FULL TITLE: Human commensal *Candida albicans* strains demonstrate substantial within-host diversity and retained pathogenic potential

AUTHORS: Faith M Anderson[¶1], Noelle Visser[¶1#a], Kevin Amses[2#b], Andrea Hodgins-Davis[1], Alexandra M Weber[3], Katura M Metzner[1], Michael J McFadden[1], Ryan E Mills[3,4] Matthew J O'Meara[3], Timothy Y James[2], Teresa R O'Meara[1*]

AFFILIATIONS:

1. Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI, USA
2. Department of Ecology and Evolution, University of Michigan, Ann Arbor, MI, USA
3. Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA
4. Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA

[#a] Current address: Department of Biology, University of Louisville, Louisville, Kentucky, USA
[#b] Current address: Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA

[¶] authors contributed equally

* Corresponding author

Short Title: Commensal Candida albicans diversity

***Address correspondence to:**
Teresa O'Meara, Ph.D
Department of Microbiology and Immunology
1150 W. Medical Center Dr
Medical Sciences Building II, Room 6751
Ann Arbor, MI, 48109, USA
Phone: 734-647-1853
Fax: 734-764-3562
Email: tromeara@umich.edu

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 undertook an examination of commensal isolates from healthy donors with a goal of identifying site-specific phenotypic adaptation and genetic variation associated with these phenotypes. We demonstrate that healthy people are reservoirs for genotypically and phenotypically diverse *C. albicans* strains, and that this genetic diversity includes both SNVs and structural rearrangements. 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. However, our commensal strains retained the capacity to cause disease in systemic models of 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.

INTRODUCTION:

Candida albicans is a common colonizer of humans, with between 20-80% of the world's population estimated to be asymptomatically 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]. 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 [6–9]. *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 [10–12]. Additionally, *C. albicans* can cause more minor mucosal infections, including oral and vaginal thrush and skin infections [13]. As a consequence, *C. albicans* represents the 2nd most common human fungal pathogen and the most common source of healthcare-associated fungal infections [14].

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 [15,16]. Recent work has highlighted intra-species variation in important aspects of *C. albicans* biology [17–20]. There are currently 17 clades of *C. albicans,* which were initially defined through multilocus sequence typing (MLST) [21–23]. More recently, genome sequencing has resolved finer population structure in the

group [24]. 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 [25], with five clades accounting for the majority of clinical isolates [26–28]. 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 [25,29–33]. 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 five different strains, each representing a different clade of *C. albicans,* have highlighted that circuit diversification is widespread between strains [17], 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 [34–40]. 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 [34–39]. In oral candidiasis, trisomic strains have been identified with a commensal phenotype [41]. Additionally, the 529L strain of *C. albicans* causes less damage and inflammation during oropharyngeal candidiasis [42], and persists at a higher fungal burden in both the mouth and the gut [20]. 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. We obtained representative isolates from healthy undergraduate student donors, including both oral and fecal isolates. These isolates included representatives from 8 clades of *C. albicans,* including instances of colonization by strains from multiple clades within a single individual, highlighting the within-host diversity of the colonizing fungal population. Synthetic long-read sequencing coupled with CHEF gel analysis revealed large structural variation between strains, consistent with early karyotyping work in the field [43]. Phenotyping of the strains revealed extensive variation in growth, stress response, biofilm formation, and interaction with macrophages, but no separation of strains by sample origin site. Instead, we identified shared structural variations and phenotypes that were shared between strains from the same phylogenetic clade, suggesting that underlying genetic background may be more predictive of certain phenotypes than selection pressure at a specific sample site. We also discovered and mapped specific nucleotide variants to the ability of strains to invade agar, identifying a new role for *ZMS1* transcription factor. Notably, 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. These new isolates complement previous work focusing on bloodstream or mucosal infection isolates [18]. In this population of students, 16% (16/98) were positive for oral *Candida* colonization and 12% (12/98) were positive for 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), with oral samples giving rise to more individual colonies per donor (Fig 1B).

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 mean and SD in number of colonies from each positive individual donor. C) Strains varied in maximum growth rate and carrying capacity in response to different carbon sources. Growth curves were performed on each isolate under three carbon sources at 30°C for 24 hours in biological duplicate. Rate and carrying capacity were determined using the GrowthcurveR analysis package. 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 20X 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) Parallel coordinate plot. Each strain was ranked for each growth condition, and the relationships between phenotypes are indicated with the lines. Blue highlights the slow growing strains.

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 unimodal distributions with a long tail of slow-growing strains for both exponential growth (μ_{max}) and saturating density (K) in rich media (Fig 1C). The five slowest-growing strains originated from five different hosts and showed significantly decreased growth rate and saturating density relative to other isolates from the same host. Although the five 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 correlated (S1B Fig), as slow growth on one carbon source was predictive of slow growth on other carbon sources.

