Lineage-specific selection and the evolution of virulence in the *Candida* clade

Sheena D. Singh-Babak[a,b], Tomas Babak[c], Hunter B. Fraser[d], and Alexander D. Johnson[a,e,f,1]

[a]Department of Microbiology and Immunology, University of California, San Francisco, CA 94158; [b]Assay Development, Lucira Health, Emeryville, CA 94608; [c]Department of Biology, Queens University, Kingston, ON K7L 3N6, Canada; [d]Department of Biology, Stanford University, Stanford, CA 94305; [e]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158; and [f]Microbiome Initiative, Chan Zuckerberg Biohub, San Francisco, CA 94158

Edited by Jasper Rine, University of California, Berkeley, CA, and approved December 16, 2020 (received for review September 22, 2020)

Candida albicans **is the most common cause of systemic fungal infections in humans and is considerably more virulent than its closest known relative, *Candida dubliniensis*. To investigate this difference, we constructed interspecies hybrids and quantified mRNA levels produced from each genome in the hybrid. This approach systematically identified expression differences in orthologous genes arising from *cis*-regulatory sequence changes that accumulated since the two species last shared a common ancestor, some 10 million y ago. We documented many orthologous gene-expression differences between the two species, and we pursued one striking observation: All 15 genes coding for the enzymes of glycolysis showed higher expression from the *C. albicans* genome than the *C. dubliniensis* genome in the interspecies hybrid. This pattern requires evolutionary changes to have occurred at each gene; the fact that they all act in the same direction strongly indicates lineage-specific natural selection as the underlying cause. To test whether these expression differences contribute to virulence, we created a *C. dubliniensis* strain in which all 15 glycolysis genes were produced at modestly elevated levels and found that this strain had significantly increased virulence in the standard mouse model of systemic infection. These results indicate that small expression differences across a deeply conserved set of metabolism enzymes can play a significant role in the evolution of virulence in fungal pathogens.**

microbial pathogenesis | natural selection | allele-specific expression

Microbial pathogens of humans typically have at least one close, nonpathogenic relative, and comparisons between the two provide a powerful entry point to identify and study disease-causing determinants. This approach has revealed numerous bacterial pathogenicity islands, clusters of genes required for a given strain or species to cause disease in humans. Because these gene clusters can be horizontally transferred, they can often be initially identified by comparing genome sequences of pathogenic and nonpathogenic strains. The situation with fungal pathogens is substantially different. Like their human hosts, these microbes are eukaryotes, and genes that work together are typically dispersed across different chromosomes rather than clustered; not surprisingly, horizontal transfer of groups of cofunctioning genes from one fungal species to another is rare. Thus, identifying and understanding how groups of genes work together to contribute to virulence remains a challenge in fungi.

In this report, we consider two fungal pathogens that are closely related but differ in their virulence. *Candida albicans* is an opportunistic pathogen; it is a component of the normal human microbiota but is also the leading cause of systemic fungal infections in humans, which can have mortality rates as high as 40%. It is highly virulent when injected into the tail vein of mice, a standardized laboratory procedure that initiates a systemic blood stream infection. *Candida dubliniensis* is the closest known relative of *C. albicans*; it was first identified from the oral cavity of an HIV-infected patient and is typically found only as secondary infections. Although it is found throughout the world, *C. dubliniensis* is much less prevalent than *C. albicans* in the clinic; it is also less virulent in the mouse model of systemic infection, based on time-to-illness measurements (1–5).

C. albicans and *C. dubliniensis* last shared a common ancestor nominally 10 million y ago, and their genomes are very similar in terms of gene content and synteny. A small number of individual genes are "missing" in one species compared to the other but there are also wide-scale differences in the expression of those genes conserved in both species (6–10). It has been proposed that both types of differences—in gene content and in regulation—contribute to the difference in pathogenicity, although neither idea has been directly demonstrated.

In this report, we used an approach known as "allele-specific expression" to highlight patterns of gene-expression differences between *C. albicans* and *C. dubliniensis* and to identify those that bear the hallmarks of selection (11–15). Specifically, we created an interspecies hybrid between *C. albicans* (strain SC5314) and *C. dubliniensis* (strain CD36) by forcing them to mate with each other (16) (Fig. 1*A*). Both strains were originally isolated from human patients, and both have been studied extensively in the laboratory. In the hybrid, both parental genomes reside in the same cell, so any bias in mRNA levels must be due to *cis*-acting sequences specific to that genome. We show that ∼40% of orthologous gene pairs show statistically significant differences in their mRNA levels in the interspecies hybrid, although many of these differences are small in magnitude. We searched for

Significance

Of the many microbial species on earth, only a small number are able to thrive in humans and cause disease. Comparison of closely related pathogenic and nonpathogenic species can therefore be useful in identifying key features that contribute to virulence. We created interspecies hybrids between *Candida albicans*, a prevalent fungal pathogen of humans, and *Candida dubliniensis*, a close, but much less pathogenic, relative. By comparing genome-wide expression differences between the two genomes in the same cell, we surmised that since the two species diverged from a common ancestor, natural selection has acted upon the expression level of an ancient metabolic pathway, illustrating that pathogenicity traits can arise over evolutionary timescales through small expression changes in deeply conserved proteins.

Author contributions: S.D.S.-B., H.B.F., and A.D.J. designed research; S.D.S.-B. performed research; S.D.S.-B., T.B., and A.D.J. contributed new reagents/analytic tools; S.D.S.-B., T.B., and H.B.F. analyzed data; and S.D.S.-B., H.B.F., and A.D.J. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: ajohnson@cgl.ucsf.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2016818118/-/DCSupplemental.

Published March 15, 2021.

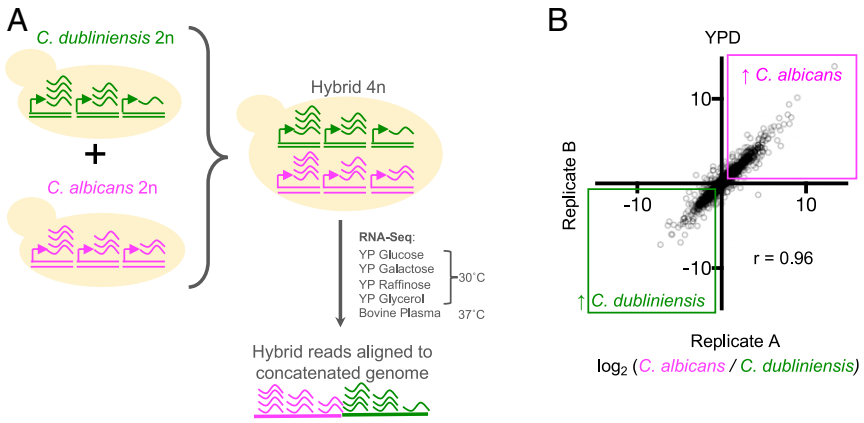

Fig. 1. RNA-seq of an interspecies hybrid systematically identifies gene-expression differences between *C. albicans* and *C. dubliniensis* that arose through *cis*-acting changes that accumulated since the two species shared a common ancestor. (*A*) *C. albicans* and *C. dubliniensis* (both diploid) were mated to a create tetraploid interspecies hybrid strain, using mating and auxotrophic marker complementation. We performed RNA-seq in the hybrid strain in the conditions listed, as well as the parent diploid strains in bovine plasma at 37 °C. We aligned RNA-seq reads to a concatenated *C. albicans*–*C. dubliniensis* genome and excluded reads mapping to both genomes or to multiple locations within a single genome (<5% of total reads). Systematic gene-expression differences of orthologous genes between the two species were assessed. Since both genomes were in the same *trans*-acting environment (that is, the hybrid strain), all measurable gene expression changes must be due to *cis*-acting changes. (*B*) Correlation between two replicate RNA-seq experiments of the interspecies hybrid grown in YPD at 30 °C. Each point represents the $\log_2$ ratio of reads from the *C. albicans* genome versus reads from the *C. dubliniensis* genome in the interspecies hybrid, for each orthologous gene pair, with experimental replicates plotted on each axis. Points in the upper right quadrant represent genes with higher expression from the *C. albicans* genome than *C. dubliniensis* genome in the interspecies hybrid, and points in the lower left quadrant represent genes with higher expression from the *C. dubliniensis* genome than the *C. albicans* genome in that same hybrid, with Spearman correlation values shown in the lower right. We observed very high correlation between replicate experiments for all conditions tested (*SI Appendix*, Figs. S1 and S2).

pathways that showed a systematic up-regulation from one genome compared to the other, and were led to the genes for glycolysis, all 15 of which exhibited increased expression from the C. *albicans* genome compared with the *C. dubliniensis* genome. This pattern requires *cis*-acting changes to have accumulated at each of the 15 genes, all resulting in higher mRNA expression from the *C. albicans* alleles in the hybrid. The likelihood of this pattern emerging by chance is extremely low, and the observation strongly implies that this nonrandom pattern is the result of natural selection.

