


## RESEARCH ARTICLE

# *Candida albicans* selection for human commensalism results in substantial within-host diversity without decreasing fitness for invasive disease

Faith M. Anderson<sup>1</sup>, Noelle D. Visser<sup>1</sup><sup>Ⓜ</sup><sup>Ⓝ</sup>, Kevin R. Amses<sup>2</sup><sup>Ⓜ</sup><sup>Ⓝ</sup>, Andrea Hodgins-Davis<sup>1</sup>, Alexandra M. Weber<sup>3</sup>, Katura M. Metzner<sup>1</sup>, Michael J. McFadden<sup>1</sup>, Ryan E. Mills<sup>3,4</sup>, Matthew J. O'Meara<sup>3</sup>, Timothy Y. James<sup>2</sup>, Teresa R. O'Meara<sup>1</sup><sup>\*</sup>

**1** Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, **2** Department of Ecology and Evolution, University of Michigan, Ann Arbor, Michigan, United States of America, **3** Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, United States of America, **4** Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, United States of America

 These authors contributed equally to this work.

<sup>Ⓜ</sup> Current address: Department of Biology, University of Louisville, Louisville, Kentucky, United States of America

<sup>Ⓜ</sup><sup>Ⓝ</sup> Current address: Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

\* [tromeara@umich.edu](mailto:tromeara@umich.edu)



## OPEN ACCESS

**Citation:** Anderson FM, Visser ND, Amses KR, Hodgins-Davis A, Weber AM, Metzner KM, et al. (2023) *Candida albicans* selection for human commensalism results in substantial within-host diversity without decreasing fitness for invasive disease. PLoS Biol 21(5): e3001822. <https://doi.org/10.1371/journal.pbio.3001822>

**Academic Editor:** Aaron P. Mitchell, University of Georgia, UNITED STATES

**Received:** August 31, 2022

**Accepted:** April 12, 2023

**Published:** May 19, 2023

**Copyright:** © 2023 Anderson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Relevant data are within the paper and its [Supporting Information](#) files. Sequences are available at SRA under PRJNA875200.

**Funding:** Funding for this project included mCubed grant to TRO and TYJ, NIH grant NIH KA1137299 (NIAID) to TRO, NIAID T32 AI007528 to FMA, NIH 1F31HG010569 to AMW, NIH T32GM007544 to KRA, University of Michigan Postdoctoral Pioneer Program to MJM. TYJ is a fellow of the Canadian

## Abstract

*Candida albicans* is a frequent colonizer of human mucosal surfaces as well as an opportunistic pathogen. *C. albicans* is remarkably versatile in its ability to colonize diverse host sites with differences in oxygen and nutrient availability, pH, immune responses, and resident microbes, among other cues. It is unclear how the genetic background of a commensal colonizing population can influence the shift to pathogenicity. Therefore, we examined 910 commensal isolates from 35 healthy donors to identify host niche-specific adaptations. We demonstrate that healthy people are reservoirs for genotypically and phenotypically diverse *C. albicans* strains. Using limited diversity exploitation, we identified a single nucleotide change in the uncharacterized *ZMS1* transcription factor that was sufficient to drive hyper invasion into agar. We found that SC5314 was significantly different from the majority of both commensal and bloodstream isolates in its ability to induce host cell death. However, our commensal strains retained the capacity to cause disease in the *Galleria* model of systemic infection, including outcompeting the SC5314 reference strain during systemic competition assays. This study provides a global view of commensal strain variation and within-host strain diversity of *C. albicans* and suggests that selection for commensalism in humans does not result in a fitness cost for invasive disease.

Institute for Advanced Research program Fungal Kingdom: Threats & Opportunities. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

**Abbreviations:** GRACE, gene replacement and conditional expression; MLST, multilocus sequence typing; PI, propidium iodide; SNV, single-nucleotide variant; UMAP, uniform manifold approximation projection; YPD, yeast extract peptone dextrose.

## Introduction

*Candida albicans* is a common colonizer of humans, with between 20% and 80% of the world's population estimated to be asymptotically colonized at any given time [1,2], although this depends on many factors, including host health status and diet [3–5]. Colonization occurs at multiple body sites including the mouth, skin, GI tract, and vaginal tract [6,7]. These sites present a wide range of physiological stresses to colonizing fungi, including variation in pH, temperature, and oxygen levels, as well as nutrient limitation and host immune responses [6,8–10]. *C. albicans*–host interactions are generally commensal, but *C. albicans* can also act as an opportunistic pathogen, resulting in an estimated 400,000 serious bloodstream infections per year [11–13]. Additionally, *C. albicans* can cause more minor mucosal infections, including oral and vaginal thrush and skin infections [14]. As a consequence, *C. albicans* represents the second most common human fungal pathogen and the most common source of healthcare-associated fungal infections [15].

While host immune status is known to be an important predictor of disease outcomes, whether genetic variation between *C. albicans* strains also contributes to differences in virulence remains an open question. In the model yeast *Saccharomyces cerevisiae*, there is extensive variability in genotype and phenotype among different isolates that has been linked to the ability of *S. cerevisiae* to adapt to a wide range of environmental conditions [16,17]. Recent work has highlighted intraspecies variation in important aspects of *C. albicans* biology [18–21]. There are currently 17 clades of *C. albicans*, which were initially defined through multilocus sequence typing (MLST) [22–24]. More recently, genome sequencing has resolved finer population structure in the group [25]. Genetic epidemiology studies suggest that the major clades of *C. albicans* may differ in how frequently they are isolated from bloodstream infections or asymptomatic colonization [26], with 5 clades accounting for the majority of clinical isolates [27–29]. These clinical isolates have demonstrated significant variation in murine models of systemic infection, biofilm formation, cell wall remodeling, secretion of toxins and proteolytic enzymes, and morphological plasticity [18,26,30–35]. However, our primary understanding of the genotype–phenotype relationship in *C. albicans* results from analyses of a relatively limited set of pathogenic clinical isolates and their laboratory derivatives, with the majority of the work performed in the SC5314 genetic background. Detailed analysis of the genetic determinants of biofilm regulation between 5 different strains, each representing a different clade of *C. albicans*, have highlighted that circuit diversification is widespread between strains [18], adding complexity to our understanding of *C. albicans* biology.

Recent work suggests that *C. albicans* experiences fitness tradeoffs between invasive and colonizing growth, with selection for commensal behavior during colonization [36–42]. In serial passage experiments or competitive fitness experiments in the gut, mutations in key transcription factors controlling hyphal formation, *EFG1* and *FLO8*, resulted in increased fitness in the gut and decreased fitness in systemic models of infection [36–41]. In oral candidiasis, trisomic strains have been identified with a commensal phenotype [43]. Additionally, the 529L strain of *C. albicans* causes less damage and inflammation during oropharyngeal candidiasis [44] and persists at a higher fungal burden in both the mouth and the gut [21]. These potentially divergent selection pressures imply that commensal *C. albicans* strains can differ from isolates that cause invasive disease, but the genetic determinants underlying this difference have not been defined.

Here, we demonstrate that healthy people are reservoirs for genotypically and phenotypically diverse *C. albicans* strains that retain their capacity to cause disease. Our results indicate that individuals can be colonized by strains from multiple clades and that these commensal isolates have extensive variation in growth, stress response, biofilm formation, and interaction

with macrophages, but this variation was not explainable by sample origin site. We found that although the commensal isolates had a reduced capacity to cause macrophage cell death, the majority of the strains showed increased competitive fitness in invertebrate models of systemic candidiasis compared with the reference SC5314 strain and retained their capacity to cause disease during monotypic infections. Together, these data suggest that the selective pressures experienced by *C. albicans* during commensal colonization do not necessarily result in decreased pathogenic potential.

## Results

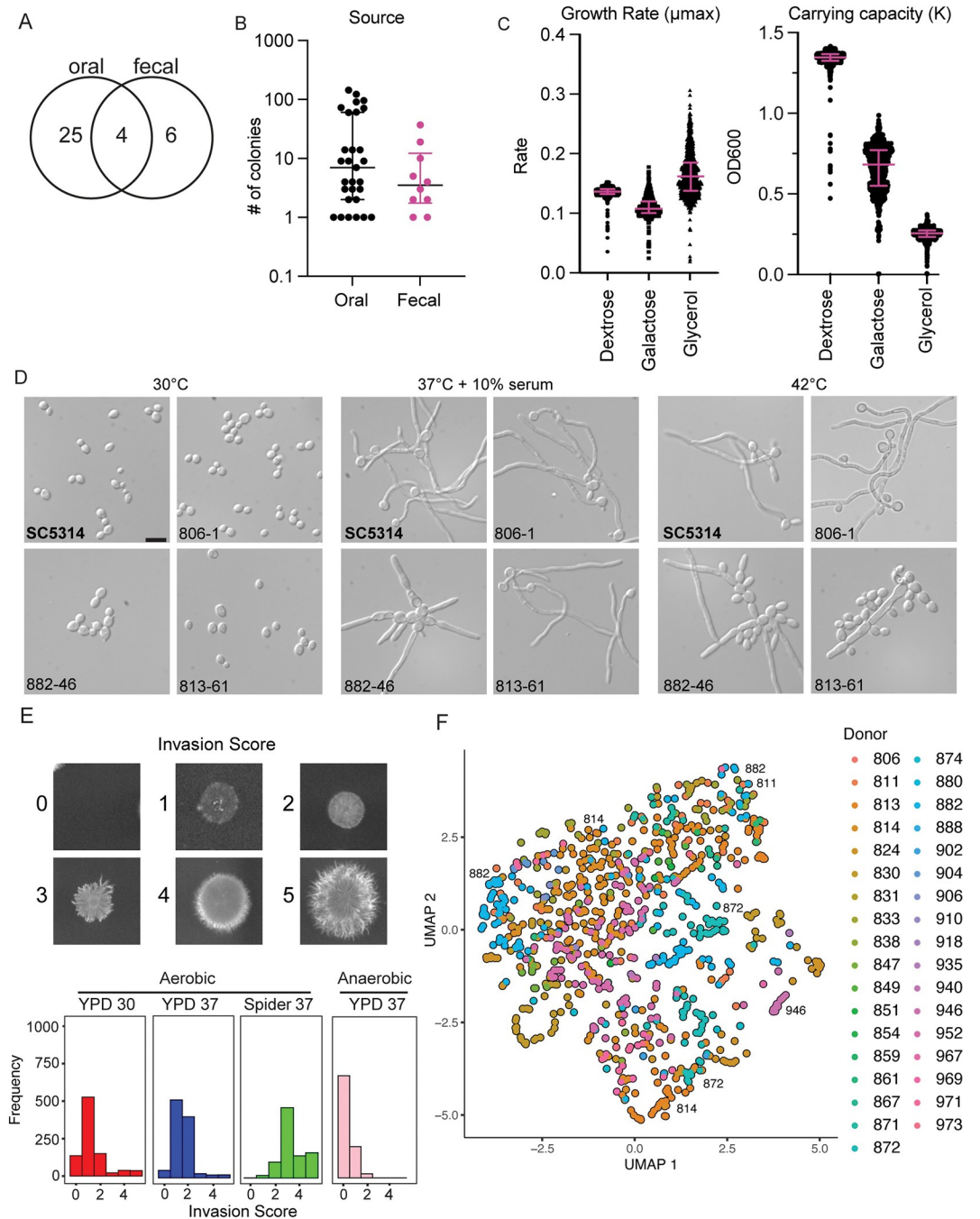
### Phenotypic characterization of commensal *C. albicans* strains

We sourced *C. albicans* from oral and fecal samples from undergraduate student donors, using these as a representative sample of colonizing *C. albicans* strains from healthy individuals. In this population of students, 29% (29/98) were positive for oral *Candida* colonization and 10% (10/98) were positive for fecal colonization, including 4 students who exhibited both oral and fecal colonization (Fig 1A). From each host and site, we collected every individual colony present on the BD ChromAgar plates and confirmed species identity through ITS amplicon sequencing. Overall, we obtained 910 *C. albicans* isolates (fecal:  $n = 84$  colonies, oral:  $n = 826$  colonies) (Fig 1B).

We then performed growth assays on all 910 isolates and the SC5314 reference strain in rich medium (yeast peptone) with dextrose, galactose, or glycerol as the carbon source. We observed primarily unimodal distributions with a long tail of slow-growing strains for both exponential growth ( $\mu_{max}$ ) and saturating density ( $K$ ) in rich media (Fig 1C). Although the 5 slowest-growing isolates were all obtained from oral samples, we did not observe a significant difference in growth rate between oral and fecal samples (S1A Fig), and overall, the growth rates between the different carbon sources were not correlated (S1B Fig).

As filamentation has been tightly associated with virulence, and because previous work in murine models has suggested that gut adapted strains may lose their ability to filament [21,36,38], we examined each isolate in the collection for their ability to form hyphae using 10% serum and febrile temperatures as 2 inducing cues. Additionally, we examined the morphology of each strain under the non-inducing condition of yeast extract peptone dextrose (YPD) at 30°C to identify constitutive filamentation, as has been observed from isolates collected from sputum samples from cystic fibrosis patients [45]. We observed that while none of the strains were constitutively filamentous, several strains aggregated at YPD 30°C (Fig 1D; represented by 882–46). These aggregative isolates were more likely to reach a lower carrying capacity; however, it is possible that the aggregation interfered with the accuracy of the OD reads.

All strains were able to filament in response to the standard laboratory inducing conditions of serum and high temperature, albeit with some variation in the number of hyphal cells, shape, and aggregation of the cells (Fig 1D). Under hyphal-inducing conditions, the aggregating strains formed filaments while the mother yeast cells remained connected, resulting in a star-like pattern. Some isolates showed different filamentation patterns under the different inducing conditions. For example, isolate 813–61 formed filaments that resembled those of SC5314 under 37°C + 10% serum, but under the 42°C inducing cue, this isolate formed fewer filaments with many yeast cells aggregating around the filament. The ability of all our commensal isolates to form filaments may be consistent with the hypothesis that interaction with bacteria in the gut maintains selection for the hyphal program [36,46]. Although there were isolates, such as 813–61, that showed more of a mixture of yeast and hyphae than the SC5314 parent that are potentially consistent with the yeast-locked phenotypes observed in Tso and



**Fig 1. Characterization of isolates from healthy donors reveals extensive phenotypic heterogeneity.** (A) The number of healthy donors from whom *C. albicans* colonies were obtained by isolation site with 4 donors exhibiting both oral and fecal colonization. (B) Number of *C. albicans* colonies isolated per donor from each sample site. Error bars represent median and interquartile range in number of colonies from each positive individual donor. Significance determined by Wilcoxon rank sum test. (C) Strains varied in maximum growth rate and carrying capacity in response to different carbon sources. Growth curves were performed on each isolate under 3 carbon sources at 30°C for 24 h for dextrose and 48 h for glycerol and galactose in biological duplicate. Rate and carrying capacity were determined using the GrowthcurveR analysis package. Error bars represent median and interquartile range of the growth parameter. (D) All strains retained the capacity to filament in liquid inducing cues, but some demonstrated altered morphology and aggregation. Strains were incubated in the indicated conditions and imaged at 40× magnification. Scale = 10 μM. (E) Strains varied in their capacity to invade into solid agar. Colonies were incubated on the indicated conditions for 5 days before gentle washing and imaging for invasion. (F) UMAP plot for strain phenotypic similarity. All growth conditions and invasion phenotypes were nonlinearly projected into 2D space and colored by donor.

