Multiple *GAL* pathway gene clusters evolved independently and by different mechanisms in fungi

Jason C. Slot and Antonis Rokas[1]

Department of Biological Sciences, Vanderbilt University, Nashville, TN 37235

Edited* by John Doebley, University of Wisconsin, Madison, WI, and approved April 23, 2010 (received for review December 14, 2009)

A notable characteristic of fungal genomes is that genes involved in successive steps of a metabolic pathway are often physically linked or clustered. To investigate how such clusters of functionally related genes are assembled and maintained, we examined the evolution of gene sequences and order in the galactose utilization (*GAL*) pathway in whole-genome data from 80 diverse fungi. We found that *GAL* gene clusters originated independently and by different mechanisms in three unrelated yeast lineages. Specifically, the *GAL* cluster found in *Saccharomyces* and *Candida* yeasts originated through the relocation of native unclustered genes, whereas the *GAL* cluster of *Schizosaccharomyces* yeasts was acquired through horizontal gene transfer from a *Candida* yeast. In contrast, the *GAL* cluster of *Cryptococcus* yeasts was assembled independently from the *Saccharomyces*/*Candida* and *Schizosaccharomyces GAL* clusters and coexists in the *Cryptococcus* genome with unclustered *GAL* paralogs. These independently evolved *GAL* clusters represent a striking example of analogy at the genomic level. We also found that species with *GAL* clusters exhibited significantly higher rates of *GAL* pathway loss than species with unclustered *GAL* genes. These results suggest that clustering of metabolic genes might facilitate fungal adaptation to changing environments both through the acquisition and loss of metabolic capacities.

metabolic gene cluster | gene relocation | horizontal gene transfer | independent evolution

Fungal genes involved in successive steps of a pathway are frequently physically clustered (1–12). Evolutionary analysis of several different fungal gene clusters has shown that they have originated through different mechanisms (13–16). For example, the *DAL* cluster for allantoin utilization in *Saccharomyces cerevisiae* and close relatives originated through adaptive gene relocation (13), whereas a nitrate assimilation cluster in *Trichoderma* originated through horizontal gene transfer (14). However, it is unclear how these mechanisms shape the assembly of gene clusters, or how the evolutionary trajectories differ between species containing clusters and species in which the genes in a pathway are scattered across the genome.

To better understand the origins and maintenance of gene clusters over a large evolutionary timescale and a range of ecological conditions, we investigated the evolution of the Leloir galactose utilization (*GAL*) pathway in fungi. The detailed knowledge on *GAL* pathway function in *S. cerevisiae* (15, 16), and the abundance of genomic data from a wide diversity of fungal species (17), make it an excellent model pathway to address these questions. Furthermore, the relative galactose content varies substantially among different plant substrata (from hundreds of mg/g legume seeds and algal mats to less than 1 mg/g in some fruits) (18–21), suggesting that the natural substrates of fungi have likely provided ample opportunity for populations to evolve niche-dependent adaptations for galactose utilization.

The protein products of three *GAL* genes (*GAL1*, *GAL7*, and *GAL10*) are involved in four enzymatic steps (22). In the first step, the spontaneous conversion of β-D-galactose to α-D-galactose is accelerated by the mutarotase (aldose-1-epimerase) domain of Gal10p. Notably, the mutarotase domain of Gal10p in most Ascomycota yeasts is fused with the epimerase (UDP-galactose-4-epimerase) domain, but not in other fungal lineages. In the second step, α-D-galactose is phosphorylated into α-D-galactose-1-phosphate by Gal1p (galactokinase), whereas in the third step, UDP is transferred from UDP-α-D-glucose-1-phosphate to α-D-galactose-1-phosphate via Gal7p (galactose-1-phosphate uridylyl transferase), thereby freeing glucose-1-phosphate. In the fourth and final step, UDP-α-D-galactose-1-phosphate is used to regenerate UDP-α-D-glucose-1-phosphate by the epimerase domain of Gal10p.

Whereas *GAL1*, *GAL7*, and *GAL10* are unclustered in most fungi (23–27), they are clustered in *S. cerevisiae*, *Candida albicans*, *Schizosaccharomyces pombe*, *Cryptococcus neoformans*, and their relatives (23, 28, 29). Interestingly, the three *GAL* gene cluster-containing lineages (*Saccharomyces*/*Candida*, *Schizosaccharomyces*, and *Cryptococcus*) are very distantly related (29). Specifically, *Saccharomyces*, *Candida*, and relatives (budding yeasts) are members of Saccharomycotina (phylum Ascomycota), whereas the *Schizosaccharomyces* fission yeasts are nested within the early diverged clade Taphrinomycotina (phylum Ascomycota) (30). Saccharomycotina and Taphrinomycotina are estimated to have diverged ≈400 million years ago (31). The *Cryptococcus* encapsulated yeasts are even more distantly related to the other three lineages (*Saccharomyces*, *Candida*, and *Schizosaccharomyces*), as they belong to an entirely different phylum (Basidiomycota).

To better understand the origins and evolutionary maintenance of these patchily distributed *GAL* clusters, we identified and reconstructed the evolution of the *GAL1*, *GAL7*, and *GAL10* genes and their relative physical locations in a diverse set of 80 fungal genomes. Our evolutionary analyses indicate that *GAL* clusters originated by at least two different mechanisms, horizontal gene transfer (HGT) and gene relocation, twice independently. Furthermore, our results suggest that *GAL* genes in clusters are significantly more likely to be lost than their unclustered orthologs. These results suggest that metabolic gene clusters can have multiple origins and that they likely facilitate fungal adaptation both through the acquisition and loss of pathways.