To test whether increased glycolysis gene expression could account for at least some of the virulence difference between *C. albicans* and *C. dubliniensis*, we created a *C. dubliniensis* strain in which the glycolysis genes were modestly overexpressed and found that this engineered strain showed enhanced virulence (compared with the parental strain) in the standard mouse model for disseminated candidiasis. This result indicates the importance of metabolic fitness in the establishment of fungal infections. More broadly, our results suggest that subtle changes in the expression of genes that are deeply conserved in all branches of life can play important roles in the evolution of specific pathogens.

Results

We performed Illumina sequencing of poly-A-selected RNA (RNA-seq) extracted from freshly constructed *C. albicans*–*C. dubliniensis* hybrids grown in a range of standard laboratory conditions (2% glucose, 2% galactose, 2% raffinose, and 2% glycerol at 30 °C) as well as a condition that more closely resembles the environment in vivo (bovine plasma supplemented with 2% glucose at 37 °C). The *C. albicans* and *C. dubliniensis* genome have sufficiently diverged such that greater than 99% of the sequencing reads could unambiguously be assigned to one genome or the other.

We found that ∼40% of all orthologous genes showed statistically significant allele-specific expression differences (*Materials and Methods*, Fig. 1*B*, and Dataset S1). We note that, for highly expressed genes with many sequence reads, very small differences in mRNA levels can reach statistical significance, although such small differences—taken on their own—are not necessarily biologically meaningful. Our overall results indicate that *cis*-acting changes underlie much of the previously documented expression differences between *C. albicans* and *C. dubliniensis* when compared as separate species (6–10). This conclusion rests on the assumption that the "*trans*" acting factors from both species are expressed in the hybrid. We also sequenced the RNA from the two parental strains (grown under conditions identical to the hybrid) and found that most of the changes observed in the hybrid were also observed, to at least some degree, in the parents (*SI Appendix*, Fig. S3). The high degree of *cis*-regulatory divergence is also consistent with allele-specific expression analyses carried out with other fungal hybrid species, (for example, *Saccharomyces cerevisiae* × *Saccharomyces paradoxus* or *S. cerevisiae* × *Saccharomyces bayanus*) (11, 15, 17–22). Although it is difficult to compare our results quantitively with the *Saccharomyces* studies (due to differences in the depth of RNA-seq and the methodologies behind the analyses; see, for example ref. 23), the broad trends appear to be similar.

Using gene ontology (GO) categories, we searched for sets of genes whose expression differences between the *C. albicans* and *C. dubliniensis* genome in the hybrid bear signatures of lineage-specific selection; that is, they show nonrandom directional expression from one genome compared with the other. We carried out this analysis in two different ways, a rank-sum test (which considers both the magnitudes and the directionalities of expression differences) and a sign test (which considers only the directionalities of the differences) (23) (Table 1). Both analyses revealed strong directional expression biases in genes involved in filamentous growth and in glycolysis (Table 1). Filamentous growth is represented by several GO terms (containing overlapping sets of genes), and it is well documented in the literature that *C. albicans* and *C. dubliniensis* differ in this phenotype (24). Although it has been proposed that this difference contributes to the virulence difference between the species, mutations that increase filamentous growth of *C. dubliniensis* did not result in increased virulence (25). Moreover, there is only a poor

Table 1. GO terms under lineage-specific selection between *C. albicans* and *C. dubliniensis*

GO term annotation	Rank sum test		Sign test	
	P value	FDR	*P* value	FDR
glycolysis_15	8.44E-09	3.81E-06	4.32E-04	1.15E-01
Glycolysis	1.96E-08	1.76E-05	7.25E-04	1.54E-01
Filamentous growth of a population of unicellular organisms	1.11E-07	4.99E-05	3.28E-04	1.16E-01
Filamentous growth	1.42E-05	3.20E-03	2.03E-04	1.08E-01
Negative regulation of filamentous growth of a population of unicellular organisms	8.65E-05	9.72E-03	1.21E-03	1.84E-01
Cytokinetic cell separation	3.02E-03	1.18E-01	9.54E-04	1.69E-01

Searching all GO terms from the Candida Genome Database (CGD) that contained five or more genes (1,068), we assessed departure from null nondirectional expression using a rank-sum test (*Materials and Methods*) and a sign test (23), and show all GO terms with an FDR of less than 0.2 in both tests (in YPD, complete datasets available in Datasets S2 and S3). The same GO categories discussed above were analyzed using the binomial test with our expectation set to the total number of genes differentially expressed in one direction or the other across all of the orthologous gene pairs (usually within 5% of the expected 50:50 ratio for each condition). FDR were collected as described above. We performed the analysis separately for each of the conditions under which we grew the hybrid. In some cases, the significance of the gene set varied across conditions; in others, the gene set showed highly significant enrichment across all conditions (Datasets S2 and S3). We note that some GO terms (e.g., "filamentous growth" and "filamentous growth of a population of unicellular organisms") contain large overlapping sets of genes, and are therefore not independent categories. A complete list of all *P* values and FDRs for all GO terms in each condition is given in Datasets S2 and S3. The "classic" glycolysis GO category includes 14 enzymes of glycolysis; based on the fungal literature, we added a GO category (glycolysis_15) that includes 15 genes present in both species that code for glycolysis enzymes. When the hybrid was grown in glucose, all 14 of the classic GO gene set and all 15 of the hand-annotated set showed higher levels of mRNA produced from the *C. albicans* genome relative to the *C. dubliniensis* genome, with the same result observed across both independent experimental replicates. Based on the sign test, the likelihood of this pattern occurring by chance (i.e., in the absence of selection) is roughly $1/2^{15}$ for the 15-gene set. (As described in *Materials and Methods*, there is a small correction applied to account for a slight bias in total read counts between the genomes, resulting in a *P* value for the 15-gene set of 4×10^{-7} representing the probability that the pattern occurred by chance).

correlation between mutations in *C. albicans* that reduce filamentation and those that reduce virulence (26). The "blind" identification of filamentous growth genes by the interspecies hybrid data supports the idea that the allele-specific expression analysis is useful for uncovering phenotypic differences between the two species and, perhaps more importantly, highlighting those that arose from lineage-specific selection.