<https://doi.org/10.1371/journal.pbio.3001822.g001>

colleagues (2018) [36], the majority of our commensal strains retained full capacity to filament in our in vitro assays. This may suggest that the antibiotic or germ-free mouse models are not fully recapitulating important features of *C. albicans*–human interactions.

Filamentation programs under solid and liquid growth can involve distinct genetic programs [47], and the human host can present a variety of substrates for *C. albicans* to utilize. Therefore, we also tested each isolate for its capacity to invade into solid agar media under multiple conditions: YPD agar at 37°C under anaerobic conditions as a minimal inducing cue, YPD agar at 30°C as the baseline condition, YPD agar at 37°C as an intermediate inducing condition, and Spider agar at 37°C as a strong inducing condition (Fig 1D). As expected for a strong inducing cue, the highest degree of invasive growth was observed on Spider agar at 37°C [47]. Interestingly, although all strains showed the ability to form hyphae under liquid culture conditions, many strains failed to invade into solid agar, giving scores of 0 or 1 under multiple conditions (Fig 1E). We also observed substantial phenotypic heterogeneity among strains isolated from the same host, including in strains isolated from the same site, consistent with each individual being colonized by multiple, phenotypically distinct, strains of *C. albicans*.

To examine the relationship between the strains, observed phenotypes, and donors, we generated a uniform manifold approximation projection (UMAP) embedding, which will plot strains with similar phenotypes closer together and strains with dissimilar phenotypes farther apart based on the Euclidean metric (Fig 1F). Using this projection, we saw significant overlap in strain phenotypes from many donors, although there were some donors that had strains that clustered away from others, such as strain 946.

Additionally, some donors had 2 separate clusters of strains, such as those from donor 814, 872, and 882. We also observed that some strains from multiple donors clustered together in this projection, including the aggregating strains from donors 811 and 882. However, we did not observe clustering based on sample origin site (S1C Fig). Together, these experiments demonstrate a large variation in phenotypes between commensal isolates, even among traits, such as filamentation and invasion, that are often correlated with virulence.

## Genomic variability

Due to the extensive variation in observed phenotypes among the isolates, we next wanted to characterize the genomic variability in these strains, at both a sequence and structural variation level. To carry out these analyses, we selected a set of 45 commensal isolates, hereafter referred to as the “condensed set.” Isolates were chosen for inclusion in the condensed set where we had matched pairs from oral and fecal isolates from the same donor or isolates from the same donor that exhibited multiple phenotypes based on growth, filamentation, or invasion. For these analyses, we compared each isolate to the SC5314 reference strain. We also included the human isolate, CHN1, which has been previously phenotypically and genotypically characterized as a bona fide commensal and is able to stably colonize the murine gastrointestinal tract for long periods of time, in contrast to SC5314 [21,48].

Previous descriptions of population genomic variation in *C. albicans* have largely relied on short-read sequencing approaches, which are limited in their ability to resolve structural variation between strains. Therefore, we performed whole-genome sequencing on all 45 commensal strains, SC5314, and CHN1 using the Transposase Enzyme Linked Long-read Sequencing (TELL-Seq) method for library preparation [49]. This approach uses barcode linked-reads to produce synthetic long reads with Illumina quality sequence, thus allowing us to capture both single-nucleotide variants (SNVs) and structural variants.

To identify SNVs and compare our strains to the existing set of sequenced *C. albicans* isolates, we collected 388 previously published *C. albicans* genomes and, after filtering, we

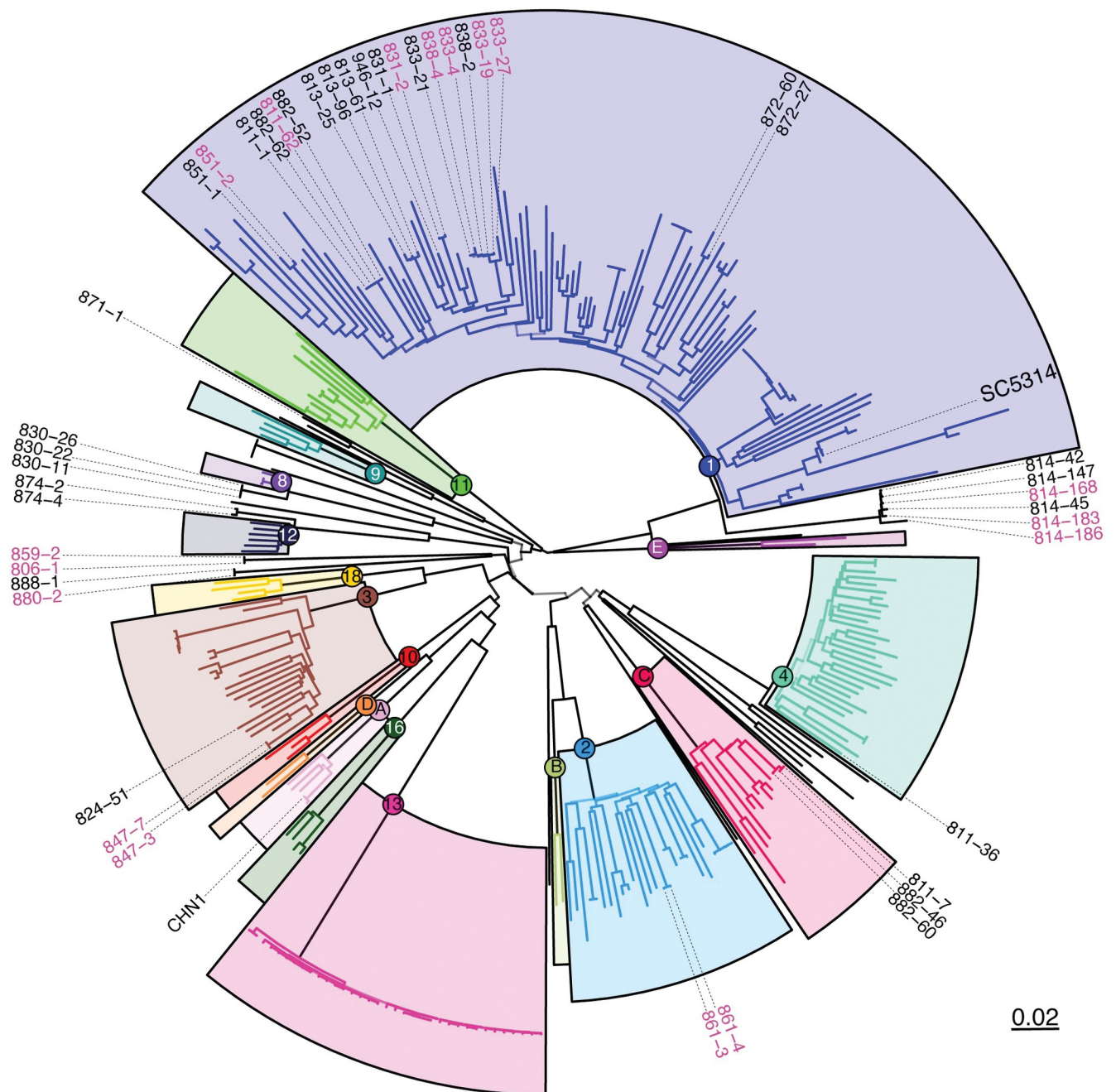
mapped them, and our 45 newly generated genomes, to the SC5314 reference genome with BWA-MEM [50]. We called variants using GATK HaplotypeCaller and after filtering, we obtained final set of 112,136 high-quality SNVs across 431 remaining samples. Of these, 90,675 (80.86%) were represented in at least 1 member of our set of newly sequenced strains. The population diversity captured in this analysis was consistent with the largest previous analysis of *C. albicans* genomic variation, which called 589,255 SNPs from 182 genomes [25]. Although our final SNV set was significantly smaller than that identified in past work, our filtering criteria were significantly more stringent and robust for inferring population level patterns.

We then wanted to place our newly sequenced isolates in the phylogenetic context of previous work on *C. albicans* strains [19,25,51]. To remove redundancy and focus on natural *C. albicans* diversification, we removed samples corresponding to resequenced strains (e.g., multiple SC5314 samples present in full data set) and those sequenced as part of experimental evolution studies (i.e., [36,52,53]). Following removal of these samples, we were left with 324 sample SNV profiles, which were then used to cluster the samples into a dendrogram of relationships. Despite the reduced size of our data set, our SNV-based clustering recovers all major accepted clades of *C. albicans* (Fig 2) [22–26]. The fact that >80% of the 112,136 high-quality SNVs we identified are represented in the genome sequences of the 45 new *C. albicans* isolates we sequenced, in addition to their clustering near 8 major clades, asserts the high degree of diversity captured in our study of commensal *C. albicans* strains actively colonizing humans. In line with past work and underpinning the validity of our SNV-based clustering, we did not collect isolates from clades of *C. albicans* known to exhibit a high degree of geographic specificity (e.g., Clade 13) (Fig 2) [25].

The majority of the new isolates (26/45) belonged to Clade 1, of which the reference strain SC5314 is also a member. We identified cases where isolates from the same donor clustered tightly together, such as donor 814, whose 6 strains included in the condensed set clustered in Clade 1. Consistent with previous reports on microevolution in the host [53], we observed primarily SNVs and short-tract loss of heterozygosity between these 6 isolates, perhaps consistent with clonal expansion and diversification during colonization. In the 4 donors colonized at both the oral and fecal sites (811, 814, 831, and 851), both oral and fecal isolates were closely genetically related. However, we also identified donors with colonizing strains from multiple clades, such as donor 882, whose 4 strains came from Clade 1 and Clade C, or donor 811, whose 4 strains came from Clade 1, Clade C, and Clade 4 (Fig 2). Interestingly, some isolates from multiple donors clustered within one another, such as those the donor pairs 838 & 833 and 882 & 811, potentially indicating transmission events between individuals, due to the phylogenetic closeness of the strains. Together, this suggests that variation in an individual's colonizing *C. albicans* strains may come from both within-host diversification and between-host transmission.

To characterize genomic variation at a structural level, we performed pulsed-field gel electrophoresis to separate the chromosomes of our condensed set of commensal *C. albicans* isolates, including SC5314 as a reference (Figs 3A and S3A). The commensal *C. albicans* strains show between 7 and 10 chromosome bands, ranging from approximately 0.8 MB to approximately 3.2 MB in size [54]. We observed many size differences in chromosomes, especially among the smaller chromosomes (corresponding with Chr 5, 6, and 7 in SC5314), but we also observed size variation in large chromosomes at approximately ~1 MB, ~1.5 MB, and ~2 MB, potentially indicating variation in Chrs 2, 4, and R. Size variation of larger chromosomes may be mostly associated with allelic expansion or contraction of the rDNA repeat array on ChrR.

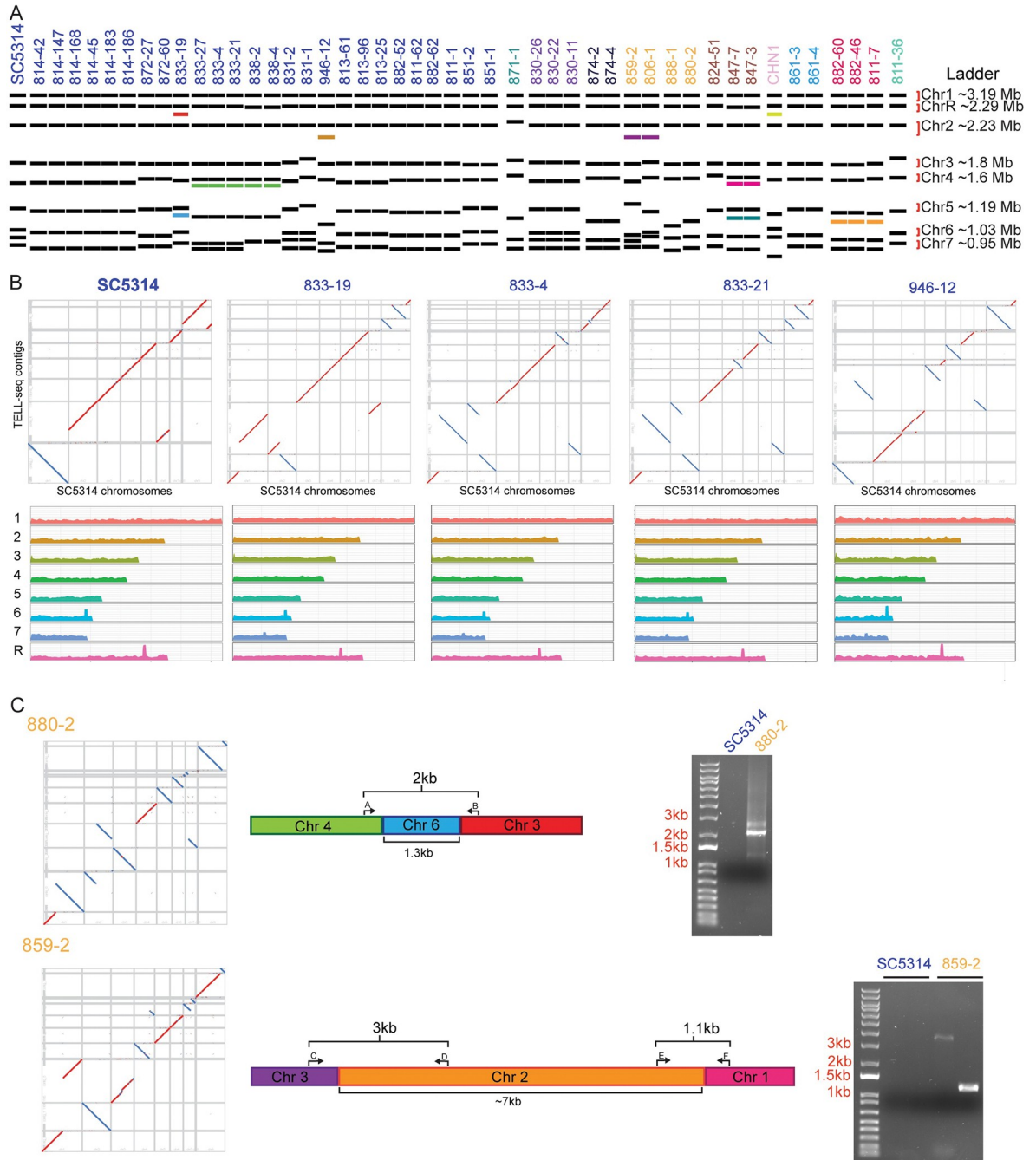
When comparing the chromosomal structural variations to our SNV tree, we identified several unique patterns and observed that a phylogenetic ordering of the isolates did not encompass the structural variation. For example, a subset of isolates within Clade 1 from donors 833



**Fig 2. The newly sequenced *C. albicans* commensal isolates include representatives from multiple clades.** A maximum likelihood tree showing the phylogenetic relationships between the 324 isolates analyzed via (neighbor joining). Previous clusters from [25] are highlighted with colored boxes. Isolates in black text are oral samples and isolates in pink text are fecal samples. Bootstrap values represented by backbone transparency. See S2 Fig for additional information on bootstrap values.