Results and Discussion

Three Distinct Syntenic Types of *GAL* Gene Clusters Are Found in Three Unrelated Fungal Lineages. To determine the distribution of *GAL* clusters across fungi, we first examined the presence and physical location of *GAL1*, *GAL7*, *GAL10*, and their gene neighbors in 80 fungal genomes (Table S1). We found that *GAL1*, *GAL7*, and *GAL10* were unclustered in several major fungal clades (Fig. 1). In contrast, the three *GAL* genes were clustered in all *Saccharomyces*/*Candida* and *Schizosaccharomyces* species exam-

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

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

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

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.0914418107/-/DCSupplemental.

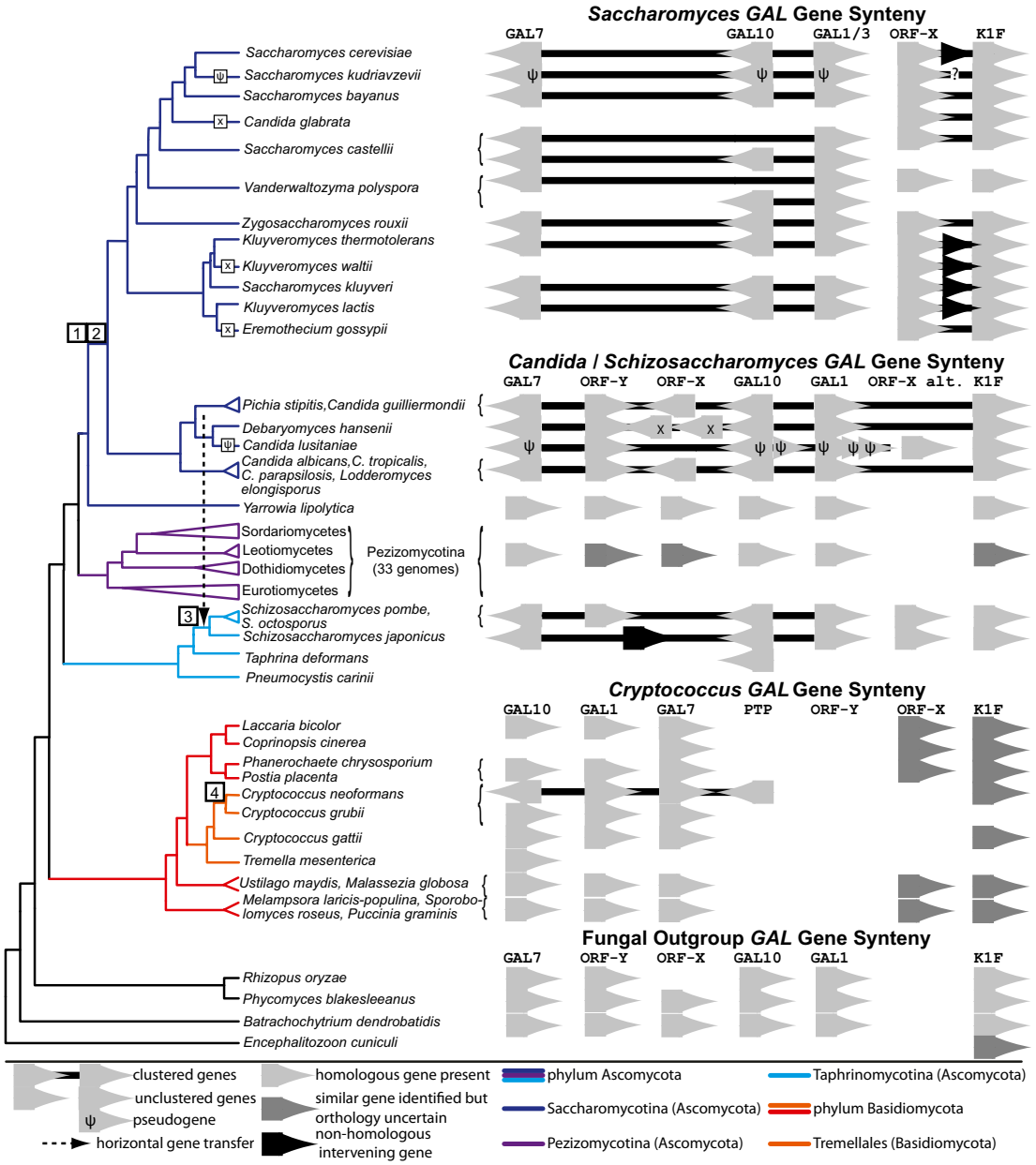

Fig. 1. Fungal *GAL* gene clusters originated twice independently by at least two different evolutionary mechanisms. The species tree of the 80 fungal genomes, constructed with information from several published multigene phylogenies (29, 62, 63), is shown on the left, whereas the gene order, orientation, and homology of genes found in *GAL* clusters are on the right. The major fungal lineages discussed in the main text have been color-coded. Boxed "x" symbols indicate species-specific *GAL* pathway losses, and boxed "ψ" symbols indicate species-specific *GAL* pathway pseudogenization events. *K1F* is a kinase that is syntenic with *ORF-X* in both *Candida* and *Saccharomyces*. "?" symbols indicate lack of information about physical linkage. The "*ORF-X alt.*" gene column shows unclustered *ORF-X* homologs. Major evolutionary events leading to the present distribution of the *GAL* cluster types across fungi are shown in numbered boxes (see also Table S5): 1. Origin of the ancestral *Candida*/*Saccharomyces GAL* cluster; 2. Origin of fused Gal10p, which contains both the epimerase and the mutarotase domains; 3. HGT of a *Candida GAL* cluster to a *Schizosaccharomyces* ancestor; and 4. Origin of the *Cryptococcus GAL* cluster.

