different Zoopagomycota groups.

This pattern, in which the gene tree in these clades largely follows the species tree, is as expected if the IPMDH-IPMI fusion was present in deep fungal ancestors and has been retained by vertical inheritance by a wide diversity of extant fungi. On the other hand, this pattern is difficult to explain otherwise. The only major possible exception to this species-like pattern is the lack of any representatives of the Dikarya (comprising Ascomycota + Basidiomycota). This could represent gene loss early in Dikarya or, depending on phylogeny (see above), gene gain after the Z+M lineage diverged from Dikarya (Fig. 5).

Is it possible that this gene fusion evolved even deeper within the fungal tree? Notably, both gene trees include fused genes from the other two deeper branching major fungal groups [Blastocladiomycota and Chytridiomyceta (Neocallimastigomycota, Chytridiomycota, and Monoblepharidomycota), as well as the sister group to fungi (Aphelidiomycota, represented by the aphelid *Amoeboaphelidium occidentale*)], including most available Neocallimastigomycota and Blastocladiomycota species (Fig. 3). However, the case is far from simple. First, no fusions besides the single aphelid were found in other Opisthosporidians. Second, only one Chytridiomycota species contains the fusion, and no fusions were found in the closely related Monoblepharidomycota clade. Third, the Blastocladiomycota and Chytridiomyceta fusions do not group into a larger fusion clade along with the Z+M clade, as would be expected by this fungal ancestor hypothesis. Indeed, the Z+M clade groups closer to paralogs whose own broad representation and semi-species-like tree structure suggests they are ancient paralogs (and thus should not fall within another ancestral gene clade). In total, then, there is no clear preponderance of evidence for a fused gene in the fungal ancestor.

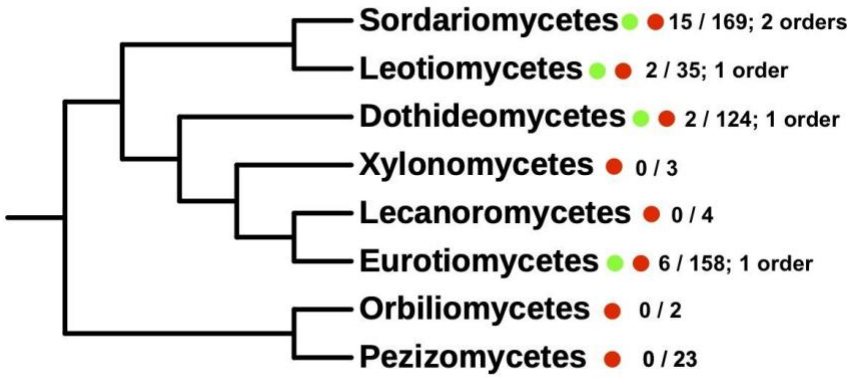

Figure 4. Presence and absence of fusions in Pezizomycotina classes. Species tree topology based on (Li et al. 2021). Red circles indicate that species without the IPMDH-IPMI fusions were found in this class. Green circles represent fusion presence. For each class, numerator and denominator give the number of species containing fusions and the number of species in that class present in the JGI Mycocosm database at the time of searching (plus any fusions found on NCBI), respectively. For classes containing fusions, the number of orders represented with fusions is also given.

Evidence for recurrent lateral gene fusion in Pezizomycotina

A strikingly different pattern is seen in the second major clade of fused genes present in both trees, containing genes from within the large Ascomycota group Pezizomycotina. For this clade, the history implied is quite different at the gene and species level. At the level of the gene (i.e., gene fusion and fission), the clade's history may be straightforward: within these clades the two domains mostly show highly similar topologies for the two genes, suggesting a shared evolutionary history for an ancestrally fused gene.

On the other hand, at the species level the tree is far less simple. First, the fused genes in this clade represent only a small fraction of studied Pezizomycotina species [21 out of 513 Pezizomycotina species searched from the downloaded JGI database (note that some fusions were found on NCBI and have been included in both numerator and denominator counts in Figure 4, e.g. 4 Pezizomycotina species were found on JGI, thus is counted as 25/517)]. However, the species that contain the fused gene are largely distantly related, being found in 4 different orders and 5 different classes (Fig. 3, Fig. 4). Moreover, the gene trees do not reflect species tree relationships (Supplemental figure S3). Both of these patterns are not at all as expected by vertical inheritance (which would require massive parallel gene loss and widespread failure of phylogenetic reconstruction specific to this clade). Instead, both patterns are as expected by multiple HGT events among Pezizomycotina species (and, notably, only among Pezizomycotina species, as non-Pezizomycotina species are absent from this clade).