Of all the GO categories that showed consistent, highly significant directional expression (Table 1), the most cohesive and best understood set of genes are those that encode the enzymes of glycolysis, and we chose these genes for further investigation. All 15 glycolysis genes showed higher expression from the *C. albicans* genome in the hybrid across both experimental replicates grown on glucose (Fig. 2). The likelihood of this pattern occurring by chance is extremely low by both the rank-sum and the sign tests (Datasets S2 and S3), and the results strongly reject the null hypothesis of equal selection pressures acting on these genes in both lineages. When this analysis was carried out on interspecies hybrids grown on carbon sources other than glucose, the majority of the glycolysis genes were still expressed at higher levels from the *C. albicans* genome (Fig. 2). We note that the glycolytic genes were not subject to compensation; that is, they

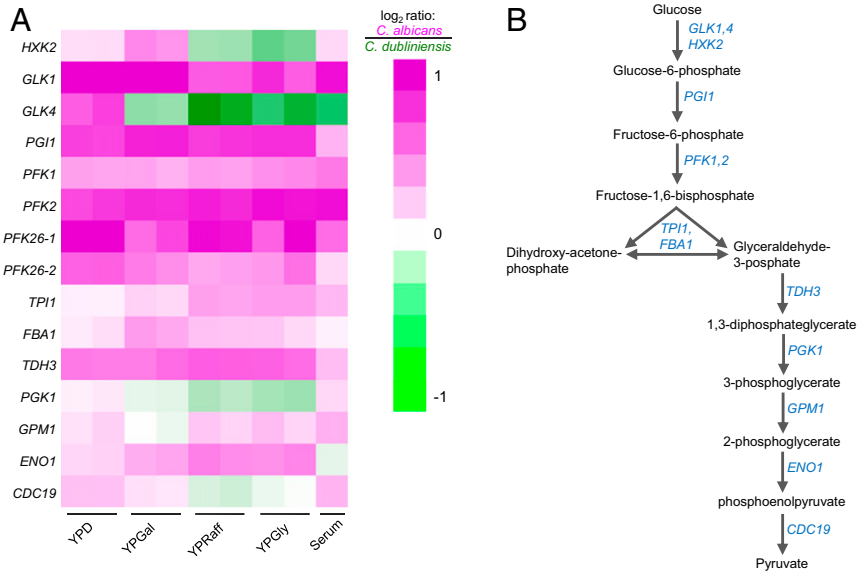

Fig. 2. mRNA levels in the interspecies hybrid identify genes in the glycolysis pathway as being under lineage-specific selection. (*A*) Heat maps of the $\log_2$ ratio of glycolysis gene reads that map to the *C. albicans* genome over those that map to the *C. dubliniensis* genome in the interspecies hybrid. These results are shown for each of the growth conditions indicated along the bottom of the figure, with independent replicate experiments shown in adjacent columns. Ratios for orthologous gene pairs across the genome, across all conditions are included in Dataset S1. (*B*) Schematic of glycolysis pathway with gene names of the corresponding enzymes (blue), and biochemical product names (black).

were also expressed at higher levels in the *C. albicans* parental strain compared with the *C. dubliniensis* parent (*SI Appendix*, Fig. S3). The higher expression of glycolysis genes in *C. albicans* compared with *C. dubliniensis* is also observed in the data reported by Palige et al. (10).

Several types of *cis*-acting changes could in principle produce higher levels of glycolysis mRNAs from the *C. albicans* genome in the hybrid, and there is no a priori reason that the same mechanism must apply to each gene. We tested one hypothesis: Namely, that differences in the binding sites for a transcriptional regulator could account for at least some of these differences. In *C. albicans,* Tye7 is the major transcriptional regulator of the glycolysis genes (27), and we found that it plays a similar role in *C. dubliniensis* (Fig. 3). We carried out a chromatin immunoprecipitation-sequencing (ChIP-seq) experiment in the interspecies hybrid using an epitope-tagged Tye7 and found that six of the glycolysis genes showed a higher signal from the *C. albicans* genome than from the *C. dubliniensis* genome (two of which are shown in Fig. 3), with three genes showing the opposite pattern. It is likely that small differences in Tye7 binding could fall below the resolution of the ChIP signal but nonetheless produce different levels of mRNA, so we cannot rigorously interpret the failure to detect a difference in the ChIP signals. We conclude that differential Tye7 binding to the glycolytic gene intergenic regions may explain some of the expression differences, but it is unlikely to be the sole explanation for the higher expression of the complete set of genes in *C. albicans*. Other possibilities include the binding of other transcription regulators and changes that affect the stability of individual mRNAs *in cis*. Irrespective of the precise cause, the differential expression of the glycolysis genes between *C. albicans* and *C dubliniensis* must be due to the accumulation of *cis*-acting changes at every gene, all acting in the same direction (higher expression in *C. albicans*) since the two species last shared a common ancestor.

We next considered the hypothesis that the increased expression of the glycolytic genes in *C. albicans* could play a role in the greater virulence of this species compared with *C. dubliniensis*. Metabolic flexibility (especially in carbon assimilation) has been implicated in *C. albicans* pathogenesis, so this idea has a historical and a logical basis (28–31). To directly assess the role of glycolysis gene expression on virulence, we sought to construct a *C. dubliniensis* strain that expressed its glycolysis genes at higher than normal levels, close to those found in *C. albicans*. We attempted this in two ways: By overexpressing the *C. albicans* Tye7 protein in *C. dubliniensis* and, in a separate construct, overexpressing the *C. albicans* Gal4, a regulator that is known to transcriptionally activate at least some of the glycolytic genes (27).

We tested for elevated expression of the glycolysis genes by RNA-seq analysis of the engineered *C. dubliniensis* strains relative to control strains. Although overexpression of Gal4 up-regulated a number of *C. dubliniensis* genes, only two of the glycolysis genes (*PFK26*-2 and *PFK1*) showed significant increases in mRNA levels (*SI Appendix*, Fig. S4). In contrast, the Tye7 overexpression strain showed significant, but modest, mRNA increases for all 15 of the glycolysis genes (Fig. 4*A*), ranging from 1.1-fold to 2.3-fold. In addition to the glycolysis genes, the engineered Tye7 strain overexpressed a number of other genes, although most of these mRNAs were present at much lower total levels than those of the glycolytic genes (Fig. 4*A*). Under anaerobic conditions (which increases the reliance on glycolysis), the engineered Tye7 *C. dubliniensis* strain showed a higher proliferation rate on glucose than did the parental *C. dubliniensis* strain, a phenotype consistent with increased expression of the glycolysis genes (*SI Appendix*, Fig. S5). We note that wild-type *C. albicans* also showed a higher proliferation rate under these same conditions compared with the *C. dubliniensis* parental strain.

We used the standard mouse model of disseminated candidiasis to test whether the engineered *C. dubliniensis* strains showed altered virulence compared to the parent strain. We infected (through tail vein injection) separate groups of eight BALB/c female mice with equal titers of *C. albicans* (strain SC5314), *C. dubliniensis* (strain CD36), the two engineered *C. dubliniensis* overexpression strains (TYE7oe, GAL4oe), or an empty vector control strain, and monitored time to illness (Fig. 3*B*). For the Gal4 overexpression strain, we observed a slight increase in time to death (i.e., slightly reduced virulence), but this did not meet statistical significance ($P = 0.21$, Mantel–Cox test; $P = 0.29$, Gehan–Breslow–Wilcoxon test). In contrast, the *C. dubliniensis* Tye7 overexpression strain showed statistically significant enhanced virulence compared with the control strain: Here, the median survival time of the mice decreased from 9 d (for the empty vector control) to 4 d for the engineered strain ($P = 0.005$). Thus, Tye7 overexpression in *C. dubliniensis* up-regulated the glycolysis genes, improved the growth rate on glucose under anaerobic laboratory conditions, and markedly increased the virulence of this strain.

To investigate whether the increased virulence was simply a consequence of improved growth, we coinjected the engineered strains (along with a control strain) and determined their relative titers in the kidney using DNA markers (*SI Appendix*, Figs. S6 and S7). In contrast to the virulence studies described above (which required a separate group of mice for each fungal strain), this experiment was carried out with a mixture of strains injected into each mouse. This strategy significantly reduces the number of mice required and also allows an accurate assessment of the competitive proliferation rates of different strains in the same animal. Of the three *C. dubliniensis* strains analyzed in this experiment, the Gal4 and Tye7 overproducer showed slightly higher titers in the kidney compared with the empty vector control strain. Because the Gal4 strain showed reduced virulence in the time-to-death experiment, these results indicate that the increased virulence of the engineered Tye7 *C. dubliniensis* strain was due to factors other than increased proliferation.