ined (when present), as well as in *C. neoformans* and *C. grubii* (Fig. 1). Examination of the gene content ([] indicates genes not present in all species of the clade), gene orientation (indicated by +/−), and gene order of these *GAL* clusters identified three distinct syntenic types:

(*i*) The *Saccharomyces* type: (−*GAL7*, −*GAL10*, +*GAL1*).
(*ii*) The *Candida*/*Schizosaccharomyces* type: (−*GAL7*, [+*ORF-Y*], [−*ORF-X*], −*GAL10*, +*GAL1*, [+*K1F*]); *ORF-X* is one of the subunits of the transport protein particle complex of the *cis*-Golgi (32), *ORF-Y* is related to dTDP-glucose 4,6-dehydratases, and *K1F* is a cytoplasmic serine/threonine protein kinase.
(*iii*) The *Cryptococcus* type: (−*GAL10*, +*GAL1*, +*GAL7*, −*PTP*) (23, 28); *PTP* is a putative trehalose permease that is only present in *Cryptococcus GAL* clusters. The product of Gal7p is used to make trehalose (33), suggesting a potential functional link between PTP and the *GAL* pathway that could account for the association.

***GAL* Gene Cluster in *Schizosaccharomyces* Originated by HGT from *Candida*.** We were surprised that the *GAL* clusters of the distantly related *Candida* and *Schizosaccharomyces* clades are of the same syntenic type whereas the *GAL* clusters of the closely re-

lated *Candida* and *Saccharomyces* clades are of different syntenic types (Fig. 1). To test whether the similarity between the *Candida* and *Schizosaccharomyces GAL* clusters was the result of HGT, we reconstructed the phylogenies of all *Candida GAL* cluster genes (*Materials and Methods*). All four gene phylogenies robustly supported the placement of *Schizosaccharomyces* sequences within the *Candida* clade (Fig. 2 and Fig. S1 *A–D*, *F*, *H–K*, and *M*), and topology tests significantly rejected the monophyly of *Saccharomyces* and *Candida* sequences (Table S2). Furthermore, *Schizosaccharomyces* species also contain an additional unclustered *GAL10* paralog (Fig. 2) whose phylogenetic placement is consistent with the species phylogeny and whose protein product contains only the epimerase domain. In contrast, the protein product of the *Schizosaccharomyces* clustered *GAL10* contains both the epimerase and the mutarotase domains, similar to *GAL10* from *Saccharomyces* and *Candida* species. These data suggest that the *Schizosaccharomyces GAL* cluster was acquired through HGT from a *Candida* yeast. Interestingly, *Schizosaccharomyces* and *Candida* species frequently co-occur in natural and controlled fermentation environments (34, 35), providing a feasible ecological route via which the HGT event could have occurred.

In *C. albicans*, expression of *GAL1*, *GAL7*, and *GAL10* is likely regulated by Cph1p and Mig1p (28, 36, 37). Using a DNA binding motif prediction algorithm (*Materials and Methods*), we found a 10-bp motif that resembles a *C. albicans* Cph1p binding site upstream of the clustered *S. pombe GAL10* (Fig. S2). We also identified a 39-bp motif upstream of *GAL1* and *GAL7* in *C. lusitaniae*, *C. guillermondii*, and *S. octosporus*, which is also found upstream of *GAL7* in *S. pombe* (Table S3). The unique sharing of these two intergenic motifs between *Candida* and *Schizosaccharomyces* species provides additional support for the HGT of a *Candida GAL* cluster to *Schizosaccharomyces*, and raises the possibility that, in addition to the coding sequences, some of the regulatory components of the *GAL* cluster might have also been retained after the HGT. However, BLAST searches using *Sac-*