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 [20,34,36], we examined each isolate in the collection for their ability

to form hyphae using 10% serum and febrile temperatures as two inducing cues. Additionally, we examined the morphology of each strain under the non-inducing condition of YPD at 30°C to identify constitutive filamentation, as has been observed from isolates collected from sputum samples from cystic fibrosis patients [44]. We observed that while none of the strains were constitutively filamentous, several strains aggregated at YPD 30°C. All strains were able to filament in response to the inducing conditions of serum and high temperature, albeit with some variation in the shape and aggregation of the cells (Fig 1D). This may be consistent with the hypothesis that interaction with bacteria in the gut maintains selection for the hyphal program [34,45]. Notably, the aggregative isolates were more likely to grow slowly and reach a lower carrying capacity. The universal ability of commensal strains isolated from healthy humans to filament stands in contrast to the observation that *C. albicans* evolves a yeast-locked phenotype in germ-free or antibiotic-treated mouse models, suggesting that selection in these models may not recapitulate important features of *C. albicans*-human interactions.

Filamentation programs under solid and liquid growth can involve distinct genetic programs [46], and the human host can present a variety of substrates for *C. albicans* to utilize. Therefore, we also tested each isolate for its colony morphology and capacity to invade into a set of solid agar media, including YPD agar at 30°C as the baseline condition, Spider agar at 37°C as a strong hyphae-inducing condition, YPD agar at 37°C as an intermediate inducing condition, and YPD agar at 37°C anaerobic conditions [46]. After several days of growth in each condition, colonies were gently rinsed from the plate and were given an invasion score from 0-5, with representative images of each

196 score provided in (Fig 1E). As expected for a strong inducing cue, the highest degree of
197 invasive growth was observed on Spider agar at 37°C [46]. Interestingly, although all
198 strains showed the ability to form hyphae under liquid culture conditions, many strains
199 failed to invade into solid agar, giving scores of 0 or 1 under multiple conditions (Fig
200 1E). Overall, there was a trend for the oral strains to invade more on Spider agar (S1C
201 Fig). We also observed substantial phenotypic heterogeneity among strains isolated
202 from the same host, including in strains isolated from the same site, consistent with
203 each individual being colonized by multiple, phenotypically-distinct, strains of *C.*
204 *albicans*.

205 To examine the relationship between the strains and the observed phenotypes,
206 we generated a parallel coordinate plot by ranking each strain for each phenotype (Fig
207 1F). From these data, we could observe that a subset of the slow growing strains in
208 YPD (blue) exhibited increased invasion on solid agar surfaces (Fig 1F), suggesting that
209 slow growth in liquid media does not necessarily correlate with defects in other growth
210 conditions. Upon closer inspection, several of these slow growing strains originated
211 from donor 882. This subset of strains showed hyper invasion (score of 5) at the
212 baseline condition of YPD agar at 30°C. Interestingly, although these strains retained
213 their hyper invasive phenotype under the strong inducing cue of Spider agar at 37°C,
214 they exhibited only moderate invasion (average score of 2) in the YPD agar at 37°C
215 condition. Together, these experiments demonstrate a large variation in phenotypes
216 between commensal isolates, even among traits, such as filamentation and invasion,
217 that are often correlated with virulence.
218

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, to represent an alternate clinical isolate that has been previously phenotypically and genotypically characterized [47].

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 and the SC5314 and CHN1 reference strains using the Transposase Enzyme Linked Long-read Sequencing (TELL-Seq) method for library preparation [48]. This approach uses barcode linked-reads to produce synthetic long reads with Illumina quality sequence, thus allowing us to capture both SNVs and structural variants. The TELL-Seq library was then sequenced using an Illumina NovaSeq, resulting in a read coverage of approximately 150X for each sample.

To identify single nucleotide variants (SNVs) and compare our strains to the existing set of sequenced *C. albicans* isolates, we collected 388 previously published *C. albicans* genomes and mapped them, and our 45 newly-generated genomes, to the

SC5314 reference genome with BWA- MEM [49]. 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 one 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 [24]. 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 [18,24,50]. 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., [34,51,52]). 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 dataset, our SNV-based clustering recovers all major accepted clades of *C. albicans* (Fig 2) [21–25]. 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 with 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, our isolates are not represented in clades of *C. albicans* known to exhibit a high degree of geographic specificity (e.g., Clade 13) (Fig 2) [24] .

Fig 2. The newly sequenced *C. albicans* commensal isolates include representatives from multiple clades.

A) A maximum likelihood tree showing the phylogenetic relationships between the 324 isolates analyzed via (neighbor joining). Previous clusters from [24] are highlighted with colored boxes. New isolates were colored based on the nearest defined cluster.