Taking a closer look at the Pezizomycotina species tree topology, we can infer at least 11-15 lateral transfer events (Fig. 4, Supplemental figure S3). The range given is due to the possibility of either two transfers within Sordariomycetes (into the Hypocreales ancestor and the distantly

related Xylariales species *Pestalotiopsis fici*, explainable with 7 losses based on the JGI Mycocosm species tree) or five transfers (transfers into each Sordariomycetes fusion species on the tips, except in Fusarium where most searched species have the fusion and thus likely was an ancestral gain explainable with 3 losses within Fusarium, discussed below), or some intermediate. Notably the Sordariomycetes species do not group in species relationships outside of Fusarium (Fig. 3). In the other Pezizomycotina groups, while fusions occurred in single classes, the JGI Mycocosm species trees were also used to examine the positions of species with fusions to look for patterns of lateral transfer or gene loss. In all cases, species are sufficiently distantly related to parsimoniously conclude lateral transfer over gene loss.

Other possible HGT events are less well-supported

Other potential lateral gene transfer events are also observed in the larger trees, with clear and sometimes striking differences between the gene tree and species tree. These include the aphelid *Amoeboaphelidium occidentale* grouping with Glomeromycotina in the IPMDH tree and the presence of a fused gene in the single Taphrinomycotina species *Neolecta irregularis* (OLL24270.1) grouping with unrelated species (Fig. 3). However, these cases do not include the second line of evidence found in Pezizomycotina, namely spotty phylogenetic distribution of the fusion gene (their phylogenetic distribution being impossible to assess given relatively paltry genomic sequence of these lineages). Thus while these cases suggest HGT, given the real possibility of phylogenetic errors, we believe that misplacement alone is insufficient evidence to conclude an instance of lateral gene transfer. Possibly ongoing sequencing of these groups will allow for distinguishing the history of these genes in the future.

Evidence for gene fission and domain loss

Among the complex changes we see across the tree, we observe evidence for multiple instances of gene fission and domain loss, particularly loss of the IPMI domain. First, we observe a clade of four species within Mucoromycotina that lack a fused gene, but which contain unfused IPMDH but not IPMI genes (Fig. 3). The unfused IPMDH genes group in species-tree positions with the fused genes found within other Mucoromycotina species, exactly as expected if the ancestrally fused gene lost the IPMI gene in the ancestor of these four species. Indeed, the species with the IPMDH-IPMI fusions form a single clade (Endogonomycetes), the earliest branching Mucoromycotina class, suggesting a single event occurring after this early divergence involving fission of the ancestral fused gene and IPMI domain loss. Elsewhere in the three, the Entomopthoromycotina species *Basidiobolus meristoporus* also has an unfused IPMDH annotated without an IPMI domain (or any evidence for IPMI nearby in the genome). This unfused gene groups with related Zoopagomycota species in the IPMDH tree, consistent with secondary domain loss.

Four other potential cases of fission could be due to the possibility of annotation errors. Both the Kickxellomycotina species *Dimargaris cristalligena* (IPMDH: RKP40269.1, IPMI: RKP40268.1) and the Blastocladiomycota species *Catenaria anguillulae* (IPMDH: ORZ31678.1, IPMI: ORZ31677.1) IPMDH and IPMI copies annotated as genomically adjacent but separate genes.

While these cases could possibly indicate gene fission, it is hard to explain simple missannotation of a single fused gene as separate genes. Both species fall within the Z+M fusion clade and while *C. anguillulae* falls in its phyla-level species tree relationship in both gene trees and *D. cristalligena* in the IPMI tree, in the IPMDH gene tree, *D. cristalligena* groups in the 3-gene Zoopagomycota group shared with *B. meristoporus* discussed above (Fig. 3). The Kickxellomycotina species *Martensiomyces pterosporus* (strain CBS 209.56 v1.0, JGI protein ID 265168) and *Furculomyces boomerangus* (PVU91561.1) are present in the IPMI but not IPMDH Z+M fusion clade, however, manual inspection revealed that these exceptions are likely due to a truncated annotated gene sequence that we could not confirm as a true fission.

