PNAS
www.pnas.org

Supplementary Information for
Lineage-specific selection and the evolution of virulence in the *Candida* clade.

Sheena D. Singh-Babak, Tomas Babak, Hunter B. Fraser, and Alexander D. Johnson

Alexander D.Johnson
Email: ajohnson@cgl.ucsf.edu

This PDF file includes:

Supplementary text
Figures S1 to Sx
Tables S1 to S3
Legends for Datasets S1 to S4
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S4

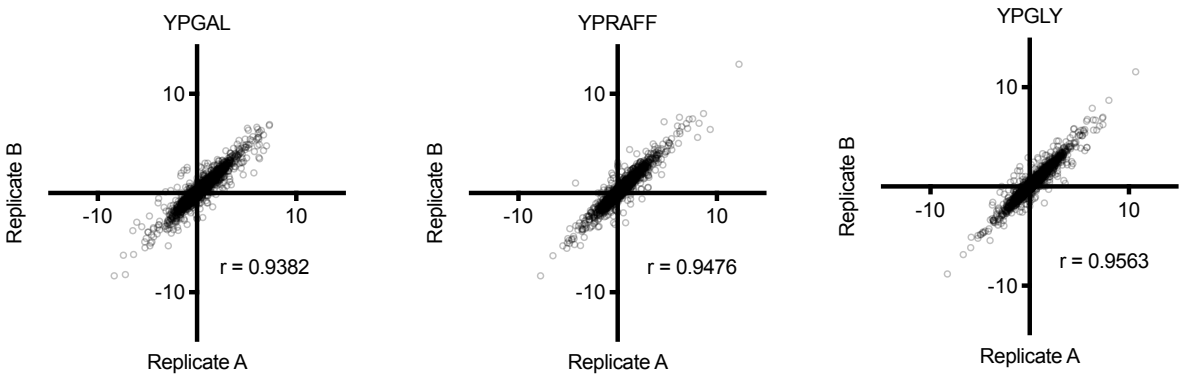

Fig. S1. Correlation of RNA-Seq results between replicates for the interspecies hybrid. Replicate correlations for RNA-Seq experiments of the interspecies hybrid grown at 30˚C in standard laboratory media containing yeast extract, peptone, and either 2% galactose, 2% raffinose, or 2% glycerol, respectively. Each point represents the $\log_2$ ratio of *C. albicans* normalized RNA-Seq read counts over *C. dubliniensis* normalized read counts for each 1:1 orthologous gene pair between *C. albicans* and *C. dubliniensis*. The Spearman correlation coefficient (r) is listed in the bottom right quadrant of each plot.