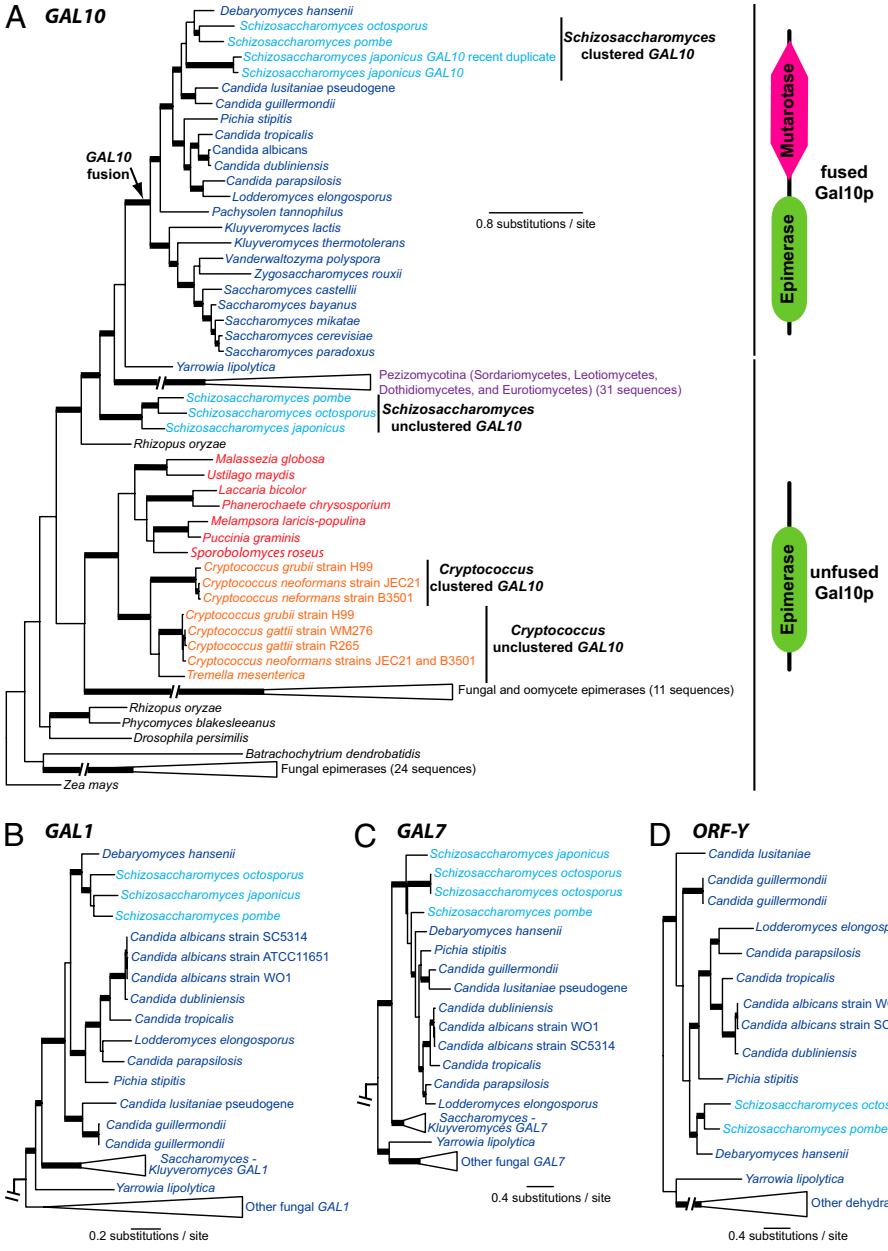

Fig. 2. Phylogeny of *GAL* gene cluster homologs across fungi. (*A*) ML phylogeny of *GAL10* homologs across fungi. The origin, distribution, and protein structure of fused and unfused Gal10 proteins across fungi are indicated both on the tree and on the right side of the panel. (*B*) ML phylogeny of *GAL1* homologs across fungi. (*C*) ML phylogeny of *GAL7* homologs across fungi. (*D*) ML phylogeny of *ORF-Y* homologs across fungi. Note that *Schizosaccharomyces* sequences are always nested within the *Candida* clade. Robustly supported clades (i.e., those supported by a maximum-likelihood bootstrap value ≥ 70% and a Bayesian posterior probability ≥ 0.95) are indicated by thickened branches. Species color-coding matches that of the major fungal lineages shown in Fig. 1.

charomyces Mig1p and *Candida* Cph1p sequences do not recover any homologs in *Schizosaccharomyces* genomes.

***GAL* Gene Clusters of *Saccharomyces* and *Candida* Are Evolutionarily Related and Originated Through Gene Relocation.** Comparison of the *Saccharomyces* and *Candida GAL* clusters suggests that the two cluster types are evolutionarily related, but distinct on account of intervening genes such as *ORF-X* and *ORF-Y*, and that the *GAL* genes in the last common ancestor of the *Saccharomyces* and *Candida* lineages were most likely clustered (Fig. 1). For example, the placement of *Saccharomyces* and *Candida GAL* genes in each gene phylogeny is consistent with their species phylogeny (Fig. 2 and Fig. S1 *A–E* and *H–L*). A common origin of the *Saccharomyces* and *Candida GAL* clusters is also supported by the unique shared presence of a mutarotase domain in all Gal10p found in the two lineages (Fig. 2), raising the possibility that the separate Gal10p domains fused after first relocating to form the cluster. Furthermore, the order and orientation of the three shared genes (*GAL1*, *GAL7*, and *GAL10*) in the *Saccharomyces* and *Candida GAL* cluster types is identical (Fig. 1), indicating that these three genes were clustered in their common ancestor.

Two lines of evidence suggest that the ancestral *Saccharomyces* and *Candida GAL* cluster likely originated through relocation of native genes. First, *GAL1*, *GAL7*, *GAL10*, and *ORF-X* phylogenies are consistent with the fungal species phylogeny (Fig. 2 and Fig. S1 *A–E* and *H–L*; except for the HGT of the *Candida GAL* cluster to *Schizosaccharomyces*), providing support for the vertical origin and inheritance of each gene in the ancestral *Saccharomyces*/*Candida GAL* cluster. Second, the *GAL* genes in the outgroups *Yarrowia lipolytica* and the entire Pezizomycotina clade are unclustered (Fig. 1), suggesting that the native gene relocation events leading to the origin of the ancestral *Saccharomyces*/*Candida GAL* cluster predated the divergence of the *Saccharomyces* and *Candida* lineages but postdated their divergence from Pezizomycotina and *Y. lipolytica*. Following the divergence of the *Saccharomyces* and *Candida* lineages, the gene content and order of the ancestral *GAL* cluster was subsequently modified in at least one of these lineages (Fig. 1), evolving into the presently distinct, but related, *Saccharomyces* and *Candida GAL* cluster types.