We also observe potential evidence for complete gene loss (e.g. as previously discussed in the dikaryon ancestor, and in the Chytrid phylum Monoblepharidomycota, depending on the species tree phylogeny and ancestral fusion timing). Another potential complete loss includes the Blastocladiomycota species *Blastocladiella britannica*, sister to *Catenaria anguillulae* which possessed unfused gene copies while grouping with the other Blastocladiomycota fusions in the IPMDH and IPMI gene trees, while *B. britannica* falls deep outside of the fusion clades. A similar pattern is seen for several Zoopagomycotina species, where one fusion groups with others in its phylum within the Z+M clade, while four searched species showed unfused genes and appear elsewhere or absent from the tree, indicating several losses. In Pezizomycotina, *Fusarium pseudograminearum, Fusarium verticillioides, and Fusarium graminearum* are our only sampled Fusarium species not present in our gene trees (See Fig. S3), thus if the IPMDH-IPMI fusion was transferred into the Fusarium ancestor as the abundance of Fusarium species with the fusion and their species-tree like grouping suggests, these may also represent complete losses. However, given phylogenetic uncertainty and potential genome incompleteness it is more challenging to be confident about these events.

Homoplasies in complex gene evolution

The history documented here includes multiple cases of secondary domain loss and complete gene loss, as well as recurrent lateral gene transfer, underscoring the remarkably complex history of these core enzymatic functions. How many times has the gene fusion occurred? If the gene trees are to be believed, the parsimonious explanation could be to infer multiple cases of fusion leading separately to the Z+M, Blastocladiomycota and Chytridiomyceta clades. However, given the possibility of phylogenetic errors, we again believe caution is in order. The initial origins of the Pezizomycotina gene fusion is particularly curious. This clade groups deeply within the tree, suggesting an ancient origin for this clade, yet the fusion is found only within small recent groups (including some single species) (Fig. 3). This anomaly could be resolved if functional changes, changes in the cellular environment of Pezizomycotina, or other factors have led to an increase in evolutionary rate of this clade, in which case long branch attraction could have led to misplacement of this clade. If so, this clade could have emerged from within the Z+M (or another) clade; alternatively, this clade could represent an independent fusion. Another interesting case involves the Chytridiomyceta, in which subclades of fused and unfused genes group as sister, while Monoblepharidomycota is absent entirely (Fig. 3, Fig. 2). This pattern could be explained either by a novel fusion or by secondary fission and domain loss

(depending on the ancestral state, which is difficult to infer given the weak bootstrap support for the deepest branches of the tree). In general, then, we do not believe that the current data provide clear evidence as to the number of gene fusion events responsible for the observed diversity.

Reconstruction of the evolutionary history of IPMI and IPMDH genes in fungi

In total, then, these results suggest a complex history, as depicted in Figure 5. This history involves dozens of unexpected events, including at least: (i) one or multiple early IPMDH-IPMI fusions; (ii) recurrent loss of the ancestral fused gene, either by complete loss or by domain loss; (iii) many events of HGT.

An independent IPMDH-IPMI fusion event in green algae

The recurrent acquisition of IPMDH-IPMI gene fusions in fungi raises the question as to whether such fusions are common across the tree of life. A search for additional annotated cases of such fusions through BlastP searches of IPMI and IPMDH domains for all of GenBank yielded no results outside of fungi beyond the single aforementioned aphelid, suggesting that IPMDH-IPMI fusions are much more common in fungi than other groups. However, an additional search of the predicted proteomes of 670 diverse eukaryotes represented within the MMETSP database (Keeling et al. 2014) revealed a single additional case in green algae. The species *Chloroparvula japonica* (MMETSP1310/RC2339) shows a very similar fusion to those observed in fungi, with the IPMDH domain upstream of the IPMI domain with a much shorter linker region between IPMDH and IPMI of 4 amino acids (Supplemental Materials). Interestingly, in contrast to the fusions observed in fungi, in this instance both domains involved in the fusion show clear similarity to unfused copies in related species (data not shown), indicating a simpler fusion occurring within these species.