<https://doi.org/10.1371/journal.pbio.3001822.g002>

and 838 all contained an additional band below Chr 4, with the exception of isolate 833–19, despite being within the same phylogenetic cluster based on SNV analyses (Fig 3A). We also observed clade-level variation in karyotypes: Isolate 871–1 was the only strain in the condensed set from Clade 9, and this strain showed unique banding patterns at Chrs 2, 3, and 4. The 3 strains from Clade C, which originated from donors 882 and 811, all contained a band between



**Fig 3. *C. albicans* commensal isolates display extensive structural variation.** (A) CHEF karyotyping gels show alterations in chromosome size and number between *C. albicans* isolates. Isolate labels were colored based on the nearest defined cluster from Fig 2 and ordered based on phylogeny. Unique band sizes are indicated with the colored bands. Gel images in S3 Fig. (B) Above: Dot plots showing the synteny alignments between new isolates and SC5314. Below: Coverage maps in 5 kb sliding windows when aligned to the SC5314 reference genome. (C) Isolates 880–2 and 859–2 chromosomal fusion events chosen for PCR validation. Bands corresponding with fusion events were not present in the SC5314 reference strain.

<https://doi.org/10.1371/journal.pbio.3001822.g003>

Chrs 5 and 6. Finally, the 2 closely related strains, 859–2 and 806–1, both contained a band between Chrs 2 and 3 (Fig 3A). The observed altered band sizes could represent either loss of genome content or chromosomal fusion events. An advantage of our TELL-seq-based



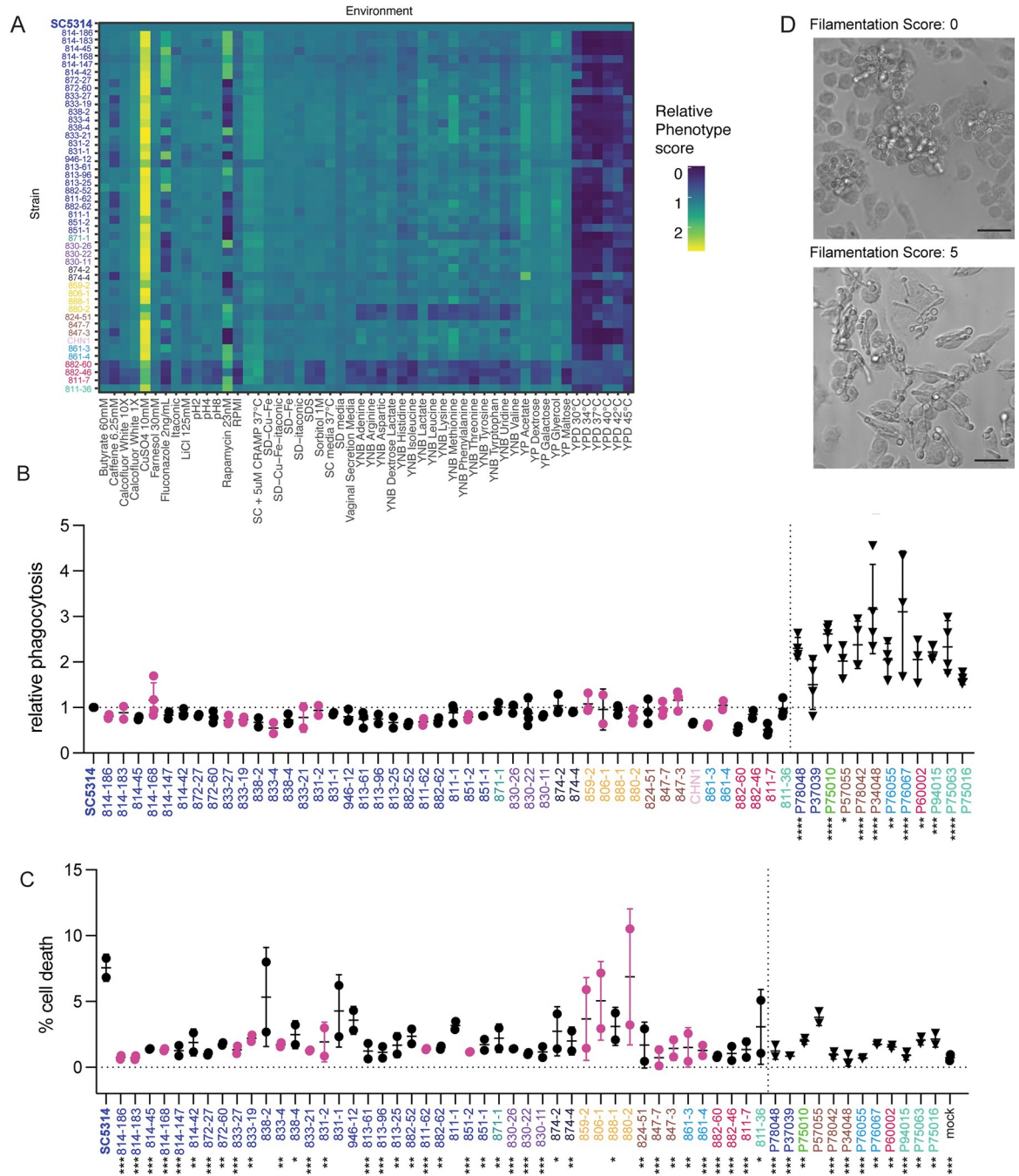
platform was the opportunity to resolve the differences between the SNV-level analysis and the structural variants that we identified through the CHEF gels. We used the synthetic long-reads to assemble contigs for each strain, using Universal Sequencing's TELL-seq pipelines: Tell-Read and Tell-Link (Fig 3B). We could observe variation in copy number across the chromosomes (Figs 3B and S3B), as well as potential inversions, duplications, or deletions (S3C Fig) [55,56]. We also see some peaks in copy number that are associated with MRS and rDNA sequences that were not resolved in our mapping. However, there may be some additional technical limitations of the TELL-seq approach, as we identified some inversions, duplications, and deletions in the resequenced SC5314 (S3C Fig). Therefore, we focused on those that were unique to the commensal isolates.

Notably, we were able to identify structural rearrangements and putative chromosome fusions that occurred in a set of the commensal isolates (Fig 3C and S4 Table) and used PCR to test for the presence of the fusion event. In strain 880–2, we observed a fusion event between chromosomes 3 and 4, connected by a 1.3 kb intervening sequence (Fig 3C). This intervening sequence had 94% sequence identity to an intergenic sub-telomeric region of chromosome 6. To determine whether this was a true event or a sequencing artifact, we designed primers to span the junction and performed PCR to amplify the fusion (Fig 3C). Using this approach, we observed that in strain 880–2, there is a bona fide structural rearrangement that links chromosomes 3 and 4. In strain 859–2, we observed a fusion event between chromosomes 1 and 3, connected by an approximately 7 kb intervening sequence with no obvious sequence identity to the SC5314 reference strain, but instead had 99.76% sequence identity to a region on chromosome 2 from *C. albicans* strain TIMM 1768, a highly virulent strain originally isolated from the feces of a candidiasis patient [57] (Fig 3C). TIMM 1768 is closely related to CHN1, a member of Clade A [56]. We were again able to use PCR to span both junctions observed the presence of the fusion between chromosomes 1, 2, and 3 (Fig 3C). This fusion event may correspond to the additional chromosomal band between chromosomes 2 and 3 that we observed in the karyotype for this strain. Importantly, neither of these fusion events were present in the SC5314 reference strain, indicating that the fusions were unique to the specific isolate (Fig 3C). These structural variations were not captured in the SNV analysis and may play important roles in gene regulation or phenotypic variation between the strains.

## Deep phenotyping of commensal isolates

The set of isolates for sequencing were initially chosen based on variation in growth rate in rich medium and alterations in invasion into agar. However, we hypothesized that we may identify site-specific adaptations, as host sites commonly colonized by *C. albicans* vary dramatically in environmental cues. Additionally, we hypothesized we may identify phenotypes associated with specific *C. albicans* clades, as we were able to identify structural variants shared between closely related isolates. To test this, we performed a set of growth analyses under multiple environmental conditions, including pH stresses, nutrient limitation, cell wall stressors, and antifungal drugs (Fig 4A). These analyses produced a dense array of quantitative phenotypic information for each strain.

From these data, we identified 3 strains, 882–60, 882–46, and 811–7, that consistently reached a lower carrying capacity than the wild type under multiple conditions; this is consistent with the aggregative phenotype exhibited by these isolates at 30°C (Fig 1D). These strains all belonged to Clade C and were closely related, despite arising from 2 donors (Fig 4A). Growth rates in the nutrient limitation conditions were generally correlated with each other. However, we did not observe a correlation between body site and growth rate, even in response to cues that would appear to be specific for a particular body site, such as anaerobic



**Fig 4. Deep phenotyping reveals heterogeneity in vitro and host response phenotypes.** (A) Growth curve analysis under multiple environmental conditions. Carrying capacity (K) was normalized to SC5314 and the fold-change plotted by heatmap. Aggregating strains (882–60, 882–46, and 811–7) demonstrate a consistently lower carrying capacity. (B) Relative macrophage phagocytosis rates of commensal isolates to reference strain SC5314. Black data points indicate an oral isolate and pink data points indicate a fecal isolate. Triangles represent bloodstream isolates. (C) Macrophage cell death rates. (D) Representative images of isolates following 4 h macrophage infection. Representative filamentation score of 0 (top). Representative filamentation score of 5 (bottom) and 20× magnification. Scale = 50 μM. For phagocytosis and cell death rates, significant differences from the SC5314 reference strain were determined by one-way ANOVA, with Dunnett’s multiple correction testing. Asterisks indicate  $P < 0.05$  (\*),  $P < 0.005$  (\*\*),  $P < 0.001$  (\*\*\*), and  $P < 0.0001$  (\*\*\*\*).

<https://doi.org/10.1371/journal.pbio.3001822.g004>

growth. Across the commensal isolates, we noted the most variation in growth in response to caffeine and the antifungals fluconazole and rapamycin. In addition to growth, we measured each of the strains for their ability to form biofilms on plastic surfaces [58]. Although we observed variation between the strains, there was no correlation between isolation site or clade with the propensity of isolates to form biofilms (S4 Fig).

Our dense array of phenotypic data across 45 *C. albicans* isolates and 8 clades reveal that commensal isolates largely retain the plasticity to grow efficiently under diverse environmental cues, even those not immediately relevant to their colonizing site, as we did not observe growth enrichment in cues specific to isolation sites.

A major stress condition and environmental factor impacting *C. albicans* in the host is the immune response. Therefore, we moved from pure growth assays to measuring host–microbe interactions, using macrophages as representative phagocytes. We first hypothesized that the oral strains may show decreased recognition by macrophages, as persistent oral isolates were recently shown to result in reduced immune recognition and inflammation in both an OPC model of infection and in cell culture [59]. We tested this by measuring phagocytosis of each strain by immortalized bone marrow-derived macrophages and determining the ratio of internalized to external cells by differential staining and microscopy (Fig 4B) [60]. Although most isolates were not significantly different from the SC5314 reference, the isolates generally had a lower phagocytic rate than SC5314. Additionally, there was no correlation between sample origin site or clade with phagocytosis rate.

As phagocytosis was not a major differentiating factor between strains, we then wanted to examine whether the strains would induce different levels of macrophage cell death. We primed bone marrow-derived macrophage for 2 h with LPS before infecting with each of our isolates for 4 h. Following infection, we stained the cells with propidium iodide (PI) as a measure of cell death (Fig 4C). Strikingly, many of the commensal isolates induced less cell death than SC5314. Our results are consistent with previous reports that commensal oral isolates of *C. albicans* are characterized by a reduced capacity to damage host cells [61]. To test whether this increased killing capacity of SC5314 was shared with other bloodstream isolates, we measured the phagocytosis and cell death rates for 12 bloodstream isolates originally described in Hirakawa and colleagues (2015) [19]. These results indicated a higher rate of phagocytosis for the bloodstream isolates (Fig 4B). However, for most of the bloodstream isolates, we also observed decreased capacity for macrophage killing (Fig 4C). The bloodstream isolate P94015, which was identified by Hirakawa and colleagues as having a loss of function of the transcription factor *EFG1*, behaved similarly to the other bloodstream isolates in these assays, despite having a decrease in virulence in their animal model [19]. These results indicate that the ability to kill host cells is not conserved across bloodstream isolates.

Recently, we showed that *C. albicans* mutants that filament in serum are not always filamentous within macrophages [62]. As filamentation is linked, but not required, for inflammatory activation within host phagocytes [62–64], and clinical isolates show variability in induction of host inflammatory responses [65,66], we examined the morphology of the commensal isolates after incubation for 4 h with macrophages. We observed considerable variation in the extent of filamentation among the natural isolates (Fig 4D), including strains that completely failed to filament and those that filamented more than the SC5314 reference strain. The aggregative strains, 882–60, 882–46, and 811–7 failed to filament during incubation with macrophages. Notably, the extent of filamentation did not correlate with colony morphology or invasion on agar, with many strains showing invasion into agar but no filamentation inside the macrophage, and vice versa (S5 Fig). Additionally, oral and fecal isolates both demonstrated defects in filamentation in macrophages, and filamentation in macrophages was not predictive of the phagocytic rate or cell death rate.

Using individual phenotypic measures, we were unable to identify associations between strains based on body site or donor. However, it is possible that the combined phenotypic and genotypic profile would identify clusters of strains with similar distinct phenotypes or reveal connections between isolates. Therefore, we turned to UMAP embedding, which will plot strains with similar phenotypes closer together and strains with dissimilar phenotypes farther apart based on the cosign metric. We did not observe any obvious clusters that segregated by isolation site, clade, or participant (S6 Fig). In sum, all of the commensal isolates showed extensive phenotypic variation, but this was not dependent on the body site or participant from which they were collected.