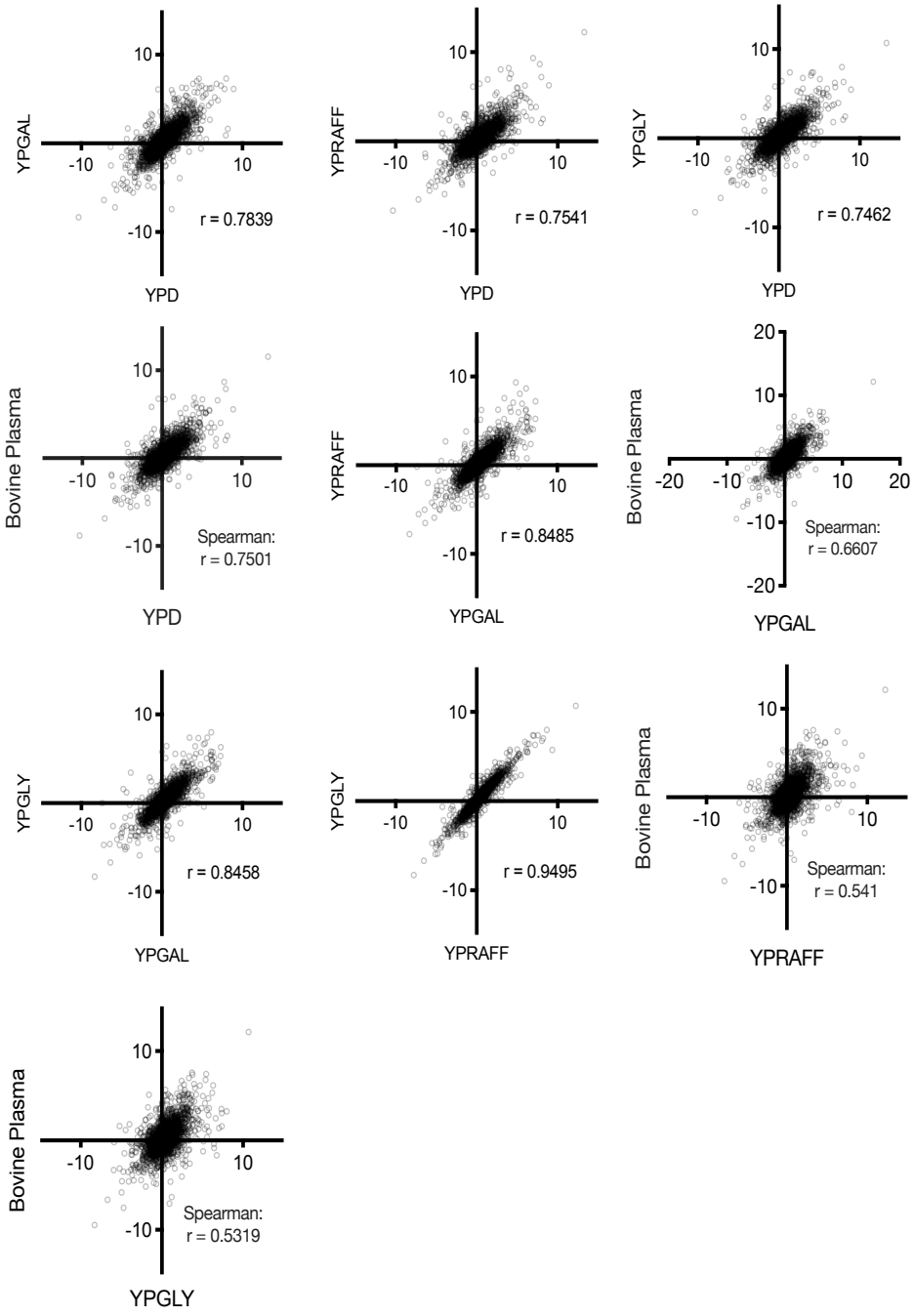

Fig. S2. Comparison of RNA-Seq results in the interspecies hybrid across different experimental conditions. Experimental condition correlations for RNA-Seq experiments of the interspecies hybrid grown in the following conditions: at 30˚C in standard laboratory medium containing 2% glucose, 2% galactose, 2% raffinose, or 2% glycerol, or in bovine plasma at 37˚C to mimic a more host-like environment. Each point represents the log2 ratio of C. albicans normalized RNA-Seq read counts over C. dubliniensis normalized read counts for each 1:1 orthologous gene pair between C. albicans and C. dubliniensis. Spearman correlation coefficient (r) is listed in the bottom right quadrant of each plot.