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 [52], we observed primarily SNVs and short-tract loss of heterozygosity between these six isolates, perhaps consistent with clonal expansion and diversification during colonization. In contrast, 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 from donors 838 and 833, likely indicating transmission between individuals. Together, this suggests that variation in an individual's colonizing *C. albicans* strains can 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 (Fig 3A, S2A Fig). The commensal *C. albicans* strains show between 7 and 10 chromosome bands, ranging from ~0.8 MB to ~3.2 MB in size [53]. 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.

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. Novel chromosomes are indicated with the colored bands. Gel images in S2 Fig. 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 one (red) suggest potential duplications while values less than one (blue) suggest potential deletions. C) Isolates 880-2 and 859-2 chromosomal fusion events identified & chosen for validation. D) PCR validation of fusion events in isolates 880-2 and 859-2. Bands corresponding with fusion events were not present in the SC5314 reference strain.

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 and 838 all contained an extra chromosome 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 a unique pattern with different banding patterns at Chrs 2, 3, and 4. The 3 strains from Clade C, which originated from donors 882 and 811, all contained a chromosome band between Chrs 5 and 6. Finally, the two closely related strains, 859-2 and 806-1, both contained a chromosome band between Chrs 2 and 3 (Fig 3A).

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. We could observe variation in copy number across the chromosomes (Fig 3B), as well as potential inversions, duplications, or deletions (S2B Fig). Across our isolates, we observed substantial copy number variation, with a trend towards increased copy number. We noted that the isolates globally exhibited copy number expansion in the telomeric regions, and that copy number increases were more common on smaller chromosomes. Overall, the considerable copy number variation among our commensal isolates is in line with past work suggesting that host environmental pressures induce changes in genome size and that increased ploidy enhances fitness within the host [54,55].

Notably, we were able to identify structural rearrangements and putative chromosome fusions, 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 3D). 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 7kb 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 (Fig 3C). We were again able to use PCR to span both junctions observed the presence of the fusion between chromosomes 1, 2, and 3 (Fig 3D). 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 3D). These structural variations were not captured in the SNV analysis, and may play important roles in gene regulation or phenotypic variation between the strains.

Together, our sequence-level and structural-level genomic analysis of commensal *C. albicans* isolates indicate that healthy individuals can harbor multiple strains of *C. albicans* from different clades. Furthermore, our structural analysis shows substantial heterogeneity in number and size of chromosomes among commensal isolates, which suggests that even within similar strains, there is potential for structural variation.

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 specific adaptations in *C. albicans* strains isolated from different sites that allow for colonizing different host microenvironments. The host sites commonly colonized by *C. albicans* vary dramatically in environmental cues, such as nutrient availability, pH, immune responses, and resident microbiomes. 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.

Fig 4. Deep phenotyping reveals heterogeneity in *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. C) Macrophage cell death rates. D) Representative images of isolates following 4 hour macrophage infection. Representative filamentation score of 0 (left). Representative filamentation score of 5 (right). 20x magnification. Scale = 50 μM. E) Parallel coordinate plot ranking

isolates for macrophage filamentation scoring, phagocytosis rates, and cell death rates. SC5314 reference strain is indicated in red. For phagocytosis rates, significant differences from the SC5314 reference strain were determined by one-way ANOVA, with Dunnett's multiple correction testing. For cell death, significant differences from the mock condition were determined by one-way ANOVA, with Dunnett's multiple correction testing. Asterisks indicate $P < 0.05$ (*), $P < 0.005$ (**), $P < 0.001$ (***) and $P < 0.0001$ (****).

From these data, we identified 3 strains, 882-60, 882-46, and 811-7, that consistently grew more poorly than the wildtype under multiple conditions; these strains were those that exhibited aggregation at 30°C and slow growth in rich media conditions. These strains all belonged to Clade C and were closely related, despite arising from two 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 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 [56]. Although we observed variation between the strains, there was no correlation between isolation site or clade with the propensity of isolates to form biofilms (S3 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 [57]. 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) [58]. 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 hours with LPS before infecting with each of our isolates for 4 hours. Following infection, we stained the cells with propidium iodide (PI) as a measure of cell death (Fig 4C). On average, the commensal isolates induced between 5% and 20% cell death, which was significantly lower than the reference strain, SC5314, which induced an average of 40% cell death. Several strains, including the 3 aggregating strains, 882-60, 882-46, and 811-7, were

not significantly different than the mock condition. Other than SC5314, only one isolate, 831-1, was significantly different ($p = 0.0411$) from the mock condition.