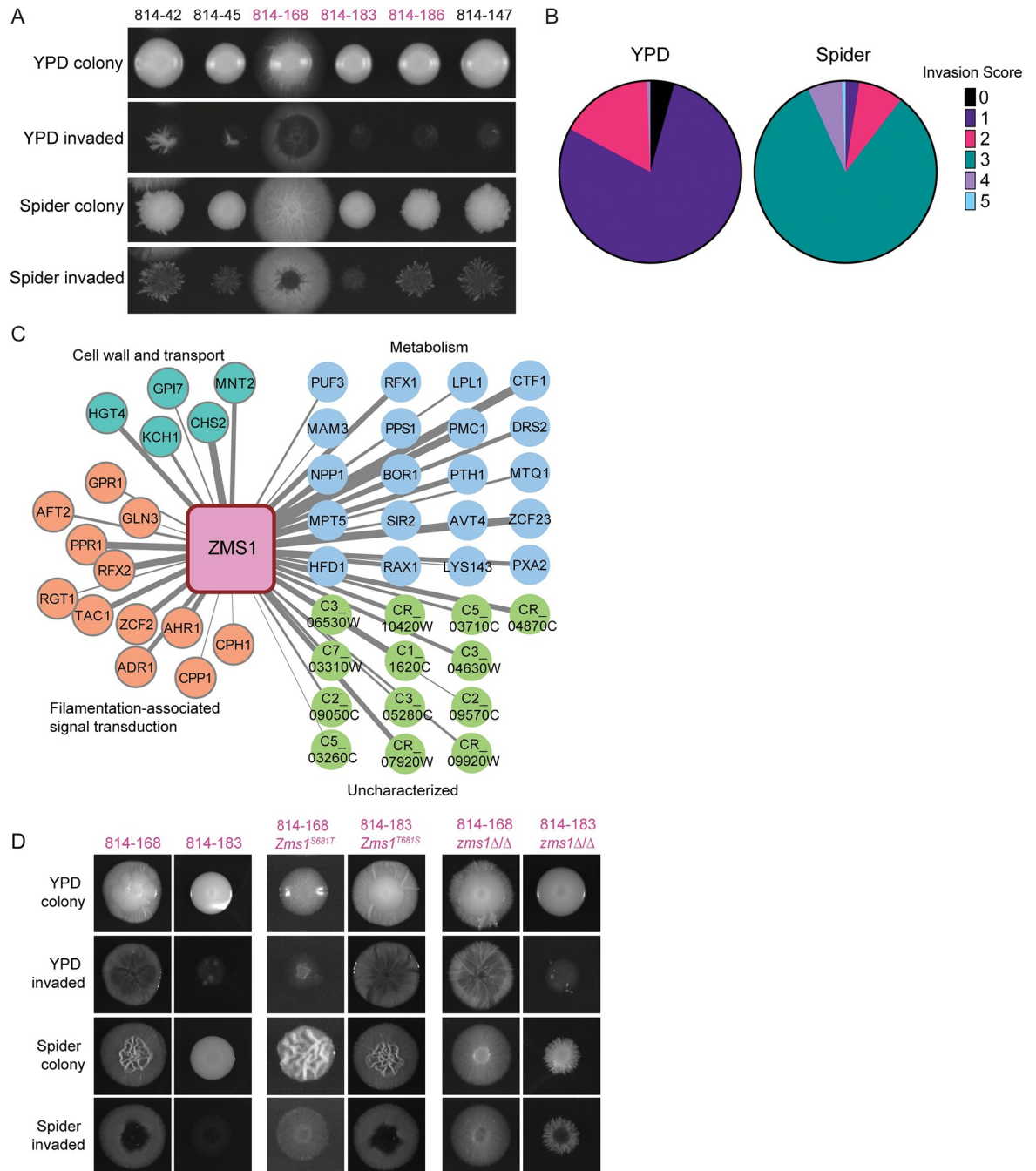
### Limited diversity exploitation

Genome-wide association studies have been a powerful tool for identifying the genetic basis of variation in phenotypes of interest in humans and other recombining species. However, the generally clonal and asexual reproduction of *C. albicans* creates a population structure that confounds traditional GWAS methods. By sampling multiple isolates from each individual, we were able to obtain phenotypically diverse strains with a limited set of unique SNPs between isolates, allowing us to identify causative variants associated with a particular phenotype.

We focused on the 6 strains from donor 814 included in the condensed set, which clustered tightly in Clade 1; we hypothesized this would allow us to identify causative variants associated with particular phenotypes that were divergent between strains. Our agar invasion analysis revealed that isolate 814–168 demonstrated hyper invasion into Spider agar at 37°C, whereas the other 5 isolates from the condensed set were less invasive (Fig 5A). Moreover, from this donor's 163 total isolates (144 oral isolates and 19 fecal isolates), only this single isolate exhibited the hyper invasive phenotype into Spider agar at 37°C (Fig 5B); this phenotype was the motivation for initially including this strain in the condensed set.

Variant analysis identified 12 genes with unique SNVs in the 814–168 strain compared with the other 5 sequenced 814 strains, including a heterozygous adenine to thymine SNV in the transcription factor *Zms1*, resulting in a change in amino acid 681 from a threonine to a serine. Moreover, co-expression analysis [67] of *ZMS1* revealed that it is highly correlated with genes involved in regulating the yeast-to-hyphal morphogenic transition (Fig 5C). To test whether this SNV can drive invasion, we generated complementation plasmids encoding each of the alleles from the 2 strain backgrounds and performed allele swap experiments between the high and low invasion strains. In the minimally invasive 814–183 background, replacing 1 copy of the endogenous *ZMS1* allele with the *ZMS1-T681S* allele was sufficient to drive hyper invasion into both YPD and Spider agar (Fig 5D). Similarly, replacing the invasive *ZMS1* allele with the *ZMS1-S681T* allele was sufficient to reduce invasion into both YPD and Spider agar (Fig 5D).

Previous work on the function of *Zms1* via deletion mutant analysis had not revealed a phenotype [68]; however, this was in the SC5314 genetic background and the impact of a specific transcription factor on a given phenotype can vary depending on the strain [18]. Therefore, we deleted *ZMS1* from both 814 backgrounds and tested the strains for invasion and filamentation. On YPD agar, as before, deletion of *ZMS1* in both genetic backgrounds had minimal effects, with the mutant strains behaving similarly to their parent strains (Fig 5D). However, on Spider agar, *ZMS1* deletion changed the colony morphology from wrinkly to smooth in the 814–168 background, although it did not decrease overall invasion. In contrast, deletion of *ZMS1* increased invasion in the otherwise minimally invasive 814–183 background, highlighting the differential impact of *ZMS1* mutation in the different genetic backgrounds (Fig 5D). Our results demonstrate that a single SNV changing amino acid 681 to a serine is a dominant



**Fig 5. SNV limited diversity exploitation analysis of donor 814 commensal isolates reveals role for Zms1 in regulating agar invasion.** (A) Agar invasion images for isolates from donor 814 included in the condensed set. Colonies were grown on YPD or Spider agar for 5 days. Invasion was determined after gentle washing. Isolates in black text are oral samples and isolates in pink text are fecal samples. (B) Agar invasion scores for all isolates from donor 814 under YPD and Spider conditions. (C) Co-expression analysis of Zms1. Width of the lines represents strength of the co-expression score. Gene names and predicted functions from Candida Genome Database. (D) Agar invasion images for allele swap and deletion strains. Colonies were grown on YPD or Spider agar for 5 days before imaging. Invasion was imaged after gentle washing. SNV, single-nucleotide variant; YPD, yeast extract peptone dextrose.

<https://doi.org/10.1371/journal.pbio.3001822.g005>

negative allele that is sufficient to drive a hyphal invasion program into Spider agar. We also identified natural variation that was distinct from deletion phenotypes. This approach, which we have termed “limited diversity exploitation,” highlights how deep phenotypic analysis of a limited set of natural isolates from a single host can be exploited to identify causative variants and identify new functions for under-characterized genes.

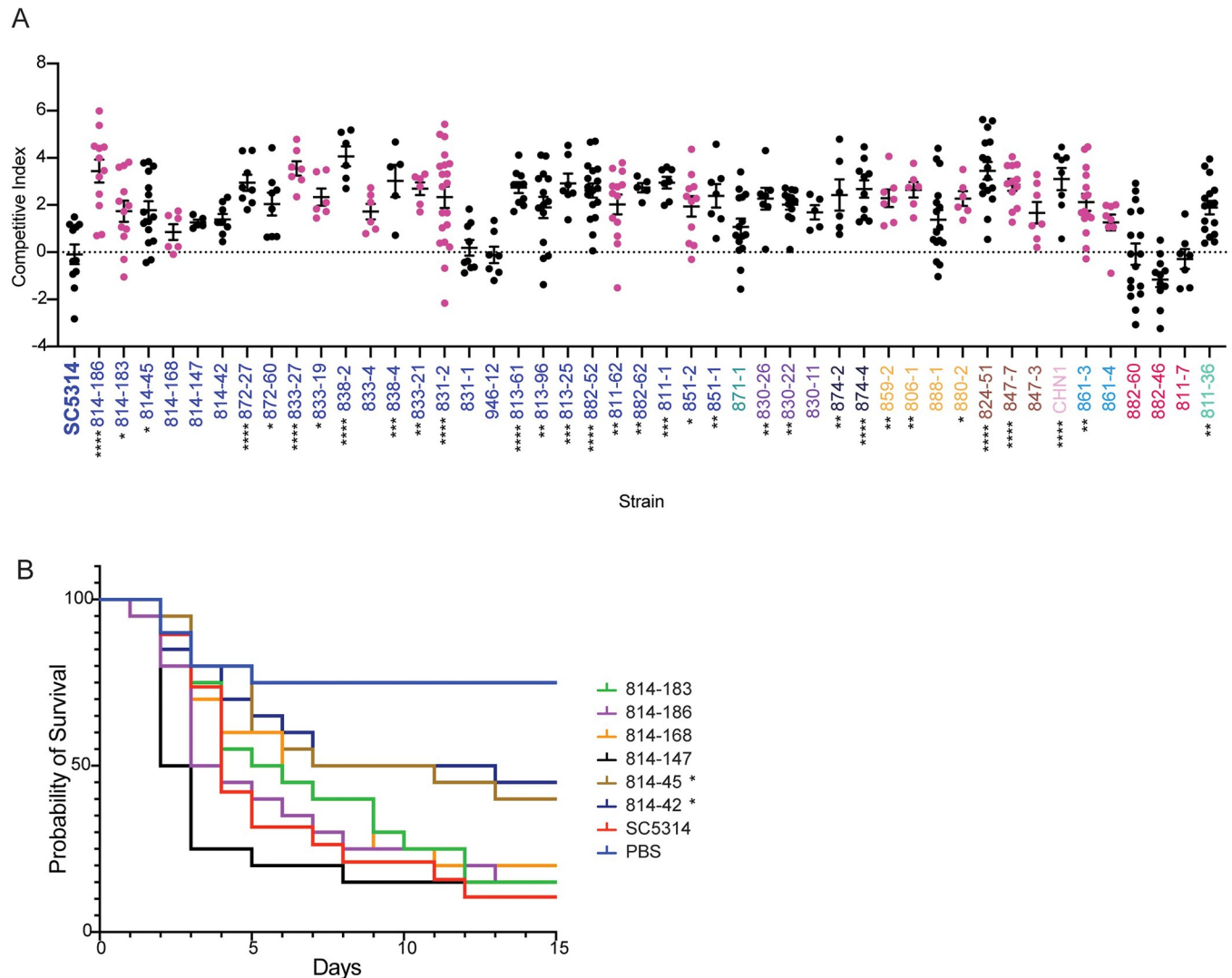
## Virulence

We next examined the fitness and virulence of the commensal isolates relative to the SC5314 reference strain; we hypothesized that the commensal isolates would have decreased virulence compared to SC5314, a clinical isolate. To test this, we turned to the *Galleria mellonella* model of systemic infection as this insect model is significantly correlated with the murine systemic infection model [19,69,70]. We first examined competitive fitness by infecting with a 1:1 mixture of fluorescently marked SC5314 and each sequenced commensal isolate [71]. After 3 days, the worms were homogenized and CFUs were plated to determine the competitive index, calculated as the log<sub>2</sub> ratio of fluorescent to nonfluorescent colonies [37]. When comparing the marked and unmarked SC5314 strains, we obtained a competitive index of 0, indicating that both strains are equally fit and that the fluorescence does not impose a fitness cost. In contrast, most commensal strains had a competitive index >2, suggesting that these strains have a competitive advantage over SC5314, even during systemic infection (Fig 6A). Notably, the 3 strains previously identified from Clade C, 882–60, 882–46, and 811–7, consistently demonstrated a competitive index of less than 0, indicating these strains are less fit than SC5314 in this model of infection. This is consistent with the aggregation exhibited by these strains in many growth conditions (Fig 4A).

The striking increased competitive fitness of the other isolates motivated us to test whether this increased ratio was correlated with increased disease. Here, we examined the survival of *G. mellonella* after performing monotypic infections. We started by using the 6 isolates from donor 814 to test the hypothesis that increased invasion is associated with increased disease. Three of the 6 isolates from 814 had increased competitive fitness compared to SC5314. However, the majority of the strains from donor 814 were not significantly different from the SC5314 reference strain ( $p > 0.05$ , log-rank test), including the hyper-invasive 814–168 isolate (Fig 6B). Isolates 814–42 and 814–45 had a slight defect in virulence ( $p < 0.05$  log-rank test). We additionally tested 2 other clusters of strains for their ability to cause systemic disease in the insect model. However, the human commensal isolates were again not significantly decreased in virulence from the SC5314 reference strain (S7 Fig), despite the increased ability of the SC5314 strain to cause host immune cell death (Fig 4C). Together, this suggests that there is not a consistent selection for avirulent behavior in the commensal human isolates in this model of systemic infection. Although we observed wide variation in *in vitro* host response phenotypes, the isolates generally retained their pathogenic potential, indicating that our *in vitro* assays may not capture the complex stresses experienced during whole-organism infections.

## Discussion

Intraspecies analyses of microbial strains can allow for the identification of variants that are associated with specific clinical outcomes, as demonstrated widely in bacterial virulence [72,73] and recently in the human fungal pathogen *Cryptococcus neoformans* [74,75]. Understanding the differences between strains can allow for insights into mechanisms of colonization and pathogenesis [18]. However, assessing the underlying genetic variation responsible for differences in virulence in *C. albicans* is challenging in the absence of clear candidate genes



**Fig 6. Commensal isolates retain pathogenic potential.** (A) Competition assays in *G. mellonella* demonstrate increased fitness of many commensal isolates compared to SC5314 reference. Isolates were competed against a fluorescent SC5314 isolate, starting at a 1:1 initial inoculum. Competitive fitness was calculated as the ratio between fluorescent and nonfluorescent colonies, normalized to the inoculum, and log<sub>2</sub> transformed. Black data points indicate an oral isolate and pink data points indicate a fecal isolate. Significant differences from the SC5314 reference strain were determined by one-way ANOVA, with Dunnett's multiple correction testing. (B) Survival assays in *G. mellonella*, comparing the SC5314 reference to 6 isolates from donor 814. Each strain was standardized to  $2 \times 10^6$  cells/mL before inoculating 20 *G. mellonella* larvae per strain with 50  $\mu$ L of prepared inoculum. Larvae were monitored daily for survival. Statistical differences were determined using a Mantel–Cox log-rank test. Asterisks indicate  $P < 0.05$  (\*),  $P < 0.005$  (\*\*),  $P < 0.001$  (\*\*\*), and  $P < 0.0001$  (\*\*\*\*) compared with SC5314.