***Cryptococcus GAL* Gene Cluster Originated Recently and Independently from Ascomycota *GAL* Clusters.** The *Cryptococcus GAL* cluster is composed of *GAL10*, *GAL1*, *GAL7*, and *PTP* and is present only in *C. neoformans* and *C. grubii* but absent from *C. gattii* (Table 1, Figs. 1 and 2, and Fig. S1 *A–D* and *H–K*). Furthermore, both *GAL* cluster-containing and cluster-lacking *Cryptococcus* species contain unclustered *GAL1*, *GAL7*, and *GAL10* paralogs (Figs. 1 and 2 and Fig. S1 *A–D* and *H–K*).

Several lines of evidence indicate that the *GAL* cluster in *Cryptococcus* originated independently from the Ascomycota *GAL* clusters. First, the order and orientation of the *GAL* genes differs between the *Cryptococcus* and both Ascomycota clusters (Fig. 1). Second, the placement of *Cryptococcus* sequences in *GAL1*, *GAL7*, and *GAL10* phylogenies is consistent with the species phylogeny (Fig. 2 and Fig. S1 *A–D* and *H–K*), suggesting that the genes found in the *Cryptococcus GAL* cluster are native and did not originate via HGT from Ascomycota or other outgroup lineages. Third, the *Cryptococcus* Gal10p protein lacks the mutarotase domain present in all clustered Gal10p proteins from Ascomycota. Fourth, ancestral state reconstruction analyses significantly support the hypothesis that the Basidiomycota common ancestor was unclustered (unclustered versus functional clustered, likelihood ratio test statistic, $\Delta l = 7.35$; unclustered versus nonfunctional or absent $\Delta l = 6.49$) (*Materials and Methods*).

The presence of both clustered and unclustered *GAL* genes in *C. neoformans* and *C. grubii* makes it difficult to determine which of the two pathway conformations is ancestral and which is derived. Interestingly, synteny in the genomic regions flanking the *GAL* cluster (Table 1) and in the regions flanking the unclustered *GAL* genes is conserved in all *Cryptococcus* species examined, suggesting that several different scenarios involving either multiple duplications and losses or HGT from an unidentified Basidiomycete could potentially explain the origin of the *Cryptococcus GAL* cluster. Furthermore, a cluster of three retrotransposon-related sequences two genes away from the *GAL* cluster indicates that retrotransposon-mediated rearrangements (38) might have also contributed to the cluster's origin. Identifying the precise mechanism of origin of the *Cryptococcus GAL* cluster will likely require extensive sampling of *GAL* genes from early-diverging Basidiomycota.

Cryptococcus is the only lineage in which certain species contain full sets of functional clustered and unclustered *GAL1*, *GAL7*, and *GAL10* genes (Fig. 2 and Fig. S1 *A–D* and *H–K*). Importantly, the presence of these two *GAL* gene sets appears to be correlated with significant differences in pathogenicity between species and in function between paralogs. For example, *C. gattii*, which contains only unclustered *GAL* genes, is pathogenic to healthy individuals, whereas *C. neoformans* and *C. grubii*, which contain both clustered and unclustered *GAL* genes, are opportunistically pathogenic to immune-compromised individuals (39). Furthermore, although the unclustered *C. neoformans GAL10* paralog participates in galactose metabolism at 37 °C and is required for pathogenicity, the clustered paralog participates in galactose metabolism at all viable temperatures and is not required for pathogenicity (23, 40).

***GAL* Pathway Loss Is Accelerated in Species That Contain *GAL* Gene Clusters.** It has been suggested that metabolic gene clustering might be widespread because it facilitates HGT of sets of functionally related genes (41, 42). However, we noted that in our case the presence of *GAL* clusters is frequently associated with loss of the pathway. For example, a number of species in the *Saccharomyces* clade (Fig. 1) appear to have either lost (*Kluyveromyces waltii*, *Eremothecium gossypii*, and *C. glabrata*) or pseudogenized (*S. kudriavzevii*) their entire *GAL* pathways (43). Furthermore, we found that *GAL1* and *GAL7* in *C. lusitaniae* each contains a single nonsense point mutation, whereas *GAL10* was pseudogenized by one of three partial duplications of the 3′ region of *GAL1*. Because the *C. lusitaniae GAL* genes are still relatively intact, this pathway pseudogenization is likely to have been much more recent than that previously described in *S. kudriavzevii* (43). In contrast, we did not observe any wholesale pathway losses or pseudogenizations

Table 1. The genomic region flanking the *Cryptococcus GAL* gene cluster

Species strain (position)	−2	−1	*GAL10*	*GAL1*	*GAL7*	*PTP*	+1	+2
C. neoformans B3501 (chr. 13: 129414–229376)	+	+	+	+	+	+	+	+
C. neoformans JEC21 (chr. 13: 119359–205757)	+	+	+	+	+	+	+	+
C. grubii H99 (chr. 12: 136817–190866)	+	+	+	+	+	+	+	+
C. gattii WM276 (ctg. 900: 2721–6092)	+	+	−	−	−	−	+	+
C. gattii R265 (ctg. 1.13: 104310–87982)	+	+	−	−	−	−	+	+

+, orthologous gene present in conserved orientation; −, orthologous gene absent.

across a broad sample of fungal genomes with unclustered *GAL* genes (Fig. 1).