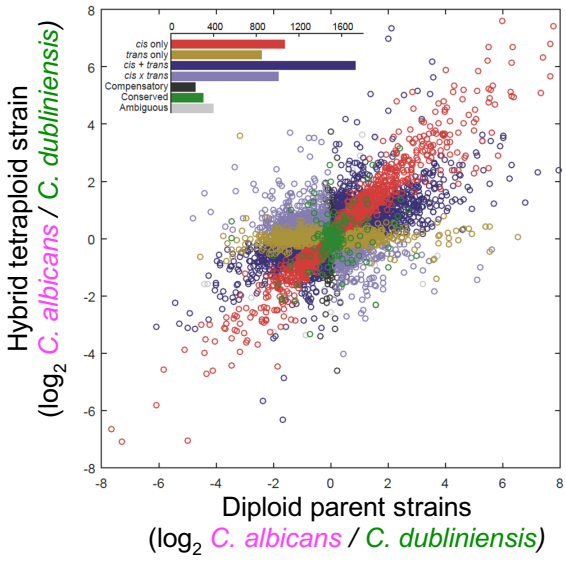

Fig. S3. Relative allele-specific expression levels in parental C. albicans and C. dubliniensis strains versus the interspecies hybrid. Each point represents log2 gene expression levels of C. albicans/C. dubliniensis for a single 1:1 orthologous gene pair. Expression ratios of gene pair for diploid parent strains of the two species are plotted on the x-axis, and the expression ratios from each species' genome in the interspecific tetraploid hybrid strain are plotted on the y-axis. As indicated in the inset, the colored dots correspond to the mechanism of regulatory evolution inferred from a hierarchical series of statistical tests as described by (*1*). Cis only refers to quantitative differences in the interspecies hybrid that were recapitulated in the parental strains. Trans only refers to differences between the parental species that were not observed in the interspecies hybrid. Cis + trans refers to genes that show a difference in the interspecies hybrid but a greater difference (in the same direction) in the parent species; cis x trans refers to genes that show a greater difference in the hybrid than the parent. Compensatory refers to genes that show differential expression in the hybrid but no difference in the parent species. Conserved refers to no differential expression in either the hybrid or the parent. The bar graph in the upper left portrays the number of genes in each category. For this analysis, all strains were grown in bovine plasma at 37˚C. Note that the frequency of compensatory cis/trans changes may be overestimated due to the effect described in (*2*), which we were unable to control for due to having only a single replicate of ASE measured in bovine plasma.

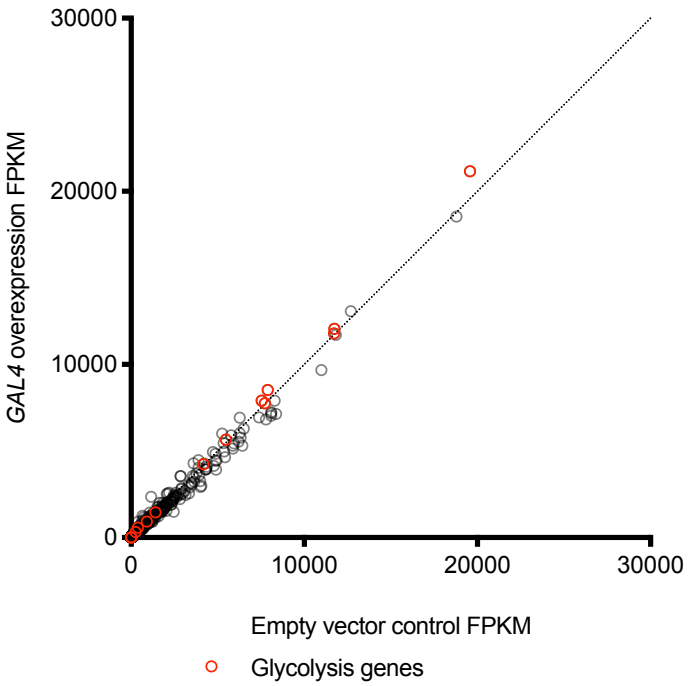

Fig. S4. Overexpression of the glycolysis regulator *GAL4* increases expression of several but not all glycolysis genes. A *C. albicans* allele of the glycolysis regulator *GAL4* was integrated into the *C. dubliniensis* genome, and genome-wide gene expression changes were assessed via RNA-Seq (y-axis), compared to an empty vector control strain (x-axis). Only two glycolysis genes, *PFK1* and *PFK26-2*, displayed a significant increase in gene expression compared to the empty vector control strain (negative binomial distribution, $P < 0.01$). FPKM is a measure of expression level (fragments per kb per million mapped reads).