Recently, we showed that *C. albicans* mutants that filament in serum are not always filamentous within macrophages [59]. As filamentation is linked, but not required, for inflammasome activation within host phagocytes [59–61], and clinical isolates show variability in induction of host inflammatory responses [62,63], we examined the morphology of the commensal isolates after incubation for four hours 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. 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 (S4 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 (Fig 4E).

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 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 cosign metric. We observed three major clusters, but they did not segregate by isolation site, clade, or participant (S5 Fig). In sum, all of the commensal

448 isolates showed extensive phenotypic variation, but this was not dependent on the body
449 site or participant from which they were collected.
450
451 **Limited Diversity Exploitation:**
452 Genome-wide association studies have been a powerful tool for identifying the
453 genetic basis of variation in phenotypes of interest in humans and other recombining
454 species. However, the generally clonal and asexual reproduction of *C. albicans* creates
455 a population structure that confounds traditional GWAS methods. By sampling multiple
456 isolates from each individual, we were able to obtain phenotypically diverse strains with
457 a limited set of unique SNPs between isolates, allowing us to identify causative variants
458 associated with a particular phenotype.
459 We focused on the strains from donor 814, as this donor's matched oral and
460 fecal samples produced the most colonies of any donor, with 144 colonies arising from
461 the oral sample, and 19 arising from the fecal sample, for a total of 163 colonies. 6 of
462 these 163 colonies were included in the condensed set, and these 6 strains clustered
463 tightly in Clade 1, which we hypothesized would allow us to identify causative variants
464 associated with particular phenotypes that were divergent between strains.
465 Our agar invasion analysis revealed that isolate 814-168 demonstrated hyper
466 invasion into Spider agar at 37°C, whereas the other 5 isolates from the condensed set
467 were less invasive (Fig 5A). Moreover, from this donor's 163 isolates, only this single
468 isolate exhibited the hyper invasive phenotype into Spider agar at 37°C (Fig 5B); this
469 phenotype was the motivation for initially including this strain in the sequenced set.

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.

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. Dark outlines indicate genes examined in Fig 5E.

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 determined after gentle washing.

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 [64] 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 an invasive phenotype, we generated complementation plasmids encoding the *ZMS1T681S* allele cloned out of the 814-168 background, and performed allele swap transformations into 814-183, a closely related strain from the same host which demonstrated minimal agar

493 invasion. In the minimally invasive 814-183 background, replacing one copy of the
494 endogenous *ZMS1* allele with one harboring a serine residue at amino acid 681
495 (T681S) resulted in hyper invasion into Spider agar (Fig 5D).
496 Previous work on the function of Zms1 via deletion mutant analysis had not
497 revealed a phenotype [65], however, this was in the SC5314 genetic background and
498 the impact of a specific transcription factor on a given phenotype can vary depending on
499 the strain [17]. Therefore, we deleted *ZMS1* from both 814 backgrounds and tested the
500 strains for invasion and filamentation. On YPD agar, deletion of *ZMS1* in both genetic
501 backgrounds had minimal effects, with the mutant strains behaving similarly to their
502 parent strains (Fig 5D). However, on Spider agar, *ZMS1* deletion changed the colony
503 morphology in the 814-168 background, although it did not decrease overall invasion.
504 Additionally, deletion of *ZMS1* increased invasion in the otherwise minimally invasive
505 814-183 background, highlighting the differential impact of *ZMS1* mutation in the
506 different genetic backgrounds (Fig 5D). Our results demonstrate that a single SNV
507 changing amino acid 681 to a serine is a dominant active allele that is sufficient to drive
508 a hyphal invasion program into Spider agar. We also identified natural variation that was
509 distinct from deletion phenotypes. This approach highlights how deep phenotypic
510 analysis of a limited set of natural isolates from a single host can be exploited to identify
511 causative variants and identify new functions for under-characterized genes.
512
513 **Virulence:**
514 We next examined the fitness and virulence of the commensal isolates relative to
515 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 two models of *Galleria mellonella* systemic infection as this insect model of virulence is significantly correlated with the murine systemic infection model [18,66,67]. We first examined competitive fitness by infecting with a 1:1 mixture of fluorescently marked SC5314 and each sequenced commensal isolate [68]. After three days, the worms were homogenized and CFUs were plated to determine the competitive index, calculated as the log2 ratio of fluorescent to nonfluorescent colonies [35]. 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 slow growth exhibited by these strains in many growth conditions (Fig 4A).