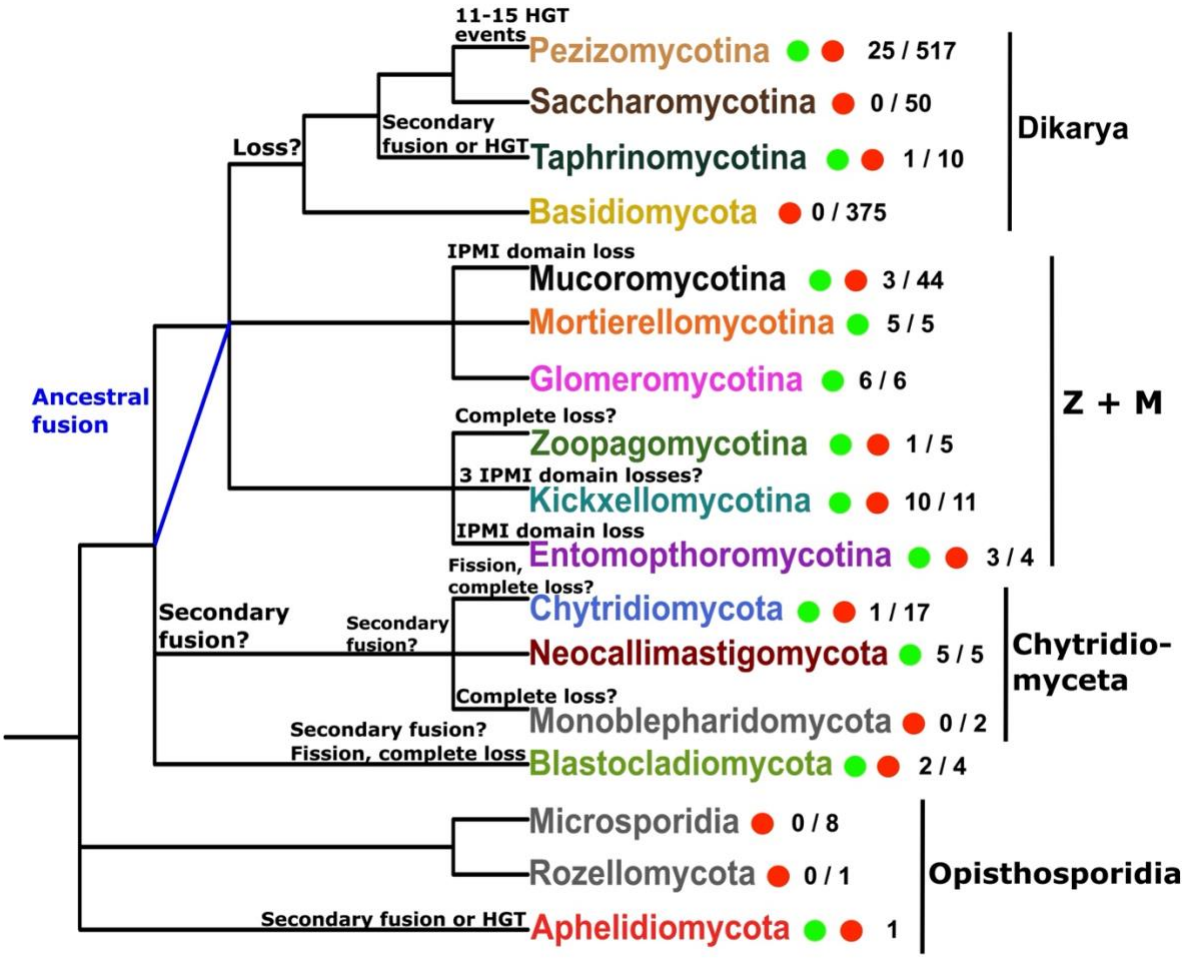

Figure 5. Summary of reconstructed evolutionary events involving the IPMDH-IPMI fusion across fungi and outgroups. HGT indicates horizontal gene transfer. “IPMI domain loss” indicates presence of a IPMDH gene related to fusions but lacking a fused IPMI, indicating gene fission and loss of the IPMI domain. Question marks indicate events with incomplete evidence (e.g., possible gene misannotation), or where different potential histories are plausible. In particular, the pattern observed in the Chytridiomyceta clade could by secondary fusion events in Chytridiomycota and Neocallimastigomycota, or alternatively by a retention of the ancestral fused gene or a single secondary fusion event, in either case followed loss in Monoblepharidomycota. The blue line indicates ambiguity as to the timing of the ancestral fusion event. As in Figure 2, green and red circles indicate presence and absence of gene fusion within each group, and numbers for fusions out of the searched JGI database are included as in Figure 3.