The simplest interpretation of these results is that the enhanced expression of the glycolytic genes in *C. albicans* relative to *C. dubliniensis* contributes to the former's increased virulence. The engineered strain of *C. dubliniensis* is not as virulent as the wild-type *C. albicans* strain, suggesting that additional differences between the species also contribute to the virulence differences (5, 25, 32–35). A caveat in our interpretation is that the engineered Tye7 strain, in addition to expressing the glycolysis gene at higher levels, up-regulated (and down-regulated) a number of other genes (Fig. 4), and these could play a role in the increased virulence of the engineered strain. Nonetheless, the most parsimonious explanation for all the observations in this paper is that differences in glycolytic gene expression contributes to the enhanced virulence of *C. albicans* compared with its closest relative, *C. dubliniensis*.

Discussion

Much of the work designed to uncover the virulence traits of *C. albicans* has relied on the effects of single-gene knockouts. Although this approach has revealed a great deal about *C. albicans* biology, most of the genes identified as being required for full virulence of *C. albicans* are also present in a large number of nonpathogenic species. It has therefore been difficult to understand how *C. albicans* evolved to a prevalent fungal pathogen of humans.

In this report, we investigated gene-expression differences between *C. albicans* and its closest relative, *C. dubliniensis*, a less pathogenic species. By comparing known mRNA levels produced by the two genomes in the same cell (an interspecies hybrid produced by mating), we documented many differences in the expression of orthologous genes that could be attributed to *cis*-acting differences that arose since the two species last shared a common ancestor, some 10 million y ago. We systematically examined this large dataset for nonrandom expression patterns that

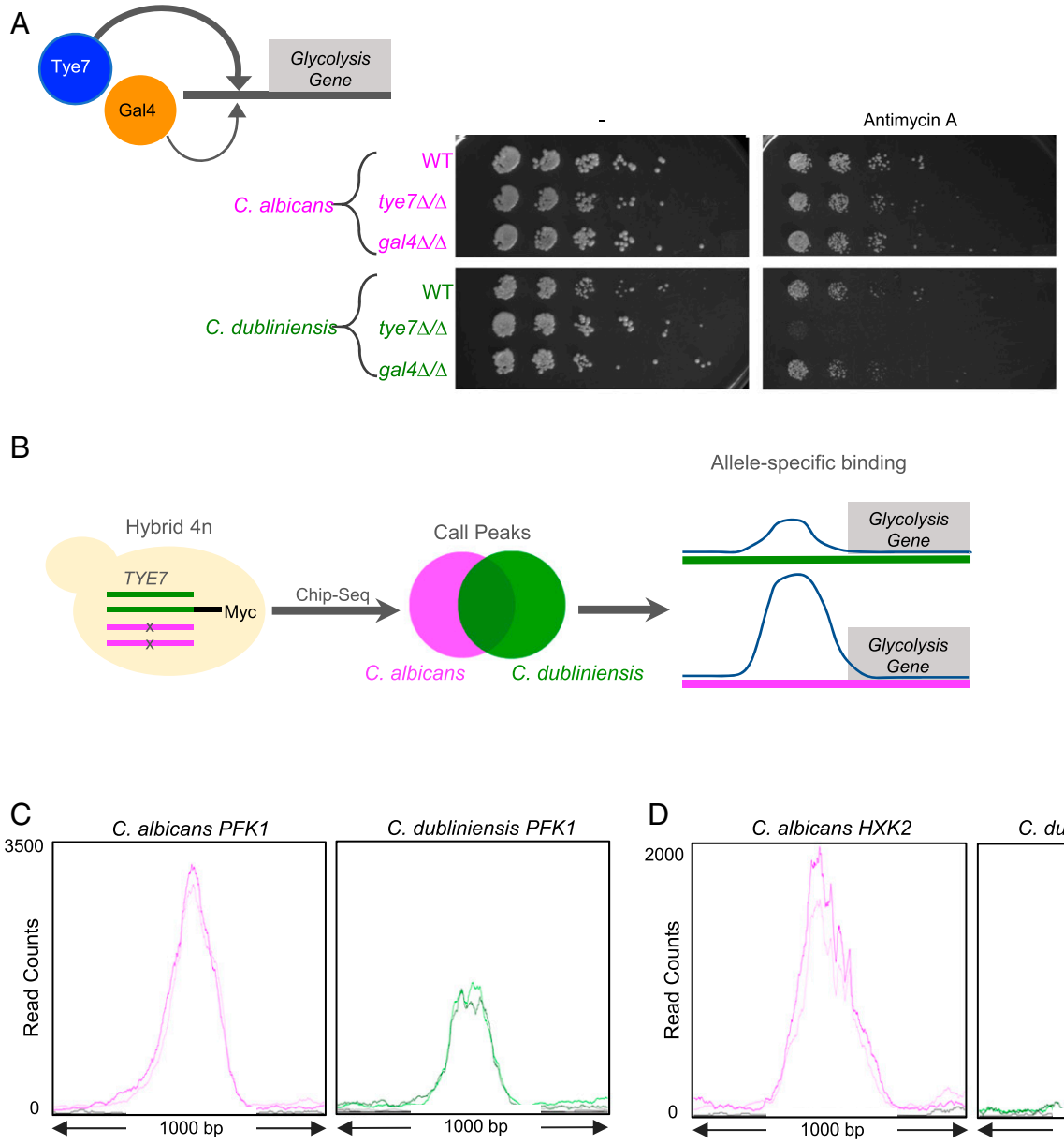

Fig. 3. The role of Tye7 in lineage-specific selection on glycolysis. (*A*) Cells were serially diluted fivefold onto solid minimal media containing 2% glucose as a carbon source with and without antimycin A to inhibit respiration and force cells to rely on glycolysis. Plates were grown at 30 °C for 72 h and pictures were taken with identical aperture and exposure settings. The results show that Tye7 is needed in *C. dubliniensis* for growth on antimycin A. (*B*) Schematic of immunoprecipitation experiments performed in the interspecies hybrid where one allele of the *C. dubliniensis* Tye7 was epitope-tagged. The ChIP-seq reads from the tagged hybrid were aligned to the concatenated genome of the two species and peaks were called in intergenic regions using MACS2 (23). Instances of allelic bias were detected using the same methods used for identifying ASE (*Materials and Methods*). (*C* and *D*) ChIP-seq of Tye7 in the interspecies hybrid shows differential binding to promoters of glycolysis genes *PFK1* and *HXK2*, respectively. The boxes depict the 1,000 base pair regions of the intergenic region upstream of the glycolysis genes in *C. albicans* (*Left*) and *C. dubliniensis* (*Right*), centered around the called ChIP peak. Darker- and lighter-colored lines represent ChIP-seq reads obtained from two replicate experiments in the interspecies hybrid. Magenta lines depict reads that map to the *C. albicans* genome, whereas green lines depict reads that map to the *C. dubliniensis* genome. Gray lines depict ChIP-Seq reads obtained from the untagged control ChIP performed in the tetraploid hybrids, with darker gray and lighter gray lines corresponding to two replicate experiments. For these genes, binding was found to be significantly enriched for the *C. albicans* alleles. Additional genes that show significant species-specific binding in the tagged hybrid are: *PGI1*, *ENO1*, *TDH3*, and *GLK1* in the direction of *C. albicans* and *GLK4, PFK26-2,* and *GPM1* in the direction of *C. dubliniensis* (negative binomial distribution, $P < 0.01$).

might have arisen through natural selection, and we were drawn to the set of genes coding for the enzymes of glycolysis, all of which showed higher mRNA levels from the *C. albicans* genome compared to the *C. dubliniensis* genome. As described in *Results*, it is extremely unlikely that this pattern arose by chance; rather, it bears all the hallmarks of natural selection due to differences in selective pressures in the two lineages. Consistent with this view, we showed that a *C. dubliniensis* strain engineered to express its glycolytic genes at slightly elevated levels was more virulent than the parent strain.