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 non-fluorescent colonies, normalized to the inoculum, and log2 transformed. Significant differences from the SC5314 reference strain were determined by one-way ANOVA, with Dunnett's multiple correction testing. B) Survival assays in

539 *G. mellonella*, comparing the SC5314 reference to 6 isolates from donor 814.
540 Each strain was standardized to 2x10^6 cells/mL before inoculating 20 *G.*
541 *mellonella* larvae per strain with 50 µL of prepared inoculum. Larvae were
542 monitored daily for survival. Statistical differences were determined using a
543 Mantel-Cox log-rank test. Asterisks indicate $P < 0.05$ (*), $P < 0.005$ (**), $P <$
544 0.001 (***) and $P < 0.0001$ (****) compared with SC5314.
545
546 The striking increased competitive fitness of the other isolates motivated us to
547 test whether this increased ratio was correlated with increased disease. Here, we
548 examined the survival of *G. mellonella* after performing monotypic infections. We started
549 by using the six isolates from donor 814, which all had increased competitive fitness
550 compared with SC5314, to test the hypothesis that increased invasion is associated
551 with increased disease. However, the majority of the strains from donor 814 were not
552 significantly different from the SC5314 reference strain ($p > 0.05$, log-rank test),
553 including the hyper-invasive 814-168 isolate (Fig 6B). Two strains had a slight defect in
554 virulence ($p < 0.05$ log-rank test). We additionally tested two other clusters of strains for
555 their ability to cause systemic disease in the insect model. However, the human
556 commensal isolates were again not significantly decreased in virulence from the
557 SC5314 reference strain (S6 Fig), despite the increased ability of the SC5314 strain to
558 cause host immune cell death (Fig 4C). Together, this suggests that there is not a
559 consistent selection for avirulent behavior in the commensal human isolates. Although
560 we observed wide variation in *in vitro* host response phenotypes, the isolates generally

561 retained their pathogenic potential, indicating that our *in vitro* assays may not capture
562 the complex stresses experienced during whole-organism infections.
563
564
565 **DISCUSSION:**
566
567 Intra-species analyses of microbial strains can allow for the identification of
568 variants that are associated with specific clinical outcomes, as demonstrated widely in
569 bacterial virulence [69,70] and recently in the human fungal pathogen *Cryptococcus*
570 *neoformans* [71,72]. Understanding the differences between strains can allow for
571 insights into mechanisms of colonization and pathogenesis [17]. However, assessing
572 the underlying genetic variation responsible for differences in virulence in *C. albicans* is
573 challenging in the absence of clear candidate genes because *C. albicans* is a primarily
574 clonal yeast that does not generally undergo meiosis and recombination. Additionally,
575 the genetic basis underlying the ability of commensal strains of *C. albicans* to transition
576 to pathogenic behavior is not well defined. Moreover, *C.albicans* exhibits high rates of
577 structural mutation and ploidy variation [41,73] as well as high heterozygosity [24].
578 Previous descriptions of population genomic variation in *C. albicans* have largely relied
579 on short read sequencing [24] or molecular typing methods [27] at relatively few loci to
580 define genetic variation in the species. Here, we have greatly expanded the number of
581 available long-read genomes for *C. albicans* and have generated a catalog of structural
582 variants that was largely absent from previous descriptions of natural diversity. We were
583 able to leverage clonal variation within a single donor to identify meaningful variation

584 that arose during microevolution. Overall, we performed a systematic phenotypic
585 analysis of commensal isolates from healthy donors, thus allowing us to examine *C.*
586 *albicans* genotypic and phenotypic diversity before the transition to virulence.
587 Strikingly, our commensal strains generally maintained their capacity to cause
588 disease, and all strains were able to filament in response to the inducing cues of 10%
589 serum or high temperature. This observation is in contrast to work suggesting that
590 passage through the mammalian gut would cause a decrease in systemic virulence and
591 filamentation [34]. Although we observed significant variation in the ability of strains to
592 invade into agar, recent intravital imaging approaches suggest that filamentation in
593 response to serum matched that seen *in vivo* more than filamentation in response to
594 solid Spider agar [74]. Together, the robust filamentation and disease-causing ability in
595 all of the commensal isolates suggests that the selective pressures that occur during
596 mouse models of colonization may not recapitulate the selection that occurs during
597 human colonization.
598 Although we were not able to associate a particular phenotype with increased
599 systemic disease, we observed extensive phenotypic and genotypic variation between
600 the commensal isolates. This variability is also consistent with recent work from clinical,
601 disease-associated strains [18]. Moreover, our commensal strains were able to
602 proliferate on a range of different environmental conditions, and we did not observe
603 significant differences in phenotypes between isolates obtained from oral or fecal sites,
604 except for a slight increase in invasion into Spider medium for oral samples. This work
605 highlights the striking ability of *C. albicans* to adapt to a wide range of environmental

606 conditions and indicates that colonization at a specific body site does not necessarily
607 predict pathogenic potential.