<https://doi.org/10.1371/journal.pbio.3001822.g006>

because *C. albicans* is a primarily clonal yeast that does not generally undergo meiosis [76,77]. Additionally, the genetic basis underlying the ability of commensal strains of *C. albicans* to transition to pathogenic behavior is not well defined. Moreover, *C. albicans* exhibits high rates of structural mutation and ploidy variation [43,78] as well as high heterozygosity [25]. Previous descriptions of population genomic variation in *C. albicans* have largely relied on short-read sequencing [25] or molecular typing methods [28] at relatively few loci to define genetic variation in the species. Here, we have greatly expanded the number of available long-read genomes for *C. albicans* and have generated a catalog of structural variants that was largely absent from previous descriptions of natural diversity. We were able to leverage clonal variation within a

single donor to identify variation within a host, potentially including variation that arose during colonization. Overall, we performed a systematic phenotypic analysis of commensal isolates from healthy donors, thus allowing us to examine *C. albicans* genotypic and phenotypic diversity.

Strikingly, we observed higher competitive fitness of the commensal strains during systemic competition assays, and when we tested 15 isolates for their ability to cause disease, we saw similar kinetics of mortality. This suggests that our commensal strains generally maintained their capacity to cause disease in our *Galleria* model of infection, despite SC5314 being able to cause more cell death in a macrophage model of infection for all but 7 commensal isolates. Importantly, other bloodstream isolates were also unable to cause host cell death. Additionally, our commensal strains were still able to filament in response to the inducing cues of 10% serum or high temperature, in contrast to work demonstrating that strains passaged through the mammalian gut resulted in a decrease in systemic virulence and defects during in vitro filamentation assays [36]. Potentially, the *Galleria* model does not recapitulate the disease process that would be caused by these isolates in a mammalian host. Although we observed significant variation in the ability of strains to invade into agar, recent intravital imaging approaches suggest that filamentation in response to serum matched that seen in vivo more than filamentation in response to solid Spider agar [79,80], although this may not recapitulate the environment seen in the gut. Potentially, the selective pressures that occur during mouse models of colonization may not recapitulate the selection that occurs during human colonization.

Although we were not able to associate a particular phenotype with increased systemic disease, we observed extensive phenotypic and genotypic variation between the commensal isolates. This variability is also consistent with recent work from clinical, disease-associated strains [19]. Moreover, our commensal strains were able to proliferate on a range of different environmental conditions, and we did not observe significant differences in phenotypes between isolates obtained from oral or fecal sites, except for a slight increase in invasion into Spider medium for oral samples. This work highlights the striking ability of *C. albicans* to adapt to a wide range of environmental conditions and indicates that colonization at a specific body site does not necessarily predict pathogenic potential.

Previous work has suggested that the *C. albicans* population within a given individual is clonal [71,81–84] and that the fungus is acquired during birth as a part of the normal microbiota [44]. In these studies, samples often came from patients with active disease; this may suggest that there is selection for the ability to cause disease, resulting in repeated isolation of representative samples of a clonal population. In contrast, we identified disparate individuals that appeared to be colonized by strains that were nearly identical, suggesting that there was some transmission between individuals. Whether these transmission events allow for long-term colonization, and how they affect the initial *C. albicans* colonizing strains, is still not fully understood. An important limitation of our study is that the undergraduate student donors were anonymous; therefore, we are unable to provide specific demographic information including age, sex, immune status, and recent antibiotic use that may allow us to expand our conclusions.

We also observed diversification within hosts, with multiple instances of closely related strains showing variation in phenotypes, such as invasion into agar. In one representative example, we were able to use comparative genome analysis coupled with co-expression, which we term “limited diversity exploitation,” to identify a candidate transcription factor that regulates invasion. Previous work on a *ZMS1* knockout strain of *C. albicans* did not show any differences in phenotype compared with the parent strain [68]. However, we observed that a single amino acid substitution in the predicted fungal transcription factor regulatory middle



homology region was sufficient to drive hyper-invasive growth. Moreover, this phenotype was distinct from the deletion phenotypes in these 2 genetic backgrounds, again contrasting with the SC5314 reference strain. It is likely that populations of colonizing *C. albicans* will vary in other important clinically relevant traits, including filamentation in macrophages or intrinsic drug tolerance and resistance, and our results suggest that studying more than a single representative isolate gives opportunities to discover new biology.

## Methods

### Human subjects and sample collection

Study volunteers were recruited through the Authentic Research Sections of the introductory biology laboratory course at the University of Michigan (BIO173). All participants provided written, informed consent for sample collection as well as isolation and characterization of microbes from feces. This study was approved by the Institutional Review Board of the University of Michigan Medical School (HUM00094242 and HUM00118951) and was conducted in compliance with the Helsinki Declaration.

Oral samples were obtained by self-administered cotton swabbing, and fecal samples were collected and diluted in PBS+DMSO before plating; in each case, BD ChromAgar was used to identify *C. albicans* colonies. Individual *C. albicans* colonies were picked from all plates with a toothpick into 100  $\mu$ L of YPD in a 96-well plate. The cultures were then inoculated onto BD Chromagar to differentiate between yeast species and ITS regions were amplified to confirm *C. albicans* strains. The *C. albicans* colonies were individually inoculated into fresh 96-well plate in YPD and after 24 h of incubation at 30°C, 50% glycerol was added to generate the stock plates. For each oral sample, individual *C. albicans* colonies were picked from the CHROMagar plates and incubated overnight in 100  $\mu$ L of YPD in a 96-well plate. After 24 h incubation at 30°C, 50% glycerol was added to the 96 wells to generate the stock plates. All strains were maintained in -80°C cryoculture.

### Reference strains

SC5314 reference strain was obtained from the gene replacement and conditional expression (GRACE) library [85]. CHN1 human isolate was obtained from the Huffnagle laboratory group [48].

### Media and culture conditions

Media for growth curves is described in [S1 Table](#).

### Growth curve analysis

Overnight cultures of each *C. albicans* isolate were grown in 200  $\mu$ L of YPD in 96-well plates at 30°C, then subcultured into fresh media using a sterile pinner. Growth curves were performed on a BioTek 800 TS Absorbance Reader in the indicated conditions for 24 h, without shaking. For the glycerol and galactose growth curves, the strains were incubated for 48 h due to the slower growth. Maximum growth rate and carrying capacity were determined using the GrowthcurveR analysis package [86]. A minimum of 2 biological replicates were performed per isolate per growth condition. All media and conditions are listed in [S1 Table](#). Summary statistics from GrowthcurveR for the entire collection of isolates are included in [S2 Table](#), raw data in [S1 Information](#). Summary statistics from GrowthcurveR for the condensed set in multiple environmental conditions are included in [S5 Table](#), raw data in [S1 Information](#).

### Filamentation analysis

Overnight cultures of each *C. albicans* isolate were grown in 100  $\mu$ L of YPD in 96-well plates at 30°C, then subcultured into 100  $\mu$ L of YPD at either 30°C or 42°C for 3 h. After 3 h, the plates were imaged on an Olympus IX80 microscope at 20 $\times$  magnification. To test the response to serum, overnight cultures were subcultured into 5 mL of YPD with 10% serum and rotated in a 37°C incubator for 3 h before vortexing and imaging on glass slides at 20 $\times$  magnification.

### Agar invasion methods

Isolates were taken from frozen glycerol stocks, incubated overnight in yeast extract peptone dextrose (YPD) in 96-well plates and inoculated onto solid YPD and Spider media using a 96-pin replicator tool (Singer Instruments). Plates were then incubated at the indicated temperature and oxygen conditions for 5 to 7 days. Colonies on plates were first imaged on a Biorad Gel Doc XR. Colonies were then gently washed off with deionized water. Plates were imaged again to capture invasion images. The invasiveness of each isolate was determined through a rubric scale of 0 to 5, as indicated in Fig 1. A 0 indicates no invasion, 1 to 2 indicates minimal invasion, a 3 indicates distinct circular hyphal invasion, while a 4 indicates an even larger circular hyphal invasion. A 5 indicates the most aggressive invasion, with a “halo” of more hyphae surrounding the initial growth. A minimum of 3 biological replicates was performed for each isolate and the median score was determined. A summary of invasion scores is in S2 Table, raw data in S1 Information.

### High molecular weight genomic DNA extraction and sequencing

Genomic DNA was extracted from the 45 commensal isolates, SC5314, and CHN1. High molecular weight (>20 kb fragment) genomic DNA was prepared using the Qiagen MagAttract HMW DNA Kit (Qiagen 67563), following the manufacturer’s protocol, with minor modification. Fungal cell walls were first digested using Zymolyase (Zymo E1005) before DNA extraction. TELL-seq [49] libraries were prepared at the University of Michigan Advanced Genomics facility. Libraries were sequenced on an Illumina NovaSeq SP 300 cycle flow cell. All sequences are uploaded to NCBI SRA at PRJNA875200.

### Sequence variation analysis

To identify SNVs across all published isolates of *C. albicans* and the newly sequenced commensal isolates, we mapped 435 read libraries to the SC5314 reference genome with BWA-MEM [50]. We used various samtools utilities to convert alignment files, remove PCR duplicates, and assign read groups [50]. Variants were called with GATK HaplotypeCaller and individual sample GVCFs were combined with GATK CombineGVCFs. Finally, we genotyped our combined GVCF across samples and loci with GATK GenotypeGVCFs [87]. GATK identified 851,355 total variable loci. Of the 741,027 diallelic loci, GATK identified 60,626 indel-derived alleles (8.18%) and 680,401 SNV alleles (91.82%). To refine our raw variant set to only include high-confidence SNVs, we removed low-quality variant calls at the (i) individual genotype call; (ii) whole locus; and (iii) whole sample levels. First, we removed individual sample genotype calls if their genotype quality (i.e., GQ) was <0.99 (out of 1.0) or the results of the map quality rank sum test (i.e., MQRS) differed from 0.0 (i.e., the perfect score). Second, we removed all variant calls with more than 1 alternative allele (i.e., not diallelic), an alternate allele longer than 1 nucleotide (i.e., indels), or with >5% missing genotype calls across samples. Finally, we removed entire samples from the combined VCF if their genotype calls across all loci were more than 90% missing, which resulted in the removal of 3 entire samples:

SRR6669899, SRR6669970, and SRR1103579. Filtering our variant set in this way yielded a final set of 112,136 high-quality SNVs across 431 remaining samples. Of these, 90,675 (80.86%) were represented in at least 1 member of our set of newly sequenced strains.

To remove redundancy and focus on natural *C. albicans* diversification, we removed additional samples corresponding to resequenced strains (e.g., multiple SC5314 samples present in full data set) and those sequenced as part of experimental evolution studies (i.e., [36,52,53]). Following removal of these samples, we were left with 324 sample SNV profiles (280 published and 44 new genome strains) (S3 Table). To generate an input matrix for distance calculation and NJ clustering, we coded genotypes as homozygous for the reference allele (I), heterozygous (H), homozygous for the alternative allele (A), or missing (N). Raw distance between 2 sample genotypes at a particular locus (i.e.  $GT_i$  and  $GT_j$ ) were calculated as 0.5 per alternative allele, with respect to the SC5314 reference such that  $D(H, A) = 0.5$ ,  $D(R, A) = 1.0$ , etc. Missing genotypes (i.e., coded as N) were ignored. Using these raw distances, we calculated the pairwise distance between all sample pairs across all 112,136 loci according to the generalized distance function for sample  $i$  and sample  $j$ .

$$D_{ij} = \frac{1}{nS - P_s - nN} \sum_{x=1}^{nSNPs} D(GT_{ix}, GT_{jx})$$

Where  $nSNPs$  is the total number of high-quality SNPs in our data set (i.e., 112,136),  $nN$  is the number of uncalled genotypes between each pair (i.e., coded as N),  $GT_{ix}$  is the genotype of sample  $i$  at locus  $x$ , and  $GT_{jx}$  is genotype of sample  $j$  at locus  $x$ . The resulting distance matrix was clustered with the  $nj$  function in phytools in R [88]. We visualized the resulting tree structure and associated data with ggtree in R [89–91].

## Bootstrapping

One hundred bootstrap replicates were generated by randomly sampling rows from the SNP table with replacement to yield a table with the same dimensions as the original SNP table. Distance matrices and neighboring-joining trees were generated for bootstrap replicates identically to the original SNP table. Bootstrap support values were computed based on the bootstrap replicate trees and annotated onto the original SNP tree with IQTree2.

## Structural variant analysis from TELL-seq

To identify structural variation genome wide in the condensed set of *C. albicans* isolates, we mapped the read libraries to the SC5314 reference genome with BWA-MEM [50]. We visualized the read depth across the genome by breaking the reference genome into 500 bp bins. Each read was assigned to a single bin based on the first coordinate in the reference genome to which the read aligned. The read counts were normalized across the samples by the average read coverage for each sample. The value for each bin ( $v_{b,s}$ ) for each sample was calculated using the following formula:

$$v_{b,s} = \frac{\text{bin read count} * \text{read length}}{\text{bin size} * \text{sample coverage}}$$

Additionally, we performed structural variant calling using lumpy-sv [92] and genotyping with svtyper [93] using smooove (version 0.2.5). We visualized the variant calls by breaking the reference genome into 500 bp bins and identified the number of variant calls of a give type that overlapped with each bin.

To identify potential translocations from the TELL-seq data, we used the Tell-Link pipeline to assemble the read libraries into contigs and aligned the assembled contigs to the SC5314

reference genome using MiniMap2 [94]. We identified candidate junctions where consecutive aligned segments longer than 10,000 bp on the same contig aligned to different chromosomes. To experimentally validate the predicted junctions, we designed primers from uniquely mapped contig segments on either side of the junction, i.e., one from each of the 2 chromosomes predicted to be joined together. Primers were selected from the 500 bp directly upstream and downstream from the predicted junction location unless the junction was directly flanked by non-uniquely mapping segments in which case, the 500 bp of the closest uniquely mapping segments were used. Primers are in [S6 Table](#).