We note that our focus on the glycolytic genes came only after we systematically analyzed the gene-expression data from the interspecies hybrid and found they represented the top-scoring deviation from random expectation. Because the glycolytic genes have been extensively studied and form a well-defined set of

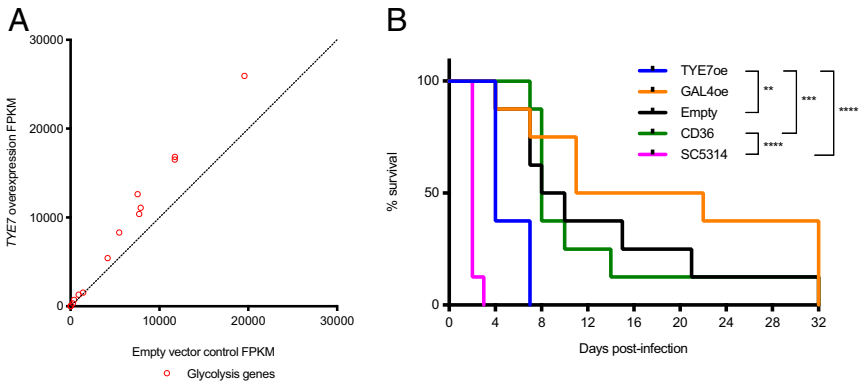

Fig. 4. Overexpression of the transcription regulator Tye7 increases glycolysis gene expression and virulence in *C. dubliniensis*. (*A*) A *C. albicans* allele of the glycolysis regulator gene *TYE7* was integrated into the *C. dubliniensis* genome, and genome-wide gene expression changes were assessed via RNA-seq (*y* axis), compared to an empty vector control construct integrated at the same position (*x* axis). All 15 *C. dubliniensis* glycolysis genes were up-regulated relative to the empty vector control strain. A total of 104 genes showed at least a twofold increase in expression and 41 genes showed at least a twofold decrease in expression in the *TYE7* overexpression strain relative to the empty vector control strain ($P < 0.01$, negative binomial distribution) (Dataset S4). The dotted line represents $x = y$, and red circles indicate glycolysis genes, and black circles all other genes. FPKM is a measure of expression level (fragments per kilobase per million mapped reads). (*B*) Overexpression of *TYE7* increases *C. dubliniensis* virulence. Groups of eight adult female BALB/c mice were infected via tail vein injection with 5×10^5 cells of the *C. albicans* wild-type strain (SC5314, magenta), the *C. dubliniensis* wild-type strain (CD36, green), the *C. dubliniensis* strain harboring an overexpressed allele of *C. albicans TYE7* (TYE7oe, blue), *GAL4* (GAL4oe, orange), or the empty vector control strain (Empty, black), and time-to-illness was monitored. Significant differences were observed between mice infected with TYE7oe versus Empty ($P = 0.0054$), between TYE7oe and CD36 ($P = 0.0004$), and between CD36 and SC5314 ($P < 0.0001$). Statistically significant differences between survival curves were determined using the log-rank test (**$P \leq 0.01$, ***$P \leq 0.001$, ****$P \leq 0.0001$).

cofunctioning genes, they readily emerged from the analysis. However, other sets of well-documented cofunctioning metabolic genes did not. We predict that, as studies continue in *C. albicans*, other sets of cofunctioning genes may be revealed from our dataset as having undergone lineage-specific selection. Our finding that all of the glycolysis genes were up-regulated from the *C. albicans* genome relative to *C. dubliniensis* is consistent with conclusions from metabolic control analysis, which stresses that increased flux through a pathway is unlikely to occur by up-regulating single "rate-limiting" enzymes [see, for example, Fell (36)]. This idea may explain why *cis* changes accumulated at all glycolysis genes rather than a few.

Several explanations are possible for the systematic difference in glycolytic gene expression that arose between *C. albicans* and *C. dubliniensis* over an evolutionary time scale: 1) The higher levels of expression were adaptive in the *C. albicans* lineage; 2) the ancestor of *C. albicans* and *C. dubliniensis* expressed the genes at the higher levels, and selection was maintained in the *C. albicans* lineage but relaxed in the *C. dubliniensis* lineage; 3) reduced expression of the glycolytic genes was somehow adaptive in the *C. dubliniensis* lineage. We favor the first two explanations, as it is more difficult to devise a coherent scenario for the third. In considering the virulence difference between the two species, it is not known whether *C. albicans* became more virulent after the *C. albicans*–*C. dubliniensis* split or *C. dubliniensis* became less virulent (or both). Regardless of the historical events, both the virulence properties and glycolytic gene-expression differences are clear distinctions between the two species that must have arisen from the different selective pressures acting on the two lineages. A possible contributor to this change is the recently documented hybridization event between *C. albicans* ancestors (either between different strains of the same species or between different species) that occurred after the branch with *C. dubliniensis* (37).

Our experiments with the engineered *C. dubliniensis* strain suggest a causal connection between increased glycolysis gene expression and virulence, but does this relationship make any conceptual sense? There is substantial literature on possible links (27–29, 31, 38–48). The concentration of glucose in the blood is relatively low (3 to 5 mM) and it drops even further during systemic infection by *C. albicans* (38, 42, 48). Deletion or conditional repression of individual glycolytic genes in *C. albicans* results in defects in both laboratory growth and in virulence studies (40, 46), as do *C. albicans* strains deleted for both of the transcriptional activators of the glycolysis genes Gal4 and Tye7 (27). And, the virulence of *C. albicans* depends on the carbon source on which the strain is grown before introducing it into the mouse (38). Finally, *C. albicans* and macrophages compete for limiting glucose and *C. albicans* can thereby cause the death of macrophages (31). All of these experiments suggest a link between glycolysis and virulence, one that is more complex than simple maintenance of a housekeeping function. Other experiments implicate additional *C. albicans* carbon metabolic pathways (such as gluconeogenesis) in pathogenesis, and it has been suggested that metabolic flexibility—specifically the ability of *C. albicans* to efficiently use multiple carbon sources at the same time—is one of the features that contributes to *C. albicans* pathogenesis (reviewed in refs. 29 and 30).

In trying to understand why *C. albicans* causes disease in humans while the vast majority of fungal species are avirulent, it has been tempting to seek out specific "virulence factors" present in *C. albicans* but missing in other species. The work presented in this report indicates that important contributions to the virulence differences across fungal species may arise through very subtle evolutionary changes that affect the expression of genes that are deeply conserved in both virulent and avirulent species.

Materials and Methods

Strain Construction. All strains and plasmids used in this study can be found in *SI Appendix*, Tables S1 and S2, respectively. All *C. albicans* strains were derived from the prototrophic wild-type strain SC5314 (49) and all *C. dubliniensis* strains were derived from the prototrophic wild-type strain CD36 (4). Mutant gene-deletion strains were constructed following published procedures (26) and overexpression strains driven by the *TDH3* promoter were constructed as described in Hernday et al. (50). To construct the tetraploid interspecies hybrid, mutant strains of both *C. albicans* and *C. dubliniensis* were first made auxotrophic for histidine and leucine, both markers that are not required for *C. albicans* virulence in the murine systemic infection model (51). Additional genetic manipulations were used to make stains homozygous at the *MTL* locus, and the resulting a- and α-strains with complementary mating types were grown together on solid rich media to allow for

mating. Mating products were subsequently selected for on the basis of auxotrophic complementation on minimal media and tetraploidy was confirmed by FACS and RNA-seq, by aligning reads across the concatenated genome to assess coverage.

Genome Sequencing. *C. albicans* SC5314 and *C. dubliniensis* CD36 were grown to midlog phase in YPD at 30 °C, pelleted, washed, and genomic DNA was extracted. The Illumina TruSeq kit was used to construct sequencing libraries and 150 base pair paired-end reads were sequenced using an Illumina MiSeq.