Discussion

Here we report a complex history of the leucine biosynthesis enzymes IPMDH and IPMI across fungi, including recurrent gene fusion, gene fission, and lateral gene transfer of fused genes (Fig. 5). In addition, the loss from the genome of ancestral enzyme-coding genes (since one of the domains/genes is generally lost in the cases of fission) suggests functional replacement of ancestral genes by paralogs or by lateral transfers, particularly given the core functions encoded

by the studied genes. This history represents, to our knowledge, among the most complex history of a single pair of genes to be reported, the most comparable being a similar story of gene fission, fusion and loss in ATP citrate lyase enzymes across eukaryotes (Gawryluk et al. 2015).

Several observations from literature on enzyme evolution, gene fusion, fission and lateral gene transfer collectively render aspects of the history reported here at once more and less surprising than it may initially seem. First, while lateral gene transfer was initially thought to be rare in eukaryotes, many studies have now shown that lateral gene transfers are not uncommon events in some eukaryotic lineages (Van Etten and Bhattacharya 2020; Gabaldón 2020). In particular, the phylogenetic concentration of HGT events observed in this study within Pezizomycotina is consistent with previous results indicating elevated rates in this group (Marcet-Houben and Gabaldón 2010). It has been demonstrated that gene fusions are transferred laterally more frequently than they occur independently or than they are inherited vertically with subsequent fission (Yanai et al. 2002). Moreover, metabolic enzymes have been shown to be frequently laterally transferred, consistent with the current study (Andersson 2005; Schönknecht et al. 2014; Van Etten and Bhattacharya 2020). On the other hand, work on HGT of metabolic enzymes has tended to concentrate on nonessential enzymes, in particular in cases of concerted transfer of entire metabolic pathways responsible for synthesizing or degrading particular compounds that often produce secondary metabolites (Wisecaver and Rokas 2015; Rokas et al. 2020). By contrast, here we describe recurrent transfer of genes involved in the essential process of leucine biosynthesis, in clear contrast to the paradigm of HGT of auxiliary enzymes. The essentiality of the reactions catalyzed also makes the finding of recurrent loss of ancestral enzymes more surprising. Presumably these species still synthesize leucine, however what genes perform the function of the lost enzyme remains obscure. However, some previous work suggests that the specific reactions studied here may be more prone to functional replacement, as in other fungi it has been found that homologous enzymes can take up this function, suggesting a route to gene loss in these species (Larson and Idnurm 2010; Aguirre-López et al. 2020).

The pattern of recurrent fusion and fission is also challenging to interpret. Fusion between adjacent enzymes in a metabolic pathway immediately suggests the possibility of optimization of the pathway due to more efficient intermediate transfers, perhaps by substrate channeling, or by coordination of gene expression (Tsoka and Ouzounis 2000; Henry et al. 2016; Hagel and Facchini 2017). Indeed, a specific fusion of IPMDH and IPMI is supported by previous work. Experimental evidence suggests that physical complexing between IPMDH and IPMI in vivo may act to regulate substrate channeling (Chen et al. 2021). This association may also be analogous to the metabolon complex of the TCA cycle (see Bulutoglu et al. 2016) where aconitase associates with citrate synthase and malate dehydrogenase and exhibits substrate channeling. Thus, fusion of the two genes encoding for the partner enzymes can lead to a structure with improved functions.

However, here there are several reasons for caution. The first is that we observe both fusion and “un-fusion”, whether by domain loss or by entire gene loss and presumably replacement by

non-fused alternatives: such recurrent reversal through evolution is not expected for a generally adaptive change. Secondly, substrate channeling may be a more major force in prokaryotes over eukaryotes, though may also simply be better studied in prokaryotes. In the absence of direct biochemical evidence, this complexity generally urges caution in interpreting the evolutionary history, caution which certainly applies to interpretations of the gene fusions.

Robustness of core findings to methodological limitations

The results presented here rely on genome annotation and phylogenetic tree building, both of which are notoriously challenging and prone to error, thus caution is in order. However, while the specifics of the reconstruction of the evolutionary history of these genes could be altered by methodological issues, we believe that our most important broad conclusions are robust to these concerns. The conclusion of an IPMDH-IPMI fusion in early fungi depends only on the concordance between the gene tree and species tree, which is not expected based on errors in gene annotation (since failure to annotate fused genes in some species would tend to cause fusions to appear more recent, not more ancient) or in phylogenetic reconstruction (which should tend to destroy rather than create concordance of gene trees with species trees). While gene annotation errors could indeed lead to a conclusion of loss of the gene fusion, the fact that the gene fusion is not annotated in large groups of species is not expected by stochastic gene annotation errors. Indeed, gene annotation errors would be most likely to manifest as absence of the fusion from a single species, which we see in 3 species discussed above with additional evidence of potential missanotation, and one singular loss in *B. meristoporus* that appears to be a genuine fission and loss. Recurrent lateral transfer of a fused gene between Pezizomycotina species is supported by the highly punctate phylogenetic pattern of presence of the fusion; by the non-species relationship reconstructed for the IPMI domain; and by a highly concordant non-species relationship reconstructed for the IPMDH domain (Fig. 3, Fig. 4). That secondary fusions have occurred is independently supported by the grouping of both fused domains outside of pan-fungal fusion clade for the same group of species (e.g., Pezizomycotina).