### CHEF gel electrophoresis

Agarose plugs for CHEF gel electrophoresis were prepared as in Selmecki 2005 [95]. Electrophoresis was performed as in Chibana [96], with minor modifications to improve chromosome separation. Plugs were run on 0.9% megabase agarose (Biorad 1613108), with run conditions of 60 to 300 s, 4.5 V/cm, 120 angle for 36 h, followed by 720 to 900 s, 2.0 V/cm, 106 for 24 h. Sizes were determined based on approximations from Selmecki and a CHEF DNA Size Marker, 0.2–2.2 Mb, *S. cerevisiae* Ladder (Biorad 1703605).

### Biofilm formation

Biofilm formation was assessed as previously described [58]. Briefly, overnights of *C. albicans* isolates were incubated and diluted to 0.5 OD<sub>600</sub>. In a 96-well plate, isolates were added to 200  $\mu$ L of RPMI media, covered with a Breathe-Easy film, and incubated at 37°C shaking at 250 rpm. After 90 min, RPMI media was removed and wells were washed once with 200  $\mu$ L of 1 $\times$  PBS and 200  $\mu$ L of fresh RPMI media was added. Plates were covered in a Breathe-Easy film and left to incubate at 37°C with shaking at 250 rpm for 24 h before a final read on the plate reader at OD<sub>600</sub>.

### Mammalian cell culture

J2-iBMDM cells were isolated from the bone marrows of C57BL/6J mice and differentiated in BMDM medium (50% DMEM, 2 mM l-glutamine, 1 mM sodium pyruvate, 30% L929-conditioned medium, 20% heat-inactivated fetal bovine serum [FBS; Invitrogen], 55  $\mu$ M 2-mercaptoethanol, and Pen/Strep), then immortalized using Cre-J2 viral supernatants.

### Phagocytosis

Phagocytosis assays were performed as previously described [60]. Briefly, iBMDM macrophages were prepared for infection in RPMI media containing 3% FBS, diluted to  $3 \times 10^6$  cells/mL and incubated overnight at 100  $\mu$ L/well in a 96-well plate. Overnight cultures of *C. albicans* isolates were incubated and diluted to  $4 \times 10^6$  cells/mL into RPMI media with 3% FBS and used to infect macrophages. Inoculated plates were centrifuged for 1 min at 1,000 rpm to synchronize. After 30 min, media was removed, and cells were fixed with 4% paraformaldehyde (PFA) for 10 min. The cells were washed 3 times in 1 $\times$  PBS. Cells were then stained with 50  $\mu$ L of FITC-Concanavalin (5  $\mu$ g/mL Sigma-Aldrich C7642) at room temperature, rocking for 30 min, wrapped in foil. The plates were then washed 3 $\times$  with 100  $\mu$ L of 1 $\times$  PBS, and then 50  $\mu$ L of 0.05% Triton-X100 was added to permeabilize the cells. Cells were then washed 3 $\times$  with 1 $\times$  PBS and a final stain of 50  $\mu$ L of calcofluor white (100  $\mu$ g/mL, Sigma-Aldrich C7642) was added to cells to incubate for 10 min. The cells were then washed 3 $\times$  with 100  $\mu$ L of 1 $\times$  PBS and then maintained in 100  $\mu$ L of 1 $\times$  PBS at 4°C before imaging on the microscope at 20 $\times$  magnification using the DIC, FITC, and DAPI channels. Images were analyzed with a

CellProfiler pipeline to determine the percentage of internalized *Candida*. The total number of *Candida* was determined through calcofluor white staining. Next, the number of external *Candida* was determined through FITC-ConA staining.  $(1 - \text{external cells})/\text{total cells} = \% \text{ of internalized } Candida$ .

### Macrophage filamentation assay

To assess filamentation of isolates in macrophages, iBMDM macrophages were prepared for infection in RPMI media containing 3% FBS, diluted to  $3 \times 10^6$  cells/mL and incubated overnight at 100  $\mu\text{L}/\text{well}$  in a 96-well plate. Overnights of *C. albicans* isolates were incubated and diluted to  $4 \times 10^6$  cells/mL into RPMI media with 3% FBS and used to infect macrophages at an MOI of approximately 1:1. The inoculated plate was centrifuged for 1 min at 1,000 rpm to synchronize phagocytosis. After 2 h, the media was removed, and cells were fixed with 4% PFA for 10 min, permeabilized with 0.05% Triton-X100, and stained with 50  $\mu\text{L}$  of calcofluor white (100  $\mu\text{g}/\text{mL}$ , Sigma-Aldrich C7642) before imaging at 40 $\times$  magnification using the DIC and DAPI channels. The extent of filamentation for each isolate was scored on a rubric from 0 to 5. A score of 0 indicates an isolate that exhibited only yeast morphology with no filamentation during macrophage infection. A score of 1 indicates an isolate that remains primarily in the yeast form, with some hyphae or pseudohyphae, while a score of 2 indicates an isolate that remains primarily in the yeast form, with more hyphae and pseudohyphae present than a score of 1. A score of 3 indicates an isolate that had approximately equal numbers of yeast and true hyphae. A score of 4 indicates an isolate with more true hyphae than yeast during infection, and a score of 5 indicates an isolate in which the majority of cells formed hyphae with very few yeast remaining. A minimum of 3 biological replicates was performed for each isolate and the median score was determined. A summary of macrophage filamentation scores is in [S5 Table](#), raw data in [S1 Information](#).

### Cell death assay

To assess macrophage cell death, J2-iBMDM macrophages were prepared for infection in RPMI media containing 3% FBS, diluted to  $4 \times 10^5$  cells/mL and incubated overnight at 100  $\mu\text{L}/\text{well}$  in a 96-well plate. Macrophages were primed with 200 ng/mL LPS in RPMI media with 3% FBS for 2 h. Overnights of *C. albicans* isolates were incubated and diluted to  $5 \times 10^5$  cells/mL into RPMI media with 3% FBS and used to infect macrophages at an MOI of approximately 1:1. The plate was centrifuged for 1 min at 1,000 rpm to synchronize phagocytosis. After 4 h, the media was removed and cells were stained with Hoechst (20  $\mu\text{M}$ , Cayman Chemical 33342) to mark nuclei and propidium iodide (1  $\mu\text{g}/\text{mL}$ , Acros Organics 440300250) to mark cell death. Plates were then imaged on an Cellomics ArrayScanVTI Objective Module at 20 $\times$  with at least 5 images taken per well. Images were analyzed with a CellProfiler pipeline (available by request) to determine the percent of dead macrophages. Macrophages were identified through Hoescht nuclear staining to determine the total number of cells in each field of view. Next, cell death events were determined based on propidium iodide staining.  $(\# \text{ of propidium iodide events})/(\# \text{ of total cells}) = \% \text{ of dead macrophages}$ .

### ZMS1 strain construction

To make the *ZMS1* allele swap strains, we used the NEBuilder HiFi assembly kit to clone the endogenous *ZMS1S681T* allele from the 814–183 strain background and the *ZMS1T681S* allele from the 814–168 strain background into the pUC19 cloning vector, along with the nourseothricin resistance cassette, to generate plasmids pTO274 and pTO192. *ZMS1* and 500 bp of putative terminator was amplified using oTO771+oTO734. NAT was amplified using oTO18

+ oTO735. pUC19 was amplified using oTO590 + oTO591. The plasmids were Sanger sequenced to confirm the presence of the specific *ZMS1* allele and the absence of additional mutations using oTO736. These plasmids were transformed into the 814–168 and 814–183 strain backgrounds using a transient CRISPR approach [97]. NAT-resistant transformants were tested for the presence of specific *ZMS1* alleles by Sanger sequencing using oTO736.

To generate the deletion strains, the *ZMS1::NAT* cassette was amplified from the NAT flipper plasmid [98] using primers oTO1215 and oTO1216, and transformed into the 814–168 and 814–183 strain backgrounds using a transient CRISPR approach [97]. Integration was tested using oTO5 and oTO1218 and loss of the wild-type *ZMS1* gene was tested using oTO736 and oTO1218.

### ***Galleria mellonella* infections**

Each isolate was examined for competitive fitness in a one-to-one ratio with a fluorescent wild-type SC5314 *C. albicans* strain in *G. mellonella* larvae. Infections were performed as previously described [98]. Briefly, *G. mellonella* larvae were purchased from [speedyworm.com](https://www.speedyworm.com) and maintained in sawdust at room temperature. Overnights were prepared for each isolate and wild-type strain in YPD at 30°C, with rotation.

To measure competitive fitness, a 1:1 ratio of fluorescent wild type to unlabeled isolate was prepared in 1× PBS at  $5 \times 10^5$  cells/mL. 10 larvae/strain were randomly chosen and infected via the last right proleg with 50 µL of the 1:1 inoculum using an exel veterinary U-40 diabetic syringe (0.5CC × 29G × ½). After injection, larvae were maintained at room temperature for 3 days before harvesting using a Benchmark D1000 homogenizer. Larvae were homogenized in 0.5 mL of 1× PBS, diluted 1:10 in PBS, and plated onto YPD plates containing gentamicin, ampicillin, and ciprofloxacin. Plates were left at 30°C for 48 h before imaging on a Typhoon FLA 9500 biomolecular Imager. The ratio of fluorescent to unmarked strains was compared with the inoculum to determine competitive index.

To measure virulence, 20 *G. mellonella* larvae/strain were infected with 50 µL of  $2 \times 10^6$  cells/mL inoculum using an exel veterinary U-40 diabetic syringe (0.5CC × 29G × ½). After injection, larvae were maintained at room temperature and monitored daily for survival. Virulence was analyzed using Kaplan–Meier survival curves in GraphPad Prism (version 9).

## **Supporting information**

### **S1 Table. Media and growth conditions.**

(XLSX)

### **S2 Table. Full collection growth curves (Tab 1) and invasion scores (Tab 2).**

(XLSX)

### **S3 Table. SRA accession numbers for strains included in phylogenetic analysis.**

(XLSX)

### **S4 Table. Genomic sequence surrounding putative fusion events.**

(XLSX)

### **S5 Table. Condensed set strain information and quantitative phenotypes.**

(XLSX)

### **S6 Table. Primers used in this study.**

(XLSX)

**S1 Fig. Comparisons of growth rate between samples and nutrient source.** (A) Enrichment of fecal vs. oral samples by carrying capacity. (B) Correlation between growth rates in different carbon sources. (C) UMAP plot for strain phenotypic similarity. All growth conditions and invasion phenotypes were nonlinearly projected into 2D space as in Fig 1F, but this time colored by sample site.

(TIF)

**S2 Fig. Cladogram of the SNP tree (branch lengths all equal to 1) that has the backbone colored by bootstrap support values.** One hundred bootstrap replicates were generated by randomly sampling rows from the SNP table with replacement to yield a table with the same dimensions as the original SNP table. Distance matrices and neighboring-joining trees were generated for bootstrap replicates identically to the original SNP table. Bootstrap support values were computed based on the bootstrap replicate trees and annotated onto the original SNP tree with IQTree2. Bloodstream isolates used in Fig 4 are indicated with red text.

(TIF)

**S3 Fig. Structural variation.** (A) CHEF karyotype gels were performed on the condensed set of *C. albicans* isolates and visualized using ethidium bromide staining. Red arrows indicate large chromosome banding patterns that differ from the SC5314 reference strain. Chromosome 5 showed especially extensive variability in size between isolates. (B) Heatmaps of read coverage across the condensed set of isolates for each chromosome. Each column represents a 500 bp bin of the reference genome and each row is an isolate from the condensed set. Values greater than 1 (red) suggest potential duplications while values less than 1 (blue) suggest potential deletions. Each row represents an isolate, with SC5314 represented by the top row. The rows are ordered, top to bottom, in the phylogenetic order used in Fig 3. (C) Heatmaps of Lumpy Structural variation calls for the condensed set of isolates. Each column represents a 500 bp bin of the reference genome and each row of the heat map for a given variant class is an isolate from the condensed set. The value for each bin indicates the number of SV calls overlapping the 500 bp window.

(TIF)

**S4 Fig. Biofilm formation.** Each condensed set isolate was tested for its ability to form a biofilm on a plastic surface. Isolate labels were colored based on the nearest defined cluster from Fig 2 and ordered based on phylogeny. Black data points indicate an oral isolate and pink data points indicate a fecal isolate. Asterisks indicate  $P < 0.05$  (\*) compared with SC5314, one-way ANOVA compared with SC5314, with Dunnett's post hoc test for multiple corrections.

(TIF)

**S5 Fig. Comparison between filamentation in macrophages and agar invasion.** (A) Images of *C. albicans* condensed set isolates after growth in macrophages for 4 h. *C. albicans* were stained with calcofluor white. Images taken using DIC and DAPI channels at 20× magnification. Scale = 50 μM. (B) Histogram of the distribution of macrophage filamentation scores of the condensed set. (C) Colony morphology and agar invasion for *C. albicans* condensed set isolates on YPD agar at 30°C and Spider agar at 37°C after 5 days.

(TIF)

**S6 Fig. UMAP embedding.** Nonlinear embedding dimensionality reduction was performed on the phenotypic data on the condensed set of isolates using UMAP. Clusters did not segregate by sample site, donor, or clade. Data used to generate the UMAP are included in S5 Table, raw data in S1 Information. Clade information is indicated by color. Pink or black outlines

indicate fecal of oral isolate, respectively.  
(TIF)

**S7 Fig. Virulence in *G. mellonella*.** (A) Survival assays in *G. mellonella*, comparing the SC5314 reference to 4 isolates from donors 882 and 811. (B) Survival assays in *G. mellonella*, comparing the SC5314 reference to 5 isolates from donors 833 and 838. Each strain was standardized to  $2 \times 10^6$  cells/mL before inoculating 20 *G. mellonella* larvae per strain with 50  $\mu$ L of prepared inoculum. Larvae were monitored daily for survival. Statistical differences were determined using a Mantel–Cox log-rank test. \*\* indicates  $P$ -value  $< 0.01$ , \* indicates  $P$ -value  $< 0.05$ .

(TIF)

**S1 Information. Raw data underlying figures.**

(XLSX)

**S1 Raw Images. Uncropped blots for Figs 3 and S3.**

(PDF)

## Acknowledgments

We thank members of the O’Meara lab for critical reading of the manuscript, the Advanced Genomics Core at the University of Michigan for assistance with TellSeq, and the University of Michigan Great Lakes research computing cluster.

## Author Contributions

**Conceptualization:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Andrea Hodgins-Davis, Timothy Y. James, Teresa R. O’Meara.

**Data curation:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Andrea Hodgins-Davis, Alexandra M. Weber, Matthew J. O’Meara, Teresa R. O’Meara.

**Formal analysis:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Andrea Hodgins-Davis, Alexandra M. Weber, Ryan E. Mills, Matthew J. O’Meara, Timothy Y. James, Teresa R. O’Meara.

**Funding acquisition:** Timothy Y. James, Teresa R. O’Meara.

**Investigation:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Andrea Hodgins-Davis, Alexandra M. Weber, Katura M. Metzner, Michael J. McFadden, Teresa R. O’Meara.