To test the hypothesis that gene clusters might also facilitate the loss of galactose assimilation, we examined the relative rate of transitions between either of two functional states [*GAL* pathway with genes clustered (C) or unclustered (U)] to a nonfunctional state [*GAL* pathway absent either through gene pseudogenization or loss (N)]. We found that C→N transitions were significantly more frequent than U→N transitions (a model in which C→N and U→N transitions were constrained to the same rate yielded a significantly lower likelihood than the unconstrained model; $\Delta l = 8.21$, P value < 0.02) (Table S4). These data suggest that rates of loss of galactose assimilation are significantly elevated in lineages containing *GAL* clusters.

Conclusion

We have found that fungal *GAL* gene clusters originated twice independently (once within Ascomycetes and once within Basidiomycetes). Analogous gene clusters have also been observed in the nitrate assimilation pathways of fungi and algae (14) and the *Iroquois* homeobox clusters in distantly related animal phyla (44). Interestingly, the specifics differ substantially in each case. For example, the animal *Iroquois* homeobox gene clusters have originated through independent series of tandem duplications (44), whereas the nitrate assimilation gene clusters arose partly through the independent assembly of functionally similar but evolutionarily unrelated (both within and between clusters) genes (14). In contrast, the *GAL* gene clusters described here originated via the assembly of orthologous (between clusters) but unrelated (within clusters) genes.

It has been suggested that metabolic gene clustering originates through selection for more precise coordination of pathway gene regulation (45) (the "coregulation" model), thus allowing the more efficient processing of potentially deleterious metabolic intermediates. In support of the coregulation model, the intergenic region between the *S. cerevisiae GAL1* and *GAL10* genes contains shared regulatory binding sites (15, 46–48), and galactose-1-phosphate, one of the intermediate products of galactose metabolism, is toxic (49).

An alternative hypothesis is that metabolic gene clusters originate and persist because of the selective advantage of the clustered state itself [the "selfish cluster" model (42, 50)]. In this model, genes in clusters propagate more readily by HGT than their unclustered counterparts (50). It has yet to be shown that clustered genes have fitness advantages over and above (and in some cases to the detriment of) the fitness benefit conferred by them to the host organism that would justify the term "selfish." However, several fungal gene clusters associated with substrate-specific and secondary metabolic pathways, including the *GAL* cluster, have spread via HGT (14, 51, 52).

Consideration of the mechanisms underlying *GAL* gene cluster evolution suggests a model for the evolution of metabolic gene clusters in fungi that contains components of both the coregulation and selfish cluster models. Under this hybrid model, gene clusters originate by positive selection for coordinated regulation (i.e., *GAL* in Saccharomycotina; see also ref. 13). Once formed, clustered nonessential genes are more likely to spread to new lineages by HGT (i.e., *GAL* in Schizosaccharomyces) than their unclustered counterparts. However, we have found that clustering has also made wholesale pathway loss significantly more likely in *GAL* cluster-containing species. This is likely due to differences between species with clustered or unclustered genes in the relative ease of pathway removal and in the potential for generating deleterious metabolic intermediates. For example, species with unclustered *GAL* genes require three independent gene deletions, which can lead to deleterious metabolites (e.g., galactose-1-phosphate). In contrast, species with *GAL* gene clusters can lose all three pathway genes in a single mutational event, bypassing the risk of toxic metabolite accumulation. This suggests that clustered genes likely face a fitness tradeoff between increased spread by HGT in favorable environments and increased loss in unfavorable ones, thus facilitating fungal adaptation to diverse and changing nutritional environments.

Materials and Methods

***GAL* Gene Cluster Identification.** We analyzed the distribution, clustering, and phylogenies of genes found in syntenic blocks in *Debaryomyces hansenii* and *C. neoformans* (Table S3), including orthologs of *GAL1*, *GAL7*, and *GAL10*, the intervening genes *ORF-X* and *ORF-Y* (28), and the flanking genes *K1F* and *PTP* (a putative trehalose permease). In synteny searches, putative gene clusters were defined as physically linked groups of genes, no one member of which is more than seven genes away from any gene that is a member of the *GAL* pathway. Homologous *GAL* clusters were inferred from a sample of 80 fungal genomes (Table S1) using a series of custom perl scripts that rely on BLAST similarity searches (53) and gene order data from publicly available genome project assemblies or from http://fungalgenomes.org (maintained by Jason Stajich).

Phylogenetic Analysis. Genomic protein sequences were combined with similar sequences retrieved from the GenBank nr database using BLASTP. Groups of homologous genes were identified using OrthoMCL (54), treating all sequences as within-genome. Each homologous gene matrix was aligned using MAFFT 6.624 (55) and then manually curated. Gene phylogenies were inferred using maximum likelihood (ML) (in RAxML version 7.2) (56) under a JTT plus GAMMA model, and Bayesian analysis (in MrBayes version 3.1.2) (57, 58) under mixed protein models. For the ML analysis, robustness of inference was assessed by running 100 bootstrap replicates. For the Bayesian analysis, two independent analyses were run, using four Markov Chain Monte Carlo (MCMC) chains (one cold and three hot) for 1,000,000 generations. Trees were sampled every 100 generations and the first 2,000 sampled trees were discarded as burn-in, by which point stationarity was already reached.

DNA Binding Motif Prediction. DNA binding motif predictions were performed in the MEME Suite (version 4.3.0) motif-based sequence analysis tools (59). De novo MEME searches were performed on palindromes of all 750-bp regions immediately upstream of the corresponding *GAL* gene open reading frames. Motif position-specific scoring matrices were used to search for similar sequences not found in the initial search using the MAST algorithm (in the MEME Suite).