Concluding remarks

These results, along with cumulative evidence from other studies, highlights the limitations of the paradigm of genes encoding core functions in eukaryotes evolving conservatively and by strictly vertical inheritance. Genes with fused and non-fused versions may be particularly useful in illuminating complex histories, since it can help to distinguish true HGTs from errors of phylogenetic reconstruction. It will be interesting to learn whether these results truly represent an extreme example or rather a typical example of an overlooked subset of eukaryotic genes with complex histories.

Acknowledgements

NA and SWR were supported on National Scientific Foundation awards 1616878 and 1751372.

Data Availability Statement

The data used in this article are publicly available through Genbank and JGI's Mycocosm database and all accessions and JGI protein IDs are in Supplemental Materials.

References:

Aguirre-López B, Escalera-Fanjul X, Hersch-González J, Rojas-Ortega E, El-Hafidi M, Lezama M, González J, Bianchi MM, López G, Márquez D, et al. 2020. In Kluyveromyces lactis a Pair of Paralogous Isozymes Catalyze the First Committed Step of Leucine Biosynthesis in Either the Mitochondria or the Cytosol. *Front Microbiol* 11:1843.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* [Internet] 215:403–410. Available from: https://pubmed.ncbi.nlm.nih.gov/2231712/

Andersson JO. 2005. Lateral gene transfer in eukaryotes. *Cellular and Molecular Life Sciences* [Internet] 62:1182–1197. Available from: https://link.springer.com/article/10.1007/s00018-005-4539-z

Bulutoglu B, Garcia KE, Wu F, Minteer SD, Banta S. 2016. Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes. *ACS Chem Biol* [Internet] 11:2847–2853. Available from: https://pubs.acs.org/doi/full/10.1021/acschembio.6b00523

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* [Internet] 10:1–9. Available from: http://dx.doi.org/10.1186/1471-2105-10-421

Castellana M, Wilson MZ, Xu Y, Joshi P, Cristea IM, Rabinowitz JD, Gitai Z, Wingreen NS. 2014. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature Biotechnology 2014 32:10* [Internet] 32:1011–1018. Available from: https://www.nature.com/articles/nbt.3018

Chen LQ, Chhajed S, Zhang T, Collins JM, Pang Q, Song W, He Y, Chen S. 2021. Protein complex formation in methionine chain-elongation and leucine biosynthesis. *Scientific Reports 2021 11:1* [Internet] 11:1–8. Available from: https://www.nature.com/articles/s41598-021-82790-4

Cote-L'Heureux A, Maurer-Alcalá XX, Katz LA. 2022. Old genes in new places: A taxon-rich analysis of interdomain lateral gene transfer events. *PLoS Genet* [Internet] 18:e1010239. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010239

Elcock AH, Andrew McCammon J. 1996. Evidence for Electrostatic Channeling in a Fusion Protein of Malate Dehydrogenase and Citrate Synthase†. *Biochemistry* [Internet] 35:12652–12658. Available from: https://pubs.acs.org/doi/full/10.1021/bi9614747

Van Etten J, Bhattacharya D. 2020. Horizontal Gene Transfer in Eukaryotes: Not if, but How Much? *Trends in Genetics* 36:915–925.