**Methodology:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Andrea Hodgins-Davis, Alexandra M. Weber, Katura M. Metzner, Michael J. McFadden, Ryan E. Mills, Matthew J. O’Meara, Timothy Y. James, Teresa R. O’Meara.

**Project administration:** Teresa R. O’Meara.

**Resources:** Teresa R. O’Meara.

**Software:** Kevin R. Amses, Alexandra M. Weber, Ryan E. Mills, Matthew J. O’Meara.

**Supervision:** Ryan E. Mills, Timothy Y. James, Teresa R. O’Meara.

**Validation:** Faith M. Anderson, Teresa R. O’Meara.

**Visualization:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Andrea Hodgins-Davis, Alexandra M. Weber, Matthew J. O’Meara, Teresa R. O’Meara.



**Writing – original draft:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Matthew J. O’Meara, Teresa R. O’Meara.

**Writing – review & editing:** Faith M. Anderson, Noelle D. Visser, Kevin R. Amses, Andrea Hodgins-Davis, Alexandra M. Weber, Katura M. Metzner, Michael J. McFadden, Ryan E. Mills, Matthew J. O’Meara, Timothy Y. James, Teresa R. O’Meara.

## References

1. Mukherjee PK, Sendid B, Hoarau G, Colombel J-F, Poulain D, Ghannoum MA. Mycobiota in gastrointestinal diseases. *Nat Rev Gastroenterol Hepatol*. 2015; 12:77–87. <https://doi.org/10.1038/nrgastro.2014.188> PMID: 25385227
2. Polvi EJ, Li X, O’Meara TR, Leach MD, Cowen LE. Opportunistic yeast pathogens: reservoirs, virulence mechanisms, and therapeutic strategies. *Cell Mol Life Sci*. 2015; 72:2261–2287. <https://doi.org/10.1007/s00018-015-1860-z> PMID: 25700837
3. Li J, Chen D, Yu B, He J, Zheng P, Mao X, et al. Fungi in Gastrointestinal Tracts of Human and Mice: from Community to Functions. *Microb Ecol*. 2018; 75:821–829. <https://doi.org/10.1007/s00248-017-1105-9> PMID: 29110065
4. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS ONE*. 2013; 8:e66019. <https://doi.org/10.1371/journal.pone.0066019> PMID: 23799070
5. Gunsalus KTW, Tornberg-Belanger SN, Matthan NR, Lichtenstein AH, Kumamoto CA. Manipulation of Host Diet To Reduce Gastrointestinal Colonization by the Opportunistic Pathogen *Candida albicans*. *mSphere*. 2016; 1. <https://doi.org/10.1128/mSphere.00020-15> PMID: 27303684
6. Noble SM, Gianetti BA, Witchley JN. *Candida albicans* cell-type switching and functional plasticity in the mammalian host. *Nat Rev Microbiol*. 2017; 15:96–108. <https://doi.org/10.1038/nrmicro.2016.157> PMID: 27867199
7. Limon JJ, Skalski JH, Underhill DM. Commensal Fungi in Health and Disease. *Cell Host Microbe*. 2017; 22:156–165. <https://doi.org/10.1016/j.chom.2017.07.002> PMID: 28799901
8. Alves R, Barata-Antunes C, Casal M, Brown AJP, Van Dijck P, Paiva S. Adapting to survive: How *Candida* overcomes host-imposed constraints during human colonization. *PLoS Pathog*. 2020; 16:e1008478. <https://doi.org/10.1371/journal.ppat.1008478> PMID: 32437438
9. Ene IV, Brunke S, Brown AJP, Hube B. Metabolism in fungal pathogenesis. *Cold Spring Harb Perspect Med*. 2014; 4:a019695. <https://doi.org/10.1101/cshperspect.a019695> PMID: 25190251
10. Polke M, Hube B, Jacobsen ID. *Candida* survival strategies. *Adv Appl Microbiol*. 2015; 91:139–235. <https://doi.org/10.1016/bs.aambs.2014.12.002> PMID: 25911234
11. Zhai B, Ola M, Rolling T, Tosini NL, Joshowitz S, Littmann ER, et al. High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. *Nat Med*. 2020. <https://doi.org/10.1038/s41591-019-0709-7> PMID: 31907459
12. Yan L, Yang C, Tang J. Disruption of the intestinal mucosal barrier in *Candida albicans* infections. *Microbiol Res*. 2013; 168:389–395. <https://doi.org/10.1016/j.micres.2013.02.008> PMID: 23545353
13. Alenazy H, Alghamdi A, Pinto R, Daneman N. *Candida* colonization as a predictor of invasive candidiasis in non-neutropenic ICU patients with sepsis: A systematic review and meta-analysis. *Int J Infect Dis*. 2021; 102:357–362. <https://doi.org/10.1016/j.ijid.2020.10.092> PMID: 33157294
14. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. *Sci Transl Med*. 2012; 4:165rv13–165rv13. <https://doi.org/10.1126/scitranslmed.3004404> PMID: 23253612
15. Magill SS O’Leary E, Janelle SJ, Thompson DL, Dumyati G, Nadle J, et al. Changes in prevalence of health care—associated infections in US hospitals. *N Engl J Med*. 2018; 379:1732–1744.
16. Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, Johnston M, et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proc Natl Acad Sci U S A*. 2011; 108:14539–14544. <https://doi.org/10.1073/pnas.1105430108> PMID: 21873232
17. Hittinger CT. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet*. 2013; 29:309–317. <https://doi.org/10.1016/j.tig.2013.01.002> PMID: 23395329
18. Huang MY, Woolford CA, May G, McManus CJ, Mitchell AP. Circuit diversification in a biofilm regulatory network. *PLoS Pathog*. 2019; 15:e1007787. <https://doi.org/10.1371/journal.ppat.1007787> PMID: 31116789

19. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, et al. Genetic and phenotypic intra-species variation in *Candida albicans*. *Genome Res.* 2015; 25:413–425. <https://doi.org/10.1101/gr.174623.114> PMID: 25504520
20. Wang JM, Woodruff AL, Dunn MJ, Fillinger RJ, Bennett RJ, Anderson MZ. Intraspecies Transcriptional Profiling Reveals Key Regulators of *Candida albicans* Pathogenic Traits. *MBio.* 2021;12. <https://doi.org/10.1128/mBio.00586-21> PMID: 33879584
21. McDonough L, Mishra AA, Tosini N, Kakade P, Penumutthu S, Liang S-H, et al. *Candida albicans* Isolates 529L and CHN1 Exhibit Stable Colonization of the Murine Gastrointestinal Tract. *bioRxiv.* 2021. <https://doi.org/10.1128/mBio.02878-21> PMID: 34724818
22. Soll DR, Pujol C. *Candida albicans* clades. *FEMS Immunol Med Microbiol.* 2003; 39:1–7. [https://doi.org/10.1016/S0928-8244\(03\)00242-6](https://doi.org/10.1016/S0928-8244(03)00242-6) PMID: 14556989
23. Blignaut E, Pujol C, Lockhart S, Joly S, Soll DR. Ca3 fingerprinting of *Candida albicans* isolates from human immunodeficiency virus-positive and healthy individuals reveals a new clade in South Africa. *J Clin Microbiol.* 2002; 40:826–836. <https://doi.org/10.1128/JCM.40.3.826-836.2002> PMID: 11880401
24. Odds FC. Molecular phylogenetics and epidemiology of *Candida albicans*. *Future Microbiol.* 2010:67–79. <https://doi.org/10.2217/fmb.09.113> PMID: 20020830
25. Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, Sertour N, et al. Gene flow contributes to diversification of the major fungal pathogen *Candida albicans*. *Nat Commun.* 2018; 9:2253. <https://doi.org/10.1038/s41467-018-04787-4> PMID: 29884848
26. MacCallum DM, Castillo L, Nather K, Munro CA, Brown AJP, Gow NAR, et al. Property differences among the four major *Candida albicans* strain clades. *Eukaryot Cell.* 2009; 8:373–387. <https://doi.org/10.1128/EC.00387-08> PMID: 19151328
27. Bournoux M-E, Aanensen DM, Morand S, Théraud M, Spratt BG, d'Enfert C. Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. *Infect Genet Evol.* 2004; 4:243–252. <https://doi.org/10.1016/j.meegid.2004.06.002> PMID: 15450203
28. Odds FC, Bournoux M-E, Shaw DJ, Bain JM, Davidson AD, Diogo D, et al. Molecular phylogenetics of *Candida albicans*. *Eukaryot Cell.* 2007; 6:1041–1052. <https://doi.org/10.1128/EC.00041-07> PMID: 17416899
29. Tavanti A, Davidson AD, Fordyce MJ, Gow NAR, Maiden MCJ, Odds FC. Population Structure and Properties of *Candida albicans*, as Determined by Multilocus Sequence Typing. *J Clin Microbiol.* 2005; 43:5601–5613. <https://doi.org/10.1128/JCM.43.11.5601-5613.2005> PMID: 16272493
30. Li X, Yan Z, Xu J. Quantitative variation of biofilms among strains in natural populations of *Candida albicans*. *Microbiology.* 2003; 149:353–362. <https://doi.org/10.1099/mic.0.25932-0> PMID: 12624197
31. Slutsky B, Staebell M, Anderson J, Risen L. “White-opaque transition”: a second high-frequency switching system in *Candida albicans*. *J Bacteriol.* 1987. [papers3://publication/uuid/02E4C532-97D3-4987-B50F-D8EAF9F2A6CF](https://doi.org/10.1128/jb.169.1.189-197.1987) <https://doi.org/10.1128/jb.169.1.189-197.1987> PMID: 3539914
32. Wu W, Lockhart SR, Pujol C, Srikantha T, Soll DR. Heterozygosity of genes on the sex chromosome regulates *Candida albicans* virulence. *Mol Microbiol.* 2007; 64:1587–1604. <https://doi.org/10.1111/j.1365-2958.2007.05759.x> PMID: 17555440
33. Calderone RA, Fonzi WA. Virulence factors of *Candida albicans*. *Trends Microbiol.* 2001; 9:327–335. [https://doi.org/10.1016/s0966-842x\(01\)02094-7](https://doi.org/10.1016/s0966-842x(01)02094-7) PMID: 11435107
34. Pappas PG, Lionakis MS, Arendrup MC, Ostrosky-Zeichner L, Kullberg BJ. Invasive candidiasis. *Nat Rev Dis Primers.* 2018; 4:18026. <https://doi.org/10.1038/nrdp.2018.26> PMID: 29749387
35. Dunn MJ, Fillinger RJ, Anderson LM, Anderson MZ. Automated quantification of *Candida albicans* biofilm-related phenotypes reveals additive contributions to biofilm production. *NPJ Biofilms Microbiomes.* 2020; 6:36. <https://doi.org/10.1038/s41522-020-00149-5> PMID: 33037223
36. Tso GHW, Reales-Calderon JA, Tan ASM, Sem X, Le GTT, Tan TG, et al. Experimental evolution of a fungal pathogen into a gut symbiont. *Science.* 2018; 362:589–595. <https://doi.org/10.1126/science.aat0537> PMID: 30385579
37. Witchley JN, Penumetcha P, Abon NV, Woolford CA, Mitchell AP, Noble SM. *Candida albicans* Morphogenesis Programs Control the Balance between Gut Commensalism and Invasive Infection. *Cell Host Microbe.* 2019; 25:432–443.e6. <https://doi.org/10.1016/j.chom.2019.02.008> PMID: 30870623
38. Pierce JV, Kumamoto CA. Variation in *Candida albicans* EFG1 expression enables host-dependent changes in colonizing fungal populations. *MBio.* 2012; 3:e00117–e00112. <https://doi.org/10.1128/mBio.00117-12> PMID: 22829676
39. Liang S-H, Anderson MZ, Hirakawa MP, Wang JM, Frazer C, Alaalm LM, et al. Hemizygosity Enables a Mutational Transition Governing Fungal Virulence and Commensalism. *Cell Host Microbe.* 2019; 25:418–431.e6. <https://doi.org/10.1016/j.chom.2019.01.005> PMID: 30824263