608 Previous work has suggested that the *C. albicans* population within a given
609 individual is clonal [68,75–78], and that the fungus is acquired during birth as a part of
610 the normal microbiota [42]. In these studies, they were often sampling from patients with
611 active disease; this may suggest that there is selection for the ability to cause disease,
612 resulting in repeated isolation of representative samples of a clonal population. In
613 contrast, we identified disparate individuals that appeared to be colonized by strains that
614 were nearly identical, suggesting that there was some transmission between individuals.
615 Whether these transmission events allow for long-term colonization, and how they affect
616 the initial *C. albicans* colonizing strains, is still not fully understood.

617 We also observed diversification within hosts, with multiple instances of closely-
618 related strains showing variation in phenotypes, such as invasion into agar. In one
619 representative example, we were able to use comparative genome analysis coupled
620 with co-expression, which we term “limited diversity exploitation”, to identify a candidate
621 transcription factor that regulates invasion. Previous work on a *ZMS1* knockout strain of
622 *C. albicans* did not show any differences in phenotype compared with the parent strain
623 [65]. However, we observed that a single amino acid substitution in the predicted fungal
624 transcription factor regulatory middle homology region was sufficient to drive hyper-
625 invasive growth. Moreover, this phenotype was distinct from the deletion phenotypes in
626 these two genetic backgrounds, again contrasting with the SC5314 reference strain. It is
627 likely that populations of colonizing *C. albicans* will vary in other important clinically-
628 relevant traits, including filamentation in macrophages or intrinsic drug tolerance and

629 resistance, and our results suggest that studying more than a single representative

630 isolate gives opportunities to discover new biology.

631 **METHODS:**

632 **Human Subjects and Sample Collection:**

633 Study volunteers were recruited through the Authentic Research Sections of the
634 introductory biology laboratory course at the University of Michigan (BIO173). All
635 participants provided written, informed consent for sample collection as well as isolation
636 and characterization of microbes from feces. This study was approved by the
637 Institutional Review Board of the University of Michigan Medical School (HUM00094242
638 and HUM00118951) and was conducted in compliance with the Helsinki Declaration.
639 Oral samples were obtained by self-administered cotton swabbing, and fecal
640 samples were collected and diluted in PBS+DMSO before plating; in each case, BD
641 ChromAgar was used to identify *C. albicans* colonies. Individual *Candida albicans*
642 colonies were picked from all plates with a toothpick into 100 µL of YPD in a 96 well
643 plate. The cultures were then inoculated onto BD Chromagar to differentiate between
644 yeast species and ITS regions were amplified to confirm *C. albicans* strains. The *C.*
645 *albicans* colonies were individually inoculated into fresh 96-well plate in YPD and after
646 24 hours of incubation at 30° C, 50% glycerol was added to generate the stock plates.
647 For each oral sample, individual *Candida albicans* colonies were picked from the
648 CHROMagar plates and incubated overnight in 100 µL of YPD in a 96 well plate. After
649 24 hours incubation in 30° C, 50% glycerol was added to the 96 wells to generate the
650 stock plates. All strains were maintained in -80 C cryoculture.
651
652 **Media and culture conditions:**
653 Media for growth curves is described in Supplemental Table 1.

654

655 **Growth Curve Analysis:**

656 Overnight cultures of each *Candida albicans* isolate were grown in 200 µL of
657 YPD in 96-well plates at 30° C, then subcultured into fresh media using a sterile pinner.
658 Growth curves were performed on a BioTek 800 TS Absorbance Reader in the
659 indicated conditions for 24 hours, without shaking. Maximum growth rate and carrying
660 capacity were determined using the GrowthcurveR analysis package [79]. All media and
661 conditions are listed in Table S1. Summary statistics from GrowthcurveR for the entire
662 collection of isolates are included in Table S2. Summary statistics from GrowthcurveR
663 for the condensed set in multiple environmental conditions are included in Table S4.

664

665 **Filamentation Analysis:**

666 Overnight cultures of each *Candida albicans* isolate were grown in 100 µL of
667 YPD in 96 well plates at 30° C, then subcultured into 100 µL of YPD at either 30° C or
668 42° C for three hours. After three hours, the plates were imaged on an Olympus IX80
669 microscope at 20X magnification. To test the response to serum, overnight cultures
670 were subcultured into 5 mL of YPD with 10% serum and rotated in a 37 °C incubator for
671 three hours before vortexing and imaging on glass slides at 20X magnification.