Fitzpatrick DA. 2012. Horizontal gene transfer in fungi. *FEMS Microbiol Lett* [Internet] 329:1–8. Available from: http://femsle.oxfordjournals.org/content/329/1/1.abstract

Gabaldón T. 2020. Patterns and impacts of nonvertical evolution in eukaryotes: a paradigm shift. *Ann N Y Acad Sci* [Internet] 1476:78–92. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/nyas.14471

Gawryluk RMR, Eme L, Roger AJ. 2015. Gene fusion, fission, lateral transfer, and loss: Not-so-rare events in the evolution of eukaryotic ATP citrate lyase. *Mol Phylogenet Evol* [Internet] 91:12–16. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1055790315001475

Geer LY, Domrachev M, Lipman DJ, Bryant SH. 2002. CDART: Protein Homology by Domain Architecture. *Genome Res* [Internet] 12:1619–1623. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12368255

Grigoriev I V., Cullen D, Goodwin SB, Hibbett D, Jeffries TW, Kubicek CP, Kuske C, Magnuson JK, Martin F, Spatafora JW, et al. 2011. Fueling the future with fungal genomics. *Mycology* [Internet] 2:192–209. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=tmyc20

Grigoriev I V., Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, Riley R, Salamov A, Zhao X, Korzeniewski F, et al. 2014. MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* [Internet] 42:D699–D704. Available from: https://academic.oup.com/nar/article/42/D1/D699/1058210

Hagel JM, Facchini PJ. 2017. Tying the knot: occurrence and possible significance of gene fusions in plant metabolism and beyond. *J Exp Bot* [Internet] 68:4029–4043. Available from: https://academic.oup.com/jxb/article/68/15/4029/3832250

Henry CS, Lerma-Ortiz C, Gerdes SY, Mullen JD, Colasanti R, Zhukov A, Frelin O, Thiaville JJ, Zallot R, Niehaus TD, et al. 2016. Systematic identification and analysis of frequent gene fusion events in metabolic pathways. *BMC Genomics* [Internet] 17:1–17. Available from: https://link.springer.com/articles/10.1186/s12864-016-2782-3

Jung S-J, Seo Y, Lee K-C, Lee D, Roe J-H. 2015. Essential function of Aco2, a fusion protein of aconitase and mitochondrial ribosomal protein bL21, in mitochondrial translation in fission yeast. *FEBS Lett* [Internet] 589:822–828. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0014579315001076

Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* [Internet] 30:3059. Available from: /pmc/articles/PMC135756/

Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): Illuminating the Functional Diversity of Eukaryotic Life in the Oceans through Transcriptome Sequencing. *PLoS Biol* [Internet] 12:e1001889. Available from: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001889

Kourtz L, Dillon K, Daughtry S, Madison LL, Peoples O, Snell KD. 2005. A novel thiolase-reductase gene fusion promotes the production of polyhydroxybutyrate in Arabidopsis. *Plant Biotechnol J* 3:435–447.

Larson EM, Idnurm A. 2010. Two Origins for the Gene Encoding α-Isopropylmalate Synthase in Fungi. *PLoS One* [Internet] 5:e11605. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011605

Leonard G, Richards TA. 2012. Genome-scale comparative analysis of gene fusions, gene fissions, and the fungal tree of life. *Proc Natl Acad Sci U S A* 109:21402–21407.