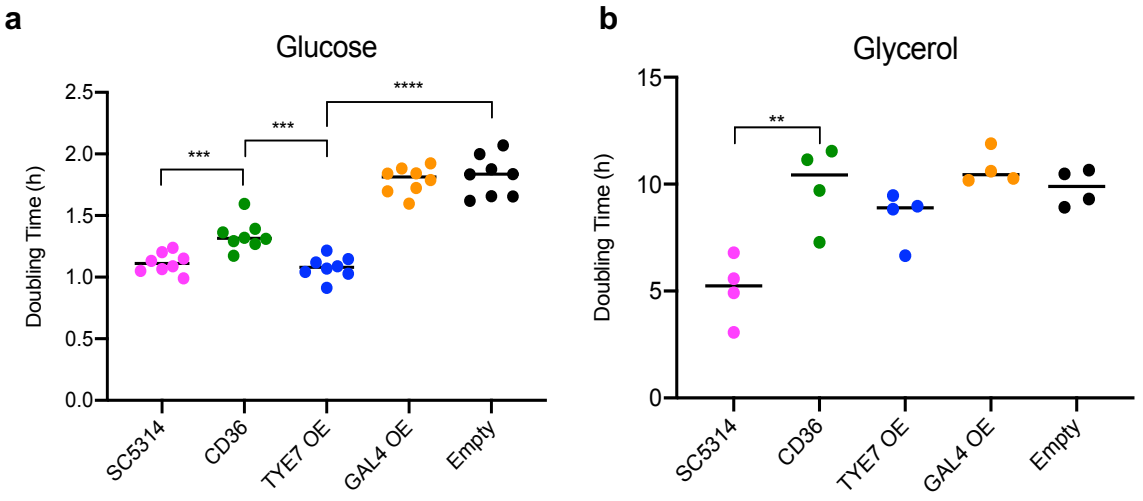

Fig. S5. Overexpression of glycolysis regulator *TYE7* in *C. dubliniensis* increases its growth rate under anaerobic conditions. a, *C. albicans* SC5314, *C. dubliniensis* CD36, and C. *dubliniensis* strains TYE7oe, GAL4oe, and Empty, described in Fig. 3, were grown in liquid media with glucose as the carbon source in anaerobic conditions to force glucose metabolism through glycolysis. Doubling time was calculated for each strain, *C. albicans* displayed significantly faster doubling times than *C. dubliniensis* ($P = 0.0007$, unpaired *t*-test, ***). The *C. dubliniensis* strain TYE7oe was observed to have a significantly faster doubling time than the Empty vector control strains ($P < 0.0001$, unpaired *t*-test, ****). **b,** Overexpression of *C. albicans TYE7* in *C. dubliniensis* does not increase the proliferation rate in non-fermentable carbon sources. The same strains as described in a, were grown in liquid media with 2% glycerol described in the presence of ambient oxygen. Doubling time was calculated as above and *C. albicans* displayed significantly shorter doubling times than *C. dubliniensis* ($P = 0.0079$, unpaired *t*-test, **); however, the *C. dubliniensis* strain TYE7oe and Empty vector control strain did not differ significantly in growth rate ($P = 0.1229$, unpaired *t*-test).