Phylogenetic Hypothesis Testing. The approximately unbiased (AU) test (60) implemented in CONSEL (61) was used to test alternative gene tree topologies (Table S2). The transition rates between different functional and clustering states of *GAL1*, *GAL7*, and *GAL10* (functional and clustered: C; functional and unclustered: U; putatively nonfunctional: N) and the ancestral states of *GAL* gene clustering in Ascomycota and Basidiomycota were analyzed using the Bayes Multistate (62) module in the BayesTraits package. The species phylogeny was constructed from an RNA polymerase II second-largest subunit (RPB2) amino acid data matrix using Bayesian inference. Two parallel MCMC chains ran for 1 million generations under mixed protein models in MrBayes, sampling every 100th tree; 17,998 trees were retained after the burn-in. Average likelihoods (l) were calculated for reconstructions across the sample of trees with C↔N and U↔N transition rates unconstrained and constrained to be equal. Alternative models were compared with the likelihood ratio test statistic $\Delta l = 2 \cdot$(log[harmonic mean(l better model)] − log[harmonic mean(l worse model)]). Log-likelihood ratio values greater than 2 were considered significant. Ancestral states for nodes were compared to "fossilized" alternative reconstructions using the methods described above. Because of the confounding effect of HGT on ancestral state reconstruction, and the limited number of full genomes available from Taphrinomycotina, clustering states in *Schizosaccharomyces* were coded as ambiguous.

ACKNOWLEDGMENTS. We thank Chris T. Hittinger, Kris McGary, Patrick Abbot, and Mark Johnston for providing comments on an earlier draft and Jason Stajich for making the gene order data from several fungal genome project assemblies publicly available. This work was conducted in part using the resources of the Advanced Computing Center for Research and Education at Vanderbilt University. This work was supported by funds provided by the Searle Scholars Program (to A.R.) and the National Science Foundation (DBI-0805625 to J.C.S. and DEB-0844968 to A.R.).