583 Li Y, Steenwyk JL, Chang Y, Wang Y, James TY, Stajich JE, Spatafora JW, Groenewald M,
584 Dunn CW, Hittinger CT, et al. 2021. A genome-scale phylogeny of the kingdom Fungi.
585 *Current Biology* 31:1653-1665.e5.
586 Lindbladh C, Rault M, Hagglund C, Small WC, Mosbach K, Buelow L, Evans C, Srere PA. 1994.
587 Preparation and kinetic characterization of a fusion protein of yeast mitochondrial citrate
588 synthase and malate dehydrogenase. *Biochemistry* [Internet] 33:11692–11698. Available
589 from: http://pubs.acs.org/doi/abs/10.1021/bi00205a004
590 Marcet-Houben M, Gabaldón T. 2010. Acquisition of prokaryotic genes by fungal genomes.
591 *Trends in Genetics* 26:5–8.
592 Meynial Salles I, Forchhammer N, Croux C, Girbal L, Soucaille P. 2007. Evolution of a
593 Saccharomyces cerevisiae metabolic pathway in Escherichia coli. *Metab Eng* 9:152–159.
594 Minh BQ, Nguyen MAT, Von Haeseler A. 2013. Ultrafast Approximation for Phylogenetic
595 Bootstrap. *Mol Biol Evol* [Internet] 30:1188–1195. Available from:
596 https://academic.oup.com/mbe/article/30/5/1188/997508
597 Naranjo-Ortiz MA, Gabaldón T. 2019. Fungal evolution: major ecological adaptations and
598 evolutionary transitions. *Biological Reviews* 94:1443–1476.
599 Naranjo-Ortiz MA, Gabaldón T. 2020. Fungal evolution: cellular, genomic and metabolic
600 complexity. *Biological Reviews* [Internet] 95:1198–1232. Available from:
601 https://onlinelibrary.wiley.com/doi/full/10.1111/brv.12605
602 Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective
603 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*
604 [Internet] 32:268. Available from: /pmc/articles/PMC4271533/
605 Nützmann H-W, Scazzocchio C, Osbourn A. 2018. Metabolic Gene Clusters in Eukaryotes.
606 *Annu Rev Genet* 52:159–183.
607 Pettersson H, Olsson P, Bülow L, Pettersson G. 2000. Kinetics of the coupled reaction
608 catalysed by a fusion protein of yeast mitochondrial malate dehydrogenase and citrate
609 synthase. *Eur J Biochem* [Internet] 267:5041–5046. Available from:
610 https://onlinelibrary.wiley.com/doi/full/10.1046/j.1432-1327.2000.01558.x
611 Richards T a., Leonard G, Soanes DM, Talbot NJ. 2011. Gene transfer into the fungi. *Fungal*
612 *Biol Rev* [Internet] 25:98–110. Available from: http://dx.doi.org/10.1016/j.fbr.2011.04.003
613 Richards TA. 2011. Genome Evolution: Horizontal Movements in the Fungi. *Current Biology*
614 21:R166–R168.
615 Richards TA, Talbot NJ. 2013. Horizontal gene transfer in osmotrophs: Playing with public
616 goods. *Nat Rev Microbiol* 11:720–727.
617 Rokas A, Mead ME, Steenwyk JL, Raja HA, Oberlies NH. 2020. Biosynthetic gene clusters and
618 the evolution of fungal chemodiversity. *Nat Prod Rep* 37:868–878.
619 Rokas A, Wisecaver JH, Lind AL. 2018. The birth, evolution and death of metabolic gene
620 clusters in fungi. *Nature Reviews Microbiology 2018 16:12* [Internet] 16:731–744. Available
621 from: https://www.nature.com/articles/s41579-018-0075-3
622 Schönknecht G, Weber APM, Lercher MJ. 2014. Horizontal gene acquisitions by eukaryotes as
623 drivers of adaptive evolution. *BioEssays* [Internet] 36:9–20. Available from:
624 https://onlinelibrary.wiley.com/doi/full/10.1002/bies.201300095
625 Slot JC. 2017. Fungal Gene Cluster Diversity and Evolution. *Adv Genet* [Internet] 100:141–178.
626 Available from: https://pubmed.ncbi.nlm.nih.gov/29153399/

627 Soucy SM, Huang J, Gogarten JP. 2015. Horizontal gene transfer: building the web of life.
628 *Nature Reviews Genetics 2015 16:8* [Internet] 16:472–482. Available from:
629 https://www.nature.com/articles/nrg3962
630 Sweetlove LJ, Fernie AR. 2018. The role of dynamic enzyme assemblies and substrate
631 channelling in metabolic regulation. *Nature Communications 2018 9:1* [Internet] 9:1–12.
632 Available from: https://www.nature.com/articles/s41467-018-04543-8
633 Tsoka S, Ouzounis CA. 2000. Prediction of protein interactions: metabolic enzymes are
634 frequently involved in gene fusion. *Nature Genetics 2000 26:2* [Internet] 26:141–142.
635 Available from: https://www.nature.com/articles/ng1000_141
636 Watkinson SC. 2016. Physiology and Adaptation. *The Fungi: Third Edition*:141–187.
637 Wisecaver JH, Rokas A. 2015. Fungal metabolic gene clusters-caravans traveling across
638 genomes and environments. *Front Microbiol* [Internet] 6:1–11. Available from:
639 http://www.ncbi.nlm.nih.gov/pubmed/25784900
640 Wisecaver JH, Slot JC, Rokas A. 2014. The Evolution of Fungal Metabolic Pathways. *PLoS*
641 *Genet* 10.
642 Yanai I, Wolf YI, Koonin E V. 2002. Evolution of gene fusions: horizontal transfer versus
643 independent events. *Genome Biology 2002 3:5* [Internet] 3:1–13. Available from:
644 https://genomebiology.biomedcentral.com/articles/10.1186/gb-2002-3-5-research0024
645 Zhaxybayeva O, Doolittle WF. 2011. Lateral gene transfer. *Current Biology* 21:R242–R246.