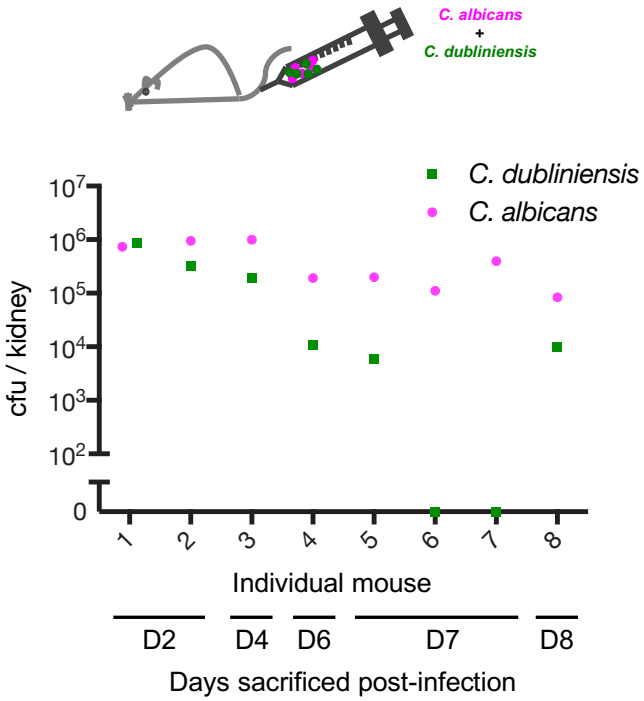

Fig. S6. *C. albicans* outcompetes *C. dubliniensis* in mice as determined by titers in the kidney. Eight female adult BALB/c mice were systemically infected through tail-vein injection with 5 x 10^5 cells of an equal mix of auxotrophically marked *C. albicans* and *C. dubliniensis*. Mice were sacrificed upon illness and kidney fungal burden was assessed. *C. albicans* and *C. dubliniensis* were distinguished by their ability to grow on media lacking histidine or leucine, respectively. For each mouse, the amount of *C. dubliniensis* recovered from the kidneys are shown with a green square and the amount of *C. albicans* recovered from the same mouse are shown with a magenta circle. The number of days post-infection at which the mice were sacrificed is given on the x-axis. For mouse 6 and 7, sacrificed 7 days post-infection, no *C. dubliniensis* colonies were recovered. Statistical analysis was performed on kidney fungal burden across all mice using the Mann-Whitney test and a significance difference between *C. dubliniensis* and *C. albicans* levels across all mice was observed (P = 0.03669).

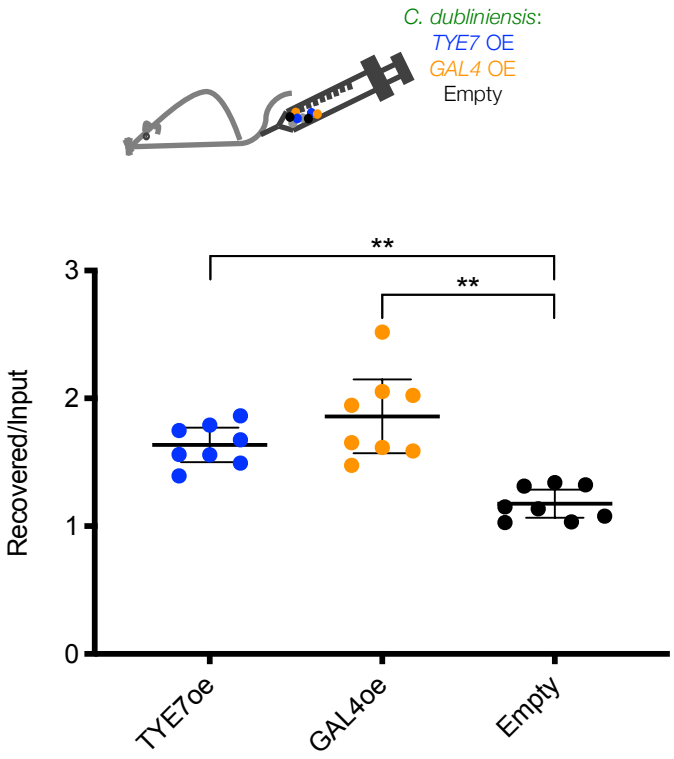

Fig. S7. Overexpression of the glycolysis regulators *TYE7* or *GAL4* increase *C. dubliniensis* proliferation in the mouse. Eight female adult BALB/c mice were systemically infected with a high inoculum (10^6 cells) of an equal mixture of *C. dubliniensis* strains TYE7oe, GAL4oe, and Empty, as described in previous experiments. Mice were sacrificed 24 hours post-infection, kidneys were dissected and genomic DNA recovered to determine relative abundance of individual strains recovered in the kidneys versus the starting inoculum by qPCR using primers specific to each strain. Strains overexpressing either *TYE7* or *GAL4* were enriched relative to the empty vector control strain in the mouse kidneys compared to the inoculum ($P = 0.0078$, Wilcoxon matched-pairs signed rank test). Bars on the graph represent mean values with 95% confidence intervals.

Table S1. Strains used in this study.