1. Keller NP, Hohn TM (1997) Metabolic pathway gene clusters in filamentous fungi. *Fungal Genet Biol* 21:17–29.
2. Lopez F, Leube M, Gil-Mascarell R, Navarro-Aviñó JP, Serrano R (1999) The yeast inositol monophosphatase is a lithium- and sodium-sensitive enzyme encoded by a non-essential gene pair. *Mol Microbiol* 31:1255–1264.
3. Schierová M, Linka M, Pazoutová S (2000) Sulfate assimilation in *Aspergillus terreus*: Analysis of genes encoding ATP-sulfurylase and PAPS-reductase. *Curr Genet* 38: 126–131.
4. Watanabe S, Saimura M, Makino K (2008) Eukaryotic and bacterial gene clusters related to an alternative pathway of nonphosphorylated L-rhamnose metabolism. *J Biol Chem* 283:20372–20382.
5. Yu J, Chang P, Bhatnagar D, Cleveland TE (2000) Cloning of a sugar utilization gene cluster in *Aspergillus parasiticus*. *Biochim Biophys Acta* 1493:211–214.
6. Pinedo C, et al. (2008) Sesquiterpene synthase from the botrydial biosynthetic gene cluster of the phytopathogen *Botrytis cinerea*. *ACS Chem Biol* 3:791–801.
7. Kawaide H (2006) Biochemical and molecular analyses of gibberellin biosynthesis in fungi. *Biosci Biotechnol Biochem* 70:583–590.
8. Kimura N, Tsuge T (1993) Gene cluster involved in melanin biosynthesis of the filamentous fungus *Alternaria alternata*. *J Bacteriol* 175:4427–4435.
9. Tudzynski P, et al. (1999) Evidence for an ergot alkaloid gene cluster in *Claviceps purpurea*. *Mol Gen Genet* 261:133–141.
10. Siewers V, Kokkelink L, Smedsgaard J, Tudzynski P (2006) Identification of an abscisic acid gene cluster in the grey mold *Botrytis cinerea*. *Appl Environ Microbiol* 72: 4619–4626.
11. Dean RA (2007) Fungal gene clusters. *Nat Biotechnol* 25:67.
12. Young C, McMillan L, Telfer E, Scott B (2001) Molecular cloning and genetic analysis of an indole-diterpene gene cluster from *Penicillium paxilli*. *Mol Microbiol* 39: 754–764.
13. Wong S, Wolfe KH (2005) Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet* 37:777–782.
14. 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.
15. Johnston M (1987) A model fungal gene regulatory mechanism: The GAL genes of *Saccharomyces cerevisiae*. *Microbiol Rev* 51:458–476.
16. Bhat PJ, Murthy TV (2001) Transcriptional control of the GAL/MEL regulon of yeast *Saccharomyces cerevisiae*: Mechanism of galactose-mediated signal transduction. *Mol Microbiol* 40:1059–1066.
17. Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B (2005) Genomics of the fungal kingdom: Insights into eukaryotic biology. *Genome Res* 15:1620–1631.
18. Nierop K, van Lagen B, Buurman P (2001) Composition of plant tissues and soil organic matter in the first stages of a vegetation succession. *Geoderma* 100:1–24.
19. Marsilio V, Campestre C, Lanza B, De Angelis M (2001) Sugar and polyol compositions of some European olive fruit varieties (*Olea europaea* L.) suitable for table olive purposes. *Food Chem* 72:485–490.
20. Gross KC, Acosta PB (1991) Fruits and vegetables are a source of galactose: Implications in planning the diets of patients with galactosaemia. *J Inherit Metab Dis* 14:253–258.
21. Acosta PB, Gross KC (1995) Hidden sources of galactose in the environment. *Eur J Pediatr* 154(7, Suppl 2):S87–S92.
22. Holden HM, Rayment I, Thoden JB (2003) Structure and function of enzymes of the Leloir pathway for galactose metabolism. *J Biol Chem* 278:43885–43888.
23. Moyrand F, Lafontaine I, Fontaine T, Janbon G (2008) UGE1 and UGE2 regulate the UDP-glucose/UDP-galactose equilibrium in *Cryptococcus neoformans*. *Eukaryot Cell* 7: 2069–2077.
24. Martin F, et al. (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. *Nature* 452:88–92.
25. Martinez D, et al. (2008) Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat Biotechnol* 26:553–560.
26. Ma LJ, et al. (2009) Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet* 5:e1000549.
27. Galagan JE, et al. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438:1105–1115.
28. Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M (2007) Transcriptional rewiring of fungal galactose-metabolism circuitry. *Curr Biol* 17:1007–1013.
29. James TY, et al. (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* 443:818–822.
30. Sipiczki M (2000) Where does fission yeast sit on the tree of life? *Genome Biol* 1: reviews1011.1–reviews1011.4.
31. Taylor JW, Berbee ML (2006) Dating divergences in the Fungal Tree of Life: Review and new analyses. *Mycologia* 98:838–849.
32. Sacher M, et al. (1998) TRAPP, a highly conserved novel complex on the *cis*-Golgi that mediates vesicle docking and fusion. *EMBO J* 17:2494–2503.
33. Cabib E, Leloir LF (1958) The biosynthesis of trehalose phosphate. *J Biol Chem* 231: 259–275.
34. Pataro C, et al. (2000) Yeast communities and genetic polymorphism of *Saccharomyces cerevisiae* strains associated with artisanal fermentation in Brazil. *J Appl Microbiol* 89:24–31.
35. Lachance MA (1995) Yeast communities in a natural tequila fermentation. *Antonie van Leeuwenhoek* 68:151–160.
36. Murad AM, et al. (2001) Transcript profiling in *Candida albicans* reveals new cellular functions for the transcriptional repressors CaTup1, CaMig1 and CaNrg1. *Mol Microbiol* 42:981–993.
37. Lavoie H, Hogues H, Whiteway M (2009) Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. *Curr Opin Microbiol* 12:655–663.
38. Sun S, Xu J, Lorenz M (2009) Chromosomal rearrangements between serotype A and D strains in *Cryptococcus neoformans*. *PLoS One* 4:e5524.
39. Xu J, Yan Z, Guo H (2009) Divergence, hybridization, and recombination in the mitochondrial genome of the human pathogenic yeast *Cryptococcus gattii*. *Mol Ecol* 18:2628–2642.
40. Moyrand F, Fontaine T, Janbon G (2007) Systematic capsule gene disruption reveals the central role of galactose metabolism on *Cryptococcus neoformans* virulence. *Mol Microbiol* 64:771–781.
41. Lawrence J (1999) Selfish operons: The evolutionary impact of gene clustering in prokaryotes and eukaryotes. *Curr Opin Genet Dev* 9:642–648.
42. Walton JD (2000) Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: An hypothesis. *Fungal Genet Biol* 30:167–171.
43. Hittinger CT, Rokas A, Carroll SB (2004) Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci USA* 101:14144–14149.
44. Irimia M, Maeso I, Garcia-Fernàndez J (2008) Convergent evolution of clustering of Iroquois homeobox genes across metazoans. *Mol Biol Evol* 25:1521–1525.
45. Hurst LD, Pal C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 5:299–310.
46. West RW, Jr, Yocum RR, Ptashne M (1984) *Saccharomyces cerevisiae* GAL1-GAL10 divergent promoter region: Location and function of the upstream activating sequence UASG. *Mol Cell Biol* 4:2467–2478.
47. Yocum RR, Hanley S, West R, Jr, Ptashne M (1984) Use of lacZ fusions to delimit regulatory elements of the inducible divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol Cell Biol* 4:1985–1998.
48. Lutfiyya LL, et al. (1998) Characterization of three related glucose repressors and genes they regulate in *Saccharomyces cerevisiae*. *Genetics* 150:1377–1391.
49. Fridovich-Keil JL (2006) Galactosemia: The good, the bad, and the unknown. *J Cell Physiol* 209:701–705.
50. Lawrence JG, Roth JR (1996) Selfish operons: Horizontal transfer may drive the evolution of gene clusters. *Genetics* 143:1843–1860.
51. Patron NJ, et al. (2007) Origin and distribution of epipolythiodioxopiperazine (ETP) gene clusters in filamentous ascomycetes. *BMC Evol Biol* 7:174.
52. Khaldi N, Collemare J, Lebrun MH, Wolfe KH (2008) Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. *Genome Biol* 9:R18.
53. Altschul SF, et al. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402.
54. Li L, Stoeckert CJ, Jr, Roos DS (2003) *Genome Res* 13:2178–2189.
55. Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* 9:286–298.
56. Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
57. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
58. Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.
59. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 2:28–36.
60. Shimodaira H (2002) An approximately unbiased test of phylogenetic tree selection. *Syst Biol* 51:492–508.
61. Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247.
62. Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral character states on phylogenies. *Syst Biol* 53:673–684.
63. Fitzpatrick DA, Logue ME, Stajich JE, Butler G (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol Biol* 6:99.
64. Suh SO, Blackwell M, Kurtzman CP, Lachance MA (2006) Phylogenetics of Saccharomycetales, the ascomycete yeasts. *Mycologia* 98:1006–1017.