RNA-Seq. The interspecies tetraploid hybrid of *C. albicans* and *C. dubliniensis,* CadSB112, was grown to midlog phase in the following laboratory conditions: YP + 2% glucose, YP + 2% galactose, YP + 2% raffinose, YP + 2% glycerol, all at 30 °C, and bovine plasma + 2% glucose at 37 °C to mimic host-like conditions. Each experiment was performed in duplicate except for the bovine plasma condition. The parental diploid strains of each species (CdSB2 and CaSB14) were grown to midlog phase in bovine plasma + 2% glucose at 37 °C. For *C. dubliniensis* overexpression experiments, TYE7oe (CdSB163), GAL4oe (CdSB164), and Empty (CdSB169), were grown in YPD at 30 °C to midlog phase. All cultures were pelleted, washed, and flash frozen in liquid nitrogen. RNA was extracted using the RiboPure RNA Purification kit (Ambion) and RNA samples were submitted to the JP Sulzberger Columbia Genome Center for library preparation and sequencing. Libraries were constructed using the Illumina TruSeq RNA library prep kit v2 and 100-base pair single-end reads were sequenced using the Illumina HiSEq. 2500.

Gene-Expression Analysis. Parental DNA-sequencing, parental RNA-seq, and hybrid RNA-seq reads were aligned against a concatenated genome of *C. albicans* SC5314 (Candida Genome Database assembly 21-s02-m05-r02) and *C. dubliniensis* CD36 (52) using SOAP2 with custom flags –e 5 –l 35 –v 5 –g 1 –s 255 (53). Fewer than 5% of hybrid reads had exactly one match in both species in any sample and fewer than 1% of parental reads cross-mapped to the other genome. All multimapped reads were removed from the analysis. Read counts were compiled for all ORFs. To account for potential alignability or sequence length differences between orthologous sets of ORFs (54), read counts were normalized using a scaling factor. The scaling factor was determined as the ratio of fragment counts-per-million of parental genomic DNA reads aligned to the same set of orthologous ORFs. Normalization was done by dividing the RNA-seq counts of the more highly represented ortholog (in DNA reads) by the scaling factor. Tiling both genomes in 100-base pair windows every 1 base pair and realigning yielded near-identical (Pearson $r > 0.99$) scaling factors. Significance of allele-specific expression (ASE) was quantified in hybrid samples using a binomial test, setting the expectation to the total fraction of *C. albicans:C. dubliniensis* aligned reads, and corrected for multiple tests (55) (Dataset S1). Significant differential expression in *C. dubliniensis* overexpression experiments was estimated with a negative binomial test as previously described (56).

***Cis*- and *trans* Regulation Analysis.** Normalized gene-level expression counts were assessed for differential expression using described methods (57). In brief, parental RNA-seq data were normalized both to the total number of aligned reads and the ortholog alignability scaling factor (see *Gene-Expression Analysis* section of *Materials and Methods*). ASE was quantified using the cumulative binomial distribution, as described above, and *trans* effects were further assessed using χ^2 tests on parental versus hybrid RNA-seq counts. *P*-values were adjusted for multiple tests (55) and significance was set at 0.5%.

Quantifying ASE of individual genes was done using two commonly applied approaches: 1) A binomial test with null expectation set to reflect overall expression from each species (58), and 2) a negative binomial test (59). It was observed, however, that both approaches yielded highly significant *P* values with nearly all expressed genes showing ASE ($P < 0.01$). However, both of the statistical approaches assume that there are no technical biases between the two alleles. edgeR and DESeq (60) were developed to sample the same gene across conditions where this assumption is likely true. In an analysis of previous work in mouse inbred hybrids (58) and hybrid yeast (61), we observed that sampling ASE at two polymorphic sites in the same mouse exon, or two polymorphic sites in the same yeast ORF, had an unexpectedly high degree of disagreement. Priming and amplification differences during library construction due to allele-specific sequence differences may be the cause. We reasoned that, as long as the two polymorphic sites were independently sampled (i.e., by separate RNA-seq fragments), the frequency with which they agree on the direction of ASE does in fact reflect presence/absence of ASE. Taking all sites separated by a distance that exceeds the read length (ensuring they are independently sampled), we observed that disagreement is a function of sequencing depth and the minimum level of measured ASE at each site. Disagreement was much more frequent than predicted using the binomial test, with similar patterns observed mouse and yeast. We formalized these observations into a mathematical model that we make available through user-friendly perl and matlab functions (ref: https://github.com/tomasbabak/trueASE). Applying this more conservative framework, we observed ~40% of genes with significant ASE.

Lineage-specific selection on pathways was assessed with two statistical approaches. To uncover biological pathways whose gene-expression patterns from the two genomes showed large departures from random expectations, we ranked each orthologous gene pair from the *C. albicans* genome compared to the *C. dubliniensis* genome in the hybrid (5,267 annotated ORFs), by the *P* value of differential gene expression between those orthologous genes (negative binomial distribution, with the expectation set to the total number of reads from the *C. albicans* genome versus *C. dubliniensis*, where genes with increased expression in *C. dubliniensis* are set to a negative value). To capture both the significance of differential gene expression, as well as direction, we ranked all GO terms (52) (with five or more genes) with higher expression from the *C. albicans* genome at the top and higher expression from the *C. dubliniensis* genome at the bottom. Rankings were based on the rank-sum test, and false-discovery rates (FDR) were computed using the Benjamini–Hochberg procedure to correct for multiple comparisons (62). Using this approach, which takes into account both the direction of the allelic bias as well as the magnitude, the glycolysis genes ranked at the top of our significance list. As a separate, independent test, a sign test was applied as in Artieri and Fraser (61). The above categories were also tested for directional bias using a binomial test, with expectation set to the total number of genes up-regulated in one species versus the other. FDR were estimated using the Benjamini–Hochberg step-up procedure (62).

ChIP. The epitope-tagged tetraploid interspecies hybrid strain (CdSB144 × CaSB133 where one copy of the *C. dubliniensis* Tye7 was tagged with Myc) and the untagged control strain lacking the C-terminal Myc tag (CdSB69 × CaSB133) were grown to midlog phase in YPD at 30 °C. Cells were cross-linked in 1% formaldehyde for 15 min, quenched with 125 mM glycine, pelleted, washed, and flash-frozen in liquid nitrogen. Cells were lysed in 50 mM Hepes/KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate, and an EDTA-free Complete Protease Inhibitor Mixture (Roche), as described in Sorrells et al. (63). ChIP was performed as described in Hernday et al. (50) with an anti–c-Myc monoclonal antibody (Invitrogen AHO0062), libraries were prepared with the NEBNext Multiplex kit for Illumina (New England Biolabs) and AMPure XP magnetic beads (Beckman Coulter), as described in Sorrells et al. (63), and sequenced on an Illumina HiSEq. 4000. All experiments were performed in duplicate from independently grown single colonies of the same strain.

Allele-Specific Binding Analysis. To enable cross-species comparison of peak position and accurate quantification of Tye7 ChIP-seq signal, liftOver (64) files were created on reciprocally-generated whole-genome pairwise alignments of *C. albicans* and *C. dubliniensis*. Lastz (64) was used to generate the alignments and maf-convert (http://last.cbrc.jp/doc/maf-convert.html) and University of California, Santa Cruz tools were used to generate liftOvers as described previously (http://genomewiki.ucsc.edu/index.php/LiftOver_Howto). ChIP-seq reads were aligned using the procedure described above in the section discussing gene-expression analysis and filtered for cross-species mappability (i.e., could be lifted over using liftOver). Peaks were called using MACS2 (65) with default settings and *C. dubliniensis* peaks were lifted over to the *C. albicans* genome coordinates. Consolidated peaks were then generated as the union of the four sets of peaks (two replicates and two species) and filtered reads were counted in these regions. Peaks were linked to genes based on proximity to closest translation start sites and allele-specific binding was assessed using the same procedures as used to quantify ASE (see discussion of ASE above).

Growth Curves and Quantification. All strains were grown overnight in standard YP + 2% glucose at 30 °C. The following day, strains were washed in PBS and diluted in either YNB + amino acids + 2% glucose or YNB + amino acids + 2% glycerol. To force glucose metabolism through glycolysis, strains were grown at 30 °C in an anaerobic chamber (Coy Laboratory Products), whereas strains grown with glycerol as the sole carbon source were grown at 30 °C aerobically. All experiments were performed in 96-well format on a Tecan Infinite M1000 (in aerobic conditions) or a Tecan Infinite M Nano (in anaerobic

conditions). Data were analyzed using the R package Growthcurver to obtain doubling times (66).