40. Witchley JN, Basso P, Brimacombe CA, Abon NV, Noble SM. Recording of DNA-binding events reveals the importance of a repurposed *Candida albicans* regulatory network for gut commensalism. *Cell Host Microbe*. 2021. <https://doi.org/10.1016/j.chom.2021.03.019> PMID: 33915113
41. White SJ, Rosenbach A, Lephart P, Nguyen D, Benjamin A, Tzipori S, et al. Self-regulation of *Candida albicans* population size during GI colonization. *PLoS Pathog*. 2007; 3:e184–e113. <https://doi.org/10.1371/journal.ppat.0030184> PMID: 18069889
42. Pande K, Chen C, Noble SM. Passage through the mammalian gut triggers a phenotypic switch that promotes *Candida albicans* commensalism. *Nat Genet*. 2013; 45:1088–1091. <https://doi.org/10.1038/ng.2710> PMID: 23892606
43. Forche A, Solis NV, Swidergall M, Thomas R, Guyer A, Beach A, et al. Selection of *Candida albicans* trisomy during oropharyngeal infection results in a commensal-like phenotype. *PLoS Genet*. 2019; 15:e1008137. <https://doi.org/10.1371/journal.pgen.1008137> PMID: 31091232
44. Schönherr FA, Sparber F, Kirchner FR, Guiducci E, Trautwein-Weidner K, Gladiator A, et al. The intra-species diversity of *C. albicans* triggers qualitatively and temporally distinct host responses that determine the balance between commensalism and pathogenicity. *Mucosal Immunol*. 2017; 10:1335–1350. <https://doi.org/10.1038/mi.2017.2> PMID: 28176789
45. Kim SH, Clark ST, Surendra A, Copeland JK, Wang PW, Ammar R, et al. Global Analysis of the Fungal Microbiome in Cystic Fibrosis Patients Reveals Loss of Function of the Transcriptional Repressor Nrg1 as a Mechanism of Pathogen Adaptation. *PLoS Pathog*. 2015; 11:e1005308. <https://doi.org/10.1371/journal.ppat.1005308> PMID: 26588216
46. Tan CT, Xu X, Qiao Y, Wang Y. A peptidoglycan storm caused by  $\beta$ -lactam antibiotic's action on host microbiota drives *Candida albicans* infection. *Nat Commun*. 2021; 12:2560.
47. Azadmanesh J, Gowen AM, Creger PE, Schafer ND, Blankenship JR. Filamentation Involves Two Overlapping, but Distinct. Programs of Filamentation in the Pathogenic Fungus *Candida albicans*. *G3*. 2017; 7:3797–3808.
48. Garg S, Ranjan P, Erb-Downward JR, Huffnagle GB. High-Quality Genome Reconstruction of *Candida albicans* CHN1 Using Nanopore and Illumina Sequencing and Hybrid Assembly. *Microbiol Resour Announc*. 2021; 10:e0029921. <https://doi.org/10.1128/MRA.00299-21> PMID: 34110235
49. Chen Z, Pham L, Wu T-C, Mo G, Xia Y, Chang PL, et al. Ultralow-input single-tube linked-read library method enables short-read second-generation sequencing systems to routinely generate highly accurate and economical long-range sequencing information. *Genome Res*. 2020; 30:898–909. <https://doi.org/10.1101/gr.260380.119> PMID: 32540955
50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. *Bioinformatics*. 2009; 25:2078–2079.
51. Sitterlé E, Coste AT, Obadia T, Maufrais C, Chauvel M, Sertour N, et al. Large-scale genome mining allows identification of neutral polymorphisms and novel resistance mutations in genes involved in *Candida albicans* resistance to azoles and echinocandins. *J Antimicrob Chemother*. 2020. <https://doi.org/10.1093/jac/dkz537> PMID: 31923309
52. Todd RT, Wikoff TD, Forche A, Selmecki A. Genome plasticity in *Candida albicans* is driven by long repeat sequences. *Elife*. 2019;8. <https://doi.org/10.7554/eLife.45954> PMID: 31172944
53. Ene IV, Farrer RA, Hirakawa MP, Agwamba K, Cuomo CA, Bennett RJ. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. *Proc Natl Acad Sci U S A*. 2018; 115:E8688–E8697. <https://doi.org/10.1073/pnas.1806002115> PMID: 30150418
54. Pittet D, Monod M, Filthuth I, Frenk E, Suter PM, Auckenthaler R. Contour-clamped homogeneous electric field gel electrophoresis as a powerful epidemiologic tool in yeast infections. *Am J Med*. 1991; 91:256S–263S. [https://doi.org/10.1016/0002-9343\(91\)90378-b](https://doi.org/10.1016/0002-9343(91)90378-b) PMID: 1928173
55. Forche A, Cromie G, Gerstein AC, Solis NV, Pisithkul T, Srida W, et al. Rapid Phenotypic and Genotypic Diversification After Exposure to the Oral Host Niche in *Candida albicans*. *Genetics*. 2018; 209:725–741. <https://doi.org/10.1534/genetics.118.301019> PMID: 29724862
56. Smith AC, Hickman MA. Host-Induced Genome Instability Rapidly Generates Phenotypic Variation across *Candida albicans* Strains and Ploidy States. *mSphere*. 2020;5. <https://doi.org/10.1128/mSphere.00433-20> PMID: 32493724
57. Panthee S, Hamamoto H, Ishijima SA, Paudel A, Sekimizu K. Utilization of Hybrid Assembly Approach to Determine the Genome of an Opportunistic Pathogenic Fungus, *Candida albicans* TIMM 1768. *Genome Biol Evol*. 2018; 10:2017–2022. <https://doi.org/10.1093/gbe/evy166> PMID: 30059981
58. Gulati M, Lohse MB, Ennis CL, Gonzalez RE, Perry AM, Bapat P, et al. In Vitro Culturing and Screening of *Candida albicans* Biofilms. *Curr Protoc Microbiol*. 2018; 50:e60. <https://doi.org/10.1002/cpmc.60> PMID: 29995344

59. Kirchner FR, Littringer K, Altmeier S, Tran VDT, Schönherr F, Lemberg C, et al. Persistence of *Candida albicans* in the Oral Mucosa Induces a Curbed Inflammatory Host Response That Is Independent of Immunosuppression. *Front Immunol*. 2019; 10:330. <https://doi.org/10.3389/fimmu.2019.00330> PMID: [30873177](https://pubmed.ncbi.nlm.nih.gov/30873177/)
60. Santana DJ, Anderson FM, O'Meara TR. Monitoring Inflammasome Priming and Activation in Response to *Candida albicans*. *Curr Protoc Microbiol*. 2020; 59:e124. <https://doi.org/10.1002/cpmc.124> PMID: [33108055](https://pubmed.ncbi.nlm.nih.gov/33108055/)
61. Lemberg C, de San M, Vicente K, Fróis-Martins R, Altmeier S, Tran VDT, et al. *Candida albicans* commensalism in the oral mucosa is favoured by limited virulence and metabolic adaptation. *PLoS Pathog*. 2022; 18:e1010012. <https://doi.org/10.1371/journal.ppat.1010012> PMID: [35404986](https://pubmed.ncbi.nlm.nih.gov/35404986/)
62. O'Meara TR, Duah K, Guo CX, Maxson ME, Gaudet RG, Koselny K, et al. High-Throughput Screening Identifies Genes Required for *Candida albicans* Induction of Macrophage Pyroptosis. *Kronstad JW, editor. MBio*. 2018; 9:a019620. <https://doi.org/10.1128/mBio.01581-18> PMID: [30131363](https://pubmed.ncbi.nlm.nih.gov/30131363/)
63. O'Meara TR, Veri AO, Ketela T, Jiang B, Roemer T, Cowen LE. Global analysis of fungal morphology exposes mechanisms of host cell escape. *Nat Commun*. 2015; 6:6741. <https://doi.org/10.1038/ncomms7741> PMID: [25824284](https://pubmed.ncbi.nlm.nih.gov/25824284/)
64. Wellington M, Koselny K, Krysan DJ. *Candida albicans* morphogenesis is not required for macrophage interleukin 1 $\beta$  production. *MBio*. 2012; 4:e00433–e00412.
65. Shankar M, Lo TL, Traven A. Natural Variation in Clinical Isolates of *Candida albicans* Modulates Neutrophil Responses. *mSphere*. 2020;5. <https://doi.org/10.1128/mSphere.00501-20> PMID: [32817378](https://pubmed.ncbi.nlm.nih.gov/32817378/)
66. Gerwien F, Dunker C, Brandt P, Garbe E, Jacobsen ID, Vylkova S. Clinical *Candida albicans* Vaginal Isolates and a Laboratory Strain Show Divergent Behaviors during Macrophage Interactions. *mSphere*. 2020;5. <https://doi.org/10.1128/mSphere.00393-20> PMID: [32817377](https://pubmed.ncbi.nlm.nih.gov/32817377/)
67. O'Meara TR, O'Meara MJ. DeORFanizing *Candida albicans* Genes using Coexpression. *mSphere*. 2021;6. <https://doi.org/10.1128/mSphere.01245-20> PMID: [33472984](https://pubmed.ncbi.nlm.nih.gov/33472984/)
68. Homann OR, Dea J, Noble SM, Johnson AD. A phenotypic profile of the *Candida albicans* regulatory network. *PLoS Genet*. 2009; 5:e1000783. <https://doi.org/10.1371/journal.pgen.1000783> PMID: [20041210](https://pubmed.ncbi.nlm.nih.gov/20041210/)
69. Brennan M, Thomas DY, Whiteway M, Kavanagh K. Correlation between virulence of *Candida albicans* mutants in mice and *Galleria mellonella* larvae. *FEMS Immunol Med Microbiol*. 2002; 34:153–157. <https://doi.org/10.1111/j.1574-695X.2002.tb00617.x> PMID: [12381467](https://pubmed.ncbi.nlm.nih.gov/12381467/)
70. Dunn MJ, Woodruff AL, Anderson MZ. The *Galleria mellonella* Waxworm Infection Model for Disseminated Candidiasis. *J Vis Exp*. 2018. <https://doi.org/10.3791/58914> PMID: [30507902](https://pubmed.ncbi.nlm.nih.gov/30507902/)
71. Hill JA, O'Meara TR, Cowen LE. Fitness trade-offs associated with the evolution of resistance to antifungal drug combinations. *Cell Rep*. 2015; 10:809–819.
72. Martin RM, Bachman MA. Colonization, Infection, and the Accessory Genome of *Klebsiella pneumoniae*. *Front Cell Infect Microbiol*. 2018; 8:4. <https://doi.org/10.3389/fcimb.2018.00004> PMID: [29404282](https://pubmed.ncbi.nlm.nih.gov/29404282/)
73. Allen JP, Snitkin E, Pincus NB, Hauser AR. Forest and Trees: Exploring Bacterial Virulence with Genome-wide Association Studies and Machine Learning. *Trends Microbiol*. 2021; 29:621–633. <https://doi.org/10.1016/j.tim.2020.12.002> PMID: [33455849](https://pubmed.ncbi.nlm.nih.gov/33455849/)
74. Gerstein AC, Jackson KM, McDonald TR, Wang Y, Lueck BD, Bohjanen S, et al. Identification of Pathogen Genomic Differences That Impact Human Immune Response and Disease during *Cryptococcus neoformans* Infection. *MBio*. 2019;10. <https://doi.org/10.1128/mBio.01440-19> PMID: [31311883](https://pubmed.ncbi.nlm.nih.gov/31311883/)
75. Beale MA, Sabiiti W, Robertson EJ, Fuentes-Cabrejo KM, O'Hanlon SJ, Jarvis JN, et al. Genotypic Diversity Is Associated with Clinical Outcome and Phenotype in Cryptococcal Meningitis across Southern Africa. *PLoS Negl Trop Dis*. 2015; 9:e0003847. <https://doi.org/10.1371/journal.pntd.0003847> PMID: [26110902](https://pubmed.ncbi.nlm.nih.gov/26110902/)
76. Forche A, Alby K, Schaefer D, Johnson AD, Berman J, Bennett RJ. The Parasexual Cycle in *Candida albicans* Provides an Alternative Pathway to Meiosis for the Formation of Recombinant Strains. *Heitman J, editor. PLoS Biol*. 2008; 6:e110–e114. <https://doi.org/10.1371/journal.pbio.0060110> PMID: [18462019](https://pubmed.ncbi.nlm.nih.gov/18462019/)
77. Mixão V, Gabaldón T. Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*. *BMC Biol*. 2020; 18:48. <https://doi.org/10.1186/s12915-020-00776-6> PMID: [32375762](https://pubmed.ncbi.nlm.nih.gov/32375762/)
78. Ene IV, Bennett RJ, Anderson MZ. Mechanisms of genome evolution in *Candida albicans*. *Curr Opin Microbiol*. 2019; 52:47–54. <https://doi.org/10.1016/j.mib.2019.05.001> PMID: [31176092](https://pubmed.ncbi.nlm.nih.gov/31176092/)
79. Wakade RS, Ristow LC, Wellington M, Krysan DJ. Intravital imaging-based genetic screen reveals the transcriptional network governing *Candida albicans* filamentation during mammalian infection. *Elife*. 2023;12. <https://doi.org/10.7554/eLife.85114> PMID: [36847358](https://pubmed.ncbi.nlm.nih.gov/36847358/)

80. Pujol C, Reynes J, Renaud F, Raymond M, Tibayrenc M, Ayala FJ, et al. The yeast *Candida albicans* has a clonal mode of reproduction in a population of infected human immunodeficiency virus-positive patients. *Proc Natl Acad Sci U S A*. 1993; 90:9456–9459. <https://doi.org/10.1073/pnas.90.20.9456> PMID: 8415722
81. Moorhouse AJ, Rennison C, Raza M, Lilic D, Gow NAR. Clonal Strain Persistence of *Candida albicans* Isolates from Chronic Mucocutaneous Candidiasis Patients. *PLoS ONE*. 2016; 11:e0145888. <https://doi.org/10.1371/journal.pone.0145888> PMID: 26849050
82. McManus BA, Maguire R, Cashin PJ, Claffey N, Flint S, Abdulrahim MH, et al. Enrichment of multilocus sequence typing clade 1 with oral *Candida albicans* isolates in patients with untreated periodontitis. *J Clin Microbiol*. 2012; 50:3335–3344. <https://doi.org/10.1128/JCM.01532-12> PMID: 22875886
83. Ford CB, Funt JM, Abbey D, Issi L, Guiducci C, Martinez DA, et al. The evolution of drug resistance in clinical isolates of *Candida albicans*. *Elife*. 2015; 4:e00662. <https://doi.org/10.7554/eLife.00662> PMID: 25646566
84. Roemer T, Jiang B, Davison J, Ketela T, Veillette K, Breton A, et al. Large-scale essential gene identification in *Candida albicans* and applications to antifungal drug discovery. *Mol Microbiol*. 2003; 167–181. <https://doi.org/10.1046/j.1365-2958.2003.03697.x> PMID: 14507372
85. Sprouffske K, Wagner A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*. 2016; 17:172. <https://doi.org/10.1186/s12859-016-1016-7> PMID: 27094401
86. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics*. 2013; 43:11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43> PMID: 25431634
87. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. John Wiley & Sons, Ltd. (10.1111); 2012.
88. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr Protoc Bioinformatics*. 2020; 69:e96. <https://doi.org/10.1002/cpbi.96> PMID: 32162851
89. Yu G, Lam TT-Y, Zhu H, Guan Y. Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Mol Biol Evol*. 2018; 35:3041–3043. <https://doi.org/10.1093/molbev/msy194> PMID: 30351396
90. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. Ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2017; 8:28–36.
91. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014; 15:R84. <https://doi.org/10.1186/gb-2014-15-6-r84> PMID: 24970577
92. Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, Garrison EP, et al. SpeedSeq: ultra-fast personal genome analysis and interpretation. *Nat Methods*. 2015; 12:966–968. <https://doi.org/10.1038/nmeth.3505> PMID: 26258291
93. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018; 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191> PMID: 29750242
94. Selmecki A, Bergmann S, Berman J. Comparative genome hybridization reveals widespread aneuploidy in *Candida albicans* laboratory strains. *Mol Microbiol*. 2005; 55:1553–1565. <https://doi.org/10.1111/j.1365-2958.2005.04492.x> PMID: 15720560
95. Chibana H, Beckerman JL, Magee PT. Fine-resolution physical mapping of genomic diversity in *Candida albicans*. *Genome Res*. 2000; 10:1865–1877. <https://doi.org/10.1101/gr.148600> PMID: 11116083
96. Veri AO, Miao Z, Shapiro RS, Tebbji F, O'Meara TR, Kim SH, et al. Tuning Hsf1 levels drives distinct fungal morphogenetic programs with depletion impairing Hsp90 function and overexpression expanding the target space. *PLoS Genet*. 2018; 14:e1007270. <https://doi.org/10.1371/journal.pgen.1007270> PMID: 29590106
97. Shen J, Guo W, Köhler JR. CaNAT1, a heterologous dominant selectable marker for transformation of *Candida albicans* and other pathogenic *Candida* species. *Infect Immun*. 2005; 73:1239–1242. <https://doi.org/10.1128/IAI.73.2.1239-1242.2005> PMID: 15664973
98. Fuchs BB, O'Brien E, Khoury JBE, Mylonakis E. Methods for using *Galleria mellonella* as a model host to study fungal pathogenesis. *Virulence*. 2010; 1:475–482. <https://doi.org/10.4161/viru.1.6.12985> PMID: 21178491