672

673 **Agar Invasion Methods**

674 Isolates were taken from frozen glycerol stocks, incubated overnight in yeast
675 extract peptone dextrose (YPD) in 96-well plates and inoculated onto solid YPD and
676 spider media using a 96-pin replicator tool (Singer Instruments). Plates were then

677 incubated at the indicated temperature and oxygen conditions for 7-9 days. Colonies on
678 plates were first imaged on a Biorad Gel Doc XR. Colonies were then gently washed off
679 with deionized water. Plates were imaged again to capture invasion images. The
680 invasiveness of each isolate was determined through a rubric scale of 0-5, as indicated
681 in Fig 1. A 0 indicates no invasion, 1-2 indicates minimal invasion, a 3 indicates distinct
682 circular hyphal invasion, while a 4 indicates an even larger circular hyphal invasion. A 5
683 indicates the most aggressive invasion, with a "halo" of more hyphae surrounding the
684 initial growth. A summary of invasion scores is in Table S2.
685
686 ***Galleria mellonella* Infections**
687 Each isolate was examined for competitive fitness in a one-to-one ratio with a
688 fluorescent wildtype SC5314 *C. albicans* strain in *Galleria mellonella* larvae. Infections
689 were performed as previously described (Fuchs et al., 2010). Briefly, *G. mellonella*
690 larvae were purchased from speedyworm.com and maintained in sawdust at room
691 temperature. Overnights were prepared for each isolate and wildtype strain in yeast
692 extract peptone dextrose (YPD) at 30°C, with rotation.
693 To measure competitive fitness, a 1:1 ratio of fluorescent wild type to unlabeled
694 isolate was prepared in 1X PBS at $5x10^5$ cells/ml. 10 larvae/strain were randomly
695 chosen and infected via the last right proleg with 50 µL of the 1:1 inoculum using an
696 exel veterinary U-40 diabetic syringe (0.5CC X 29G X ½). After injection, larvae were
697 maintained at room temperature for 3 days before harvesting using a Benchmark D1000
698 homogenizer. Larvae were homogenized in 0.5 ml of 1X PBS, diluted 1:10 in PBS, and
699 plated onto YPD plates containing gentamicin, ampicillin, and ciprofloxacin. Plates were

left at 30°C for 48 hours 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 $2x10^6$ cells/mL inoculum using an exel veterinary U-40 diabetic syringe (0.5CC X 29G X 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).

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 µM 2-mercaptoethanol, and Pen/Strep), then immortalized using Cre-J2 viral supernatants.

Phagocytosis

Phagocytosis assays were performed as previously described [58]. Briefly, iBMDM macrophages were prepared for infection in RPMI media containing 3% FBS, diluted to $3x10^6$ cells/ml and incubated overnight at 100 µL/well in a 96-well plate. Overnight cultures of *Candida albicans* isolates were incubated and diluted to $4 X 10^6$ cells/ml into RPMI media with 3% FBS and used to infect macrophages. Inoculated plates were centrifuged for 1 minute at 1,000 rpm to synchronize. After 30 minutes,

media was removed, and cells were fixed with 4% paraformaldehyde (PFA) for 10 minutes. The cells were washed three times in 1 X PBS. Cells were then stained with 50 µL of FITC-Concanavalin (5 ug/mL Sigma-Aldrich C7642) at room temperature, rocking for 30 minutes, wrapped in foil. The plates were then washed 3X with 100 µL of 1X PBS, and then 50 µL of 0.05% Triton-X100 was added to permeabilize the cells. Cells were then washed 3X with 1X PBS and a final stain of 50 µL of calcofluor white (100 ug/mL, Sigma-Aldrich C7642) was added to cells to incubate for 10 minutes. The cells were then washed 3X with 100 µL of 1X PBS and then maintained in 100 µL of 1X PBS at 4°C before imaging on the microscope at 20X 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 – external cells)/total cells = % 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 $3x10^6$ cells/ml and incubated overnight at 100 µL/well in a 96-well plate. Overnights of *Candida albicans* isolates were incubated and diluted to 4 X 10^6 cells/ml into RPMI media with 3% FBS and used to infect macrophages at an MOI of ~1:1. The inoculated plate was centrifuged for 1 minute at 1,000 rpm to synchronize phagocytosis. After 2 hours, the media was removed, and cells were fixed with 4% paraformaldehyde (PFA) for 10 minutes, permeabilized with 0.05% Triton-X100, and stained with 50 µL of calcofluor

white (100 ug/mL, Sigma-Aldrich C7642) before imaging at 40X magnification using the DIC and DAPI channels. The extent of filamentation for each isolate was scored on a rubric from 0-5. 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.

Cell death assay

To assess macrophage cell death, J2-iBMDM macrophages were prepared for infection in RPMI media containing 3% FBS, diluted to $4x10^5$ cells/ml and incubated overnight at 100 µL/well in a 96-well plate. Macrophages were primed with 200 ng/mL LPS in RPMI media with 3% FBS for 2 hours. Overnights of *Candida albicans* isolates were incubated and diluted to 5 X 10^5 cells/ml into RPMI media with 3% FBS and used to infect macrophages at an MOI of ~1:1. The plate was centrifuged for 1 minute at 1,000 rpm to synchronize phagocytosis. After 4 hours, the media was removed and cells were stained with Hoechst (20 uM, Cayman Chemical 33342) to mark nuclei and propidium iodide (1ug/mL, Acros Organics 440300250) to mark cell death. Plates were then imaged on an Cellomics ArrayScan™ VTI Objective Module at 20x 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. (# of propidium iodide events)/(# of total cells) = % of dead macrophages.