In Vivo Virulence Assay. All procedures involving animals were approved by the University of California, San Francisco Institutional Animal Care and Use Committee and performed in accordance with relevant guidelines and regulations. Strains tested were wild-type prototrophic *C. albicans* SC5314 (CaSB14), wild-type prototrophic *C. dubliniensis* CD36 (CdSB2), *C. dubliniensis* Tye7 overexpression strain TYE7oe (CdSB136), *C. dubliniensis* Gal4 overexpression strain GAL4oe (CdSB164), and *C. dubliniensis* empty vector control strain Empty (CdSB169). Strains were inoculated from overnight cultures into fresh YP + 2% glucose and grown to midlog at 30 °C. Cells were pelleted, washed, and resuspended in sterile saline and cell counts were obtained with a hemocytometer. Eight female 6- to 8-wk-old (18 to 20 g in weight), BALB/c mice were infected via tail vein injection with 5×10^5 cells (in a 0.1-mL volume) of one of the strains tested. Mice were monitored twice daily and killed upon signs of illness (minimal motor activity, weight loss of > 15% of initial body weight, and hunched posture). The log-rank test was used to determine whether survival curves were statistically significantly different.

In Vivo Proliferation Assay. In vivo proliferation protocols were adapted from Noble et al. (26, 51). For *C. albicans* versus *C. dubliniensis* proliferation experiments, *C. albicans* (CaSB16) and *C. dubliniensis* (CdSB4) were inoculated into fresh YP + 2% glucose and grown to midlog phase at 30 °C. Cells were prepared as above, except cells were resuspended in sterile saline in an equal mixture. Eight female 6- to 8-wk-old (18 to 20 g in weight), BALB/c mice were infected via tail vein injection with 5×10^5 cells (in a 0.1-mL volume) of the mixture. Mice were monitored and killed as above. Upon killing, kidneys were dissected, homogenized, diluted, and plated on Sabauraud agar with ampicillin (50 μg/mL) and gentamicin (15 μg/mL) for 48 h at 30 °C and replica plated to media either lacking the amino acid leucine (to select for *C. dubliniensis*) or lacking histidine (to select for *C. albicans*). Statistical differences in fungal burden between *C. albicans* and *C. dubliniensis* across all mice was determined using the Mann–Whitney *U* test.

For the overexpression experiments, *C. dubliniensis* strain TYE7oe (CdSB136), *C. dubliniensis* strain GAL4oe (CdSB164), and *C. dubliniensis* strain Empty (CdSB169) were inoculated from overnight cultures into fresh YP + 2% glucose and grown to midlog phase at 30 °C. Cells were pelleted, washed, and resuspended in sterile saline in an equal mixture and eight female 6- to 8-wk-old (18 to 20 g in weight), BALB/c mice were infected via tail vein injection with 10^6 cells of the mixture. Mice were monitored and killed as above. Upon killing, kidneys were dissected, homogenized, diluted, and plated on Sabouraud agar with ampicillin (50 μg/mL) and gentamicin (15 μg/mL). Recovered CFU were grown for 48 h at 30 °C and genomic DNA was recovered from CFU scraped off the plates. Individual strain abundance was determined using real-time PCR and primers specific to each strain on a StepOnePlus RT-PCR machine (Applied Biosystems), relative to the inoculum, and differences in abundance were assessed for statistical significance using the Wilcoxon signed-ranked test. The TYE7oe strain was quantified using primer pair oSB184/201, the GAL4oe strain quantified using primer pair oSB184/273, and the Empty vector strain using primer pair oSB184/275.

Data Availability. Code generated for mathematical modeling of ASE quantification data have been deposited in Github (https://github.com/tomasbabak/trueASE). All analyzed study data are included in the article and supporting information. Raw data is available on Gene Expression Omnibus at accession number GSE168112.

ACKNOWLEDGMENTS. We thank L. Noiman and M. Lohse for comments on the manuscript; Byron C. Hann and the University of California, San Francisco Preclinical Therapeutics Core for assistance with animal experiments; A.D.J. laboratory members for engaging and helpful discussion; and perceptive and useful comments from two anonymous reviewers.

1. Y. L. Chen *et al.*, Calcineurin controls drug tolerance, hyphal growth, and virulence in *Candida dubliniensis*. *Eukaryot. Cell* **10**, 803–819 (2011).
2. G. D. Gilfillan *et al.*, *Candida dubliniensis*: Phylogeny and putative virulence factors. *Microbiology (Reading)* **144**, 829–838 (1998).
3. D. J. Sullivan *et al.*, Comparison of the epidemiology, drug resistance mechanisms, and virulence of *Candida dubliniensis* and *Candida albicans*. *FEMS Yeast Res.* **4**, 369–376 (2004).
4. D. J. Sullivan, T. J. Westerneng, K. A. Haynes, D. E. Bennett, D. C. Coleman, *Candida dubliniensis* sp. nov.: Phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology (Reading)* **141**, 1507–1521 (1995).
5. M. M. Vilela *et al.*, Pathogenicity and virulence of *Candida dubliniensis*: Comparison with *C. albicans*. *Med. Mycol.* **40**, 249–257 (2002).
6. G. Moran *et al.*, Comparative genomics using *Candida albicans* DNA microarrays reveals absence and divergence of virulence-associated genes in *Candida dubliniensis*. *Microbiology (Reading)* **150**, 3363–3382 (2004).
7. D. J. Sullivan, G. P. Moran, D. C. Coleman, *Candida dubliniensis*: Ten years on. *FEMS Microbiol. Lett.* **253**, 9–17 (2005).
8. A. P. Jackson *et al.*, Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Res.* **19**, 2231–2244 (2009).
9. C. Grumaz *et al.*, Species and condition specific adaptation of the transcriptional landscapes in *Candida albicans* and *Candida dubliniensis*. *BMC Genomics* **14**, 212 (2013).
10. K. Palige *et al.*, Global transcriptome sequencing identifies chlamydospore specific markers in *Candida albicans* and *Candida dubliniensis*. *PLoS One* **8**, e61940 (2013).
11. J. H. Bullard, Y. Mostovoy, S. Dudoit, R. B. Brem, Polygenic and directional regulatory evolution across pathways in *Saccharomyces*. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5058–5063 (2010).
12. J. Chang *et al.*, The molecular mechanism of a *cis*-regulatory adaptation in yeast. *PLoS Genet.* **9**, e1003813 (2013).
13. H. B. Fraser *et al.*, Polygenic *cis*-regulatory adaptation in the evolution of yeast pathogenicity. *Genome Res.* **22**, 1930–1939 (2012).
14. H. C. Martin, J. I. Roop, J. G. Schraiber, T. Y. Hsu, R. B. Brem, Evolution of a membrane protein regulon in *Saccharomyces*. *Mol. Biol. Evol.* **29**, 1747–1756 (2012).
15. J. I. Roop, K. C. Chang, R. B. Brem, Polygenic evolution of a sugar specialization tradeoff in yeast. *Nature* **530**, 336–339 (2016).
16. C. Pujol *et al.*, The closely related species *Candida albicans* and *Candida dubliniensis* can mate. *Eukaryot. Cell* **3**, 1015–1027 (2004).
17. S. A. Signor, S. V. Nuzhdin, The evolution of gene expression in *cis* and *trans*. *Trends Genet.* **34**, 532–544 (2018).
18. I. Tirosh, S. Reikhav, A. A. Levy, N. Barkai, A yeast hybrid provides insight into the evolution of gene expression regulation. *Science* **324**, 659–662 (2009).
19. X. C. Li, J. C. Fay, *Cis*-regulatory divergence in gene expression between two thermally divergent yeast species. *Genome Biol. Evol.* **9**, 1120–1129 (2017).
20. P. J. Wittkopp, B. K. Haerum, A. G. Clark, Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**, 85–88 (2004).
21. Z. Wang *et al.*, Evolution of gene regulation during transcription and translation. *Genome Biol. Evol.* **7**, 1155–1167 (2015).
22. B. P. H. Metzger, P. J. Wittkopp, J. D. Coolon, Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol. Evol.* **9**, 843–854 (2017).
23. H. B. Fraser, Genome-wide approaches to the study of adaptive gene expression evolution: Systematic studies of evolutionary adaptations involving gene expression will allow many fundamental questions in evolutionary biology to be addressed. *BioEssays* **33**, 469–477 (2011).
24. G. P. Moran, D. C. Coleman, D. J. Sullivan, *Candida albicans* versus *Candida dubliniensis*: Why is *C. albicans* more pathogenic? *Int. J. Microbiol.* **2012**, 205921 (2012).
25. G. P. Moran, D. M. MacCallum, M. J. Spiering, D. C. Coleman, D. J. Sullivan, Differential regulation of the transcriptional repressor *NRG1* accounts for altered host-cell interactions in *Candida albicans* and *Candida dubliniensis*. *Mol. Microbiol.* **66**, 915–929 (2007).
26. S. M. Noble, S. French, L. A. Kohn, V. Chen, A. D. Johnson, Systematic screens of a *Candida albicans* homozygous deletion library decouple morphogenetic switching and pathogenicity. *Nat. Genet.* **42**, 590–598 (2010).
27. C. Askew *et al.*, Transcriptional regulation of carbohydrate metabolism in the human pathogen *Candida albicans*. *PLoS Pathog.* **5**, e1000612 (2009).
28. M. C. Lorenz, J. A. Bender, G. R. Fink, Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryot. Cell* **3**, 1076–1087 (2004).
29. A. J. Brown, G. D. Brown, M. G. Netea, N. A. Gow, Metabolism impacts upon *Candida* immunogenicity and pathogenicity at multiple levels. *Trends Microbiol.* **22**, 614–622 (2014).
30. P. Miramón, M. C. Lorenz, A feast for *Candida*: Metabolic plasticity confers an edge for virulence. *PLoS Pathog.* **13**, e1006144 (2017).
31. T. M. Tucey *et al.*, Glucose homeostasis is important for immune cell viability during *Candida* challenge and host survival of systemic fungal infection. *Cell Metab.* **27**, 988–1006.e7 (2018).
32. M. J. Spiering *et al.*, Comparative transcript profiling of *Candida albicans* and *Candida dubliniensis* identifies SFL2, a *C. albicans* gene required for virulence in a reconstituted epithelial infection model. *Eukaryot. Cell* **9**, 251–265 (2010).
33. Z. Liu, G. P. Moran, D. J. Sullivan, D. M. MacCallum, L. C. Myers, Amplification of TLO mediator subunit genes facilitate filamentous growth in *Candida* spp. *PLoS Genet.* **12**, e1006373 (2016).
34. R. P. Jordan, D. W. Williams, G. P. Moran, D. C. Coleman, D. J. Sullivan, Comparative adherence of *Candida albicans* and *Candida dubliniensis* to human buccal epithelial cells and extracellular matrix proteins. *Med. Mycol.* **52**, 254–263 (2014).
35. P. R. Flanagan *et al.*, Expansion of the TLO gene family enhances the virulence of *Candida* species. *PLoS One* **13**, e0200852 (2018).
36. D. A. Fell, Metabolic control analysis: A survey of its theoretical and experimental development. *Biochem. J.* **286**, 313–330 (1992).
37. V. Mixão, T. Gabaldón, Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*. *BMC Biol.* **18**, 48 (2020).