Accession	Alias	Genotype	Species	Source
CdSB2	CD36	WT	*C. dubliniensis*	(3)
CdSB4	CEM035	*his1Δ*::FRT/*his1Δ*::FRT	*C. dubliniensis*	(4)
CaSB14	SC5314	WT	*C. albicans*	(5)
CaSB16		*leu2*Δ::FRT/*LEU2*	*C. albicans*	This study
CaSB37		*leu2*Δ::FRT/*leu2*Δ::FRT	*C. albicans*	This study
CaSB67		*leu2*Δ::FRT/*leu2*Δ::FRT *mtl-alpha::FRT/MTL a*	*C. albicans*	This study
CdSB68		*his1Δ*::FRT/*his1Δ*::FRT *mtlA::FRT/MTL alpha*	*C. dubliniensis*	This study
CdSB69		*his1Δ*::FRT/*his1Δ*::FRT *mtl-alpha::FRT/MTL a*	*C. dubliniensis*	This study
CaSB80		*leu2*Δ::FRT/*leu2*Δ::FRT *mtl-alpha::FRT/MTL a* Opaque	*C. albicans*	This study
CadSB112		SC5314 *leu2*Δ*::FRT*/*leu2*Δ*::FRT* MTLa Op- A CD36 *his1*Δ*::*FRT/*his1*Δ*::FRT* MTL alpha/-	*C. albicans* x *C. dubliniensis* tetraploid hybrid	This study
CaSB133		tye7Δtye7Δ leu2Δ/Δ MTL alpha/- A Opaque	*C. albicans*	This study
CdSB144		*his1*Δ/Δ *TYE7/TYE7-* 13xMyc MTL a/-	*C. dubliniensis*	This study
CdSB163		*pTDH3-CaTYE7*	*C. dubliniensis*	This study
CdSB164		*pTDH3-CaGAL4*	*C. dubliniensis*	This study
CdSB169		*pTDH3*-empty vector control	*C. dubliniensis*	This study

Table S2. Plasmids used in this study.

Accession	Alias	Description	Reference
pSB1	pSFS2	A vector containing the *SAT1* resistance marker with a flippase driven by the *MAL2* maltose-inducible promoter all surrounded by FRT sites for marker recycling	(6)
pSB20	pADH57	A vector containing the constitutive *CaTDH3* promoter with homology to RP10. Can be cut with *Bgl*II, *Age*I, *Nco*I, etc within RP10 for genomic insertion at this location.	(7)
pSB25	pADH57 + *CaTYE7*	This construct is used for integrating a *C. albicans* copy of *TYE7* under the control of the *TDH3* promoter at the *RP10* locus in *C. dubliniensis*.	This Study
pSB26	pADH57 + *CaGAL4*	This construct is used for integrating a *C. albicans* copy of *GAL4* under the control of the *TDH3* promoter at the *RP10* locus in *C. dubliniensis*.	This Study

Table S3. Primers used for *in vivo* infection competition experiments.

Accession	Sequence
oSB184	TTGAAATTCCCTTCAATTGG
oSB201	GACATTGTTGTTGTTGTTAGCG
oSB273	GGCTTGTTCAATTGGGTAAG
oSB275	TGATATCGAATTCCTGCAGC

Dataset S1 (separate file). This workbook contains allele-specific expression data for all *C. albicans* – *C. dubliniensis* 1:1 orthologs in all conditions tested. Tab 1 lists all $\log_2$ ratios of *C. albicans*/*C. dubliniensis* gene expression values. Tab 2 lists all negative binomial distribution $\log_{10}$ P-values of *C. albicans* vs *C. dubliniensis* differential expression. Tab 3 contains empirically corrected ASE p-values. Tab 4 contains *C. albicans* gene expression data. Tab 5 contains *C. dubliniensis* gene expression data. Tab 6 lists descriptions of what is on each tab. Tab 7 lists experimental conditions.

Dataset S2 (separate file). This excel workbook contains rank-sum test analysis data for all GO categories tested for each of the experimental conditions tested, two replicates per condition. Tab 3 contains descriptions for each tab.

Dataset S3 (separate file). This excel workbook contains the sign test analysis data for all GO categories tested for each of the experimental conditions tested, two replicates per condition. Tab 3 contains descriptions for each tab.

Dataset S4 (separate file). This excel workbook contains the gene expression analysis data for the *C. dubliniensis TYE7* and *GAL4* overexpression strains versus the empty vector control strain. Tab 3 contains descriptions for each tab.

SI References

1. C. J. McManus *et al.*, Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Res* **20**, 816-825 (2010).
2. H. B. Fraser, Improving estimates of compensatory cis-trans regulatory divergence. *Trends Genet* **35**, 3-5 (2019).
3. 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* **141 (Pt 7)**, 1507-1521 (1995).
4. E. Mancera *et al.*, Genetic modification of closely related *Candida* species. *Front Microbiol* **10**, 357 (2019).
5. 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).
6. O. Reuss, A. Vik, R. Kolter, J. Morschhäuser, The SAT1 flipper, an optimized tool for gene disruption in Candida albicans. *Gene* **341**, 119-127 (2004).
7. 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).