38. I. V. Ene *et al*., Host carbon sources modulate cell wall architecture, drug resistance and virulence in a fungal pathogen. *Cell. Microbiol.* **14**, 1319–1335 (2012).
39. R. K. Swoboda *et al*., Fluctuations in glycolytic mRNA levels during morphogenesis in *Candida albicans* reflect underlying changes in growth and are not a response to cellular dimorphism. *Mol. Microbiol.* **13**, 663–672 (1994).
40. A. Rodaki, T. Young, A. J. Brown, Effects of depleting the essential central metabolic enzyme fructose-1,6-bisphosphate aldolase on the growth and viability of *Candida albicans*: Implications for antifungal drug target discovery. *Eukaryot. Cell* **5**, 1371–1377 (2006).
41. A. Rodaki *et al*., Glucose promotes stress resistance in the fungal pathogen *Candida albicans*. *Mol. Biol. Cell* **20**, 4845–4855 (2009).
42. M. Calahorra, N. S. Sánchez, A. Peña, Characterization of glycolytic metabolism and ion transport of *Candida albicans*. *Yeast* **29**, 357–370 (2012).
43. J. Bonhomme *et al*., Contribution of the glycolytic flux and hypoxia adaptation to efficient biofilm formation by *Candida albicans*. *Mol. Microbiol.* **80**, 995–1013 (2011).
44. D. Wilson *et al*., Identifying infection-associated genes of *Candida albicans* in the postgenomic era. *FEMS Yeast Res.* **9**, 688–700 (2009).
45. I. V. Ene, S. C. Cheng, M. G. Netea, A. J. Brown, Growth of *Candida albicans* cells on the physiologically relevant carbon source lactate affects their recognition and phagocytosis by immune cells. *Infect. Immun.* **81**, 238–248 (2013).
46. R. Laurian *et al*., Hexokinase and glucokinases are essential for fitness and virulence in the pathogenic yeast *Candida albicans*. *Front. Microbiol.* **10**, 327 (2019).
47. D. Sandai *et al*., The evolutionary rewiring of ubiquitination targets has reprogrammed the regulation of carbon assimilation in the pathogenic yeast *Candida albicans*. *MBio* **3**, e00495-12 (2012).
48. B. Spellberg, A. S. Ibrahim, J. E. Edwards Jr, S. G. Filler, Mice with disseminated candidiasis die of progressive sepsis. *J. Infect. Dis.* **192**, 336–343 (2005).
49. A. M. Gillum, E. Y. Tsay, D. R. Kirsch, Isolation of the *Candida albicans* gene for orotidine-5′-phosphate decarboxylase by complementation of S. cerevisiae *ura3* and E. coli pyrF mutations. *Mol. Gen. Genet.* **198**, 179–182 (1984).
50. A. D. Hernday, S. M. Noble, Q. M. Mitrovich, A. D. Johnson, Genetics and molecular biology in *Candida albicans*. *Methods Enzymol.* **470**, 737–758 (2010).
51. S. M. Noble, A. D. Johnson, Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. *Eukaryot. Cell* **4**, 298–309 (2005).
52. M. S. Skrzypek *et al*., The Candida Genome Database (CGD): Incorporation of assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res.* **45**, D592–D596 (2017).
53. R. Li *et al*., SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* **25**, 1966–1967 (2009).
54. D. A. Fitzpatrick, P. O'Gaora, K. P. Byrne, G. Butler, Analysis of gene evolution and metabolic pathways using the Candida Gene Order Browser. *BMC Genomics* **11**, 290 (2010).
55. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9440–9445 (2003).
56. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
57. C. J. McManus *et al*., Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res.* **20**, 816–825 (2010).
58. T. Babak *et al*., Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat. Genet.* **47**, 544–549 (2015).
59. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
60. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
61. C. G. Artieri, H. B. Fraser, Evolution at two levels of gene expression in yeast. *Genome Res.* **24**, 411–421 (2014).
62. Y. Klipper-Aurbach *et al*., Mathematical formulae for the prediction of the residual beta cell function during the first two years of disease in children and adolescents with insulin-dependent diabetes mellitus. *Med. Hypotheses* **45**, 486–490 (1995).
63. T. R. Sorrells, L. N. Booth, B. B. Tuch, A. D. Johnson, Intersecting transcription networks constrain gene regulatory evolution. *Nature* **523**, 361–365 (2015).
64. W. J. Kent *et al*., The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
65. Y. Zhang *et al*., Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
66. K. Sprouffske, A. Wagner, Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* **17**, 172 (2016).