Biofilm Formation

Biofilm formation was assessed as previously described [56] . Briefly, overnights of *C. albicans* isolates were incubated and diluted to 0.5 OD_{600}. In a 96-well plate, isolates were added to 200 µL of RPMI media, covered with a Breathe-Easy film, and incubated at 37°C shaking at 250 rpm. After 90 minutes, RPMI media was removed and wells were washed once with 200 µL of 1X PBS and 200 µL 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 hours before a final read on the plate reader at OD_{600}.

CHEF gel electrophoresis

Agarose plugs for CHEF gel electrophoresis were prepared as in Selmecki 2005 [80]. Electrophoresis was performed as in Chibana [81], with minor modifications to improve chromosome separation. Plugs were run on 0.9% megabase agarose (Biorad 1613108), with run conditions of 60 to 300 sec, 4.5 V/cm, 120 angle for 36 hr, followed by 720 to 900 sec, 2.0 V/cm, 106 for 24 hr. Sizes were determined based on approximations from Selmecki and a CHEF DNA Size Marker, 0.2–2.2 Mb, *S. cerevisiae* Ladder (Biorad 1703605).

792 **High Molecular Weight Genomic DNA extraction and sequencing**

793 High molecular weight (>20 kb fragment) genomic DNA was prepared using the
794 Qiagen MagAttract HMW DNA Kit (Qiagen 67563), following the manufacturer's
795 protocol, with minor modification. Fungal cell walls were first digested using Zymolyase
796 (Zymo E1005) before DNA extraction. TELL-seq [48] libraries were prepared at the
797 University of Michigan Advanced Genomics facility. Libraries were sequenced on an
798 Illumina NovaSeq SP 300 cycle flow cell. All sequences are uploaded to NCBI SRA at
799 PRJNA875200.

800

801 **Sequence Variation Analysis**

802 To identify single nucleotide variants (SNVs), we mapped 435 read libraries to
803 the SC5314 reference genome with BWA- MEM [49]. We used various samtools utilities
804 to convert alignment files, remove PCR duplicates, and assign read groups [49].
805 Variants were called with GATK HaplotypeCaller and individual sample GVCFs were
806 combined with GATK CombineGVCFs. Finally, we genotyped our combined GVCF
807 across samples and loci with GATK GenotypeGVCFs [82]. GATK identified 851,355
808 total variable loci. Of the 741,027 diallelic loci, GATK identified 60,626 indel-derived
809 alleles (8.18%), and 680,401 SNV alleles (91.82%). To refine our raw variant set to only
810 include high-confidence SNVs, we removed low quality variant calls at the (i) individual
811 genotype call, (ii) whole locus, and (iii) whole sample levels. First, we removed
812 individual sample genotype calls if their genotype quality (i.e., GQ) was < 0.99 (out of
813 1.0) or the results of the map quality rank sum test (i.e., MQRS) differed from 0.0 (i.e.,
814 the perfect score). Second, we removed all variant calls with more than one 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 4 entire samples. 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 one member of our set of newly sequenced strains.

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., [34,51,52]). Following removal of these samples, we were left with 324 sample SNV profiles (280 published, and 44 new genomes strains) (Table S3). To generate an input matrix for distance calculation and NJ clustering, we coded genotypes as homozygous for the reference allele (R), heterozygous (H), homozygous for the alternative allele (A), or missing (N). Raw distance between two 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*...

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

836 …where *nSNPs* is the total number of high-quality SNPs in our dataset (i.e.,
837 112,136), *nN* is the number of uncalled genotypes between each pair (i.e., coded as N),
838 GT_{ix} is the genotype of sample *i* at locus *x*, and GT_{jx} is genotype of sample *j* at locus *x*.
839 The resulting distance matrix was clustered with the *nj* function in phytools in R [83]. We
840 visualized the resulting tree structure and associated data with ggtree in R [84–86].
841
842 **Structural Variant Analysis from TELL-seq**
843 To identify structural variation genome wide in the condensed set of *C. albicans*
844 isolates, we mapped the read libraries to the SC5314 reference genome with BWA-
845 MEM [49]. We visualized the read depth across the genome by breaking the reference
846 genome into 500bp bins. Each read was assigned to a single bin based on the first
847 coordinate in the reference genome to which the read aligned. The read counts were
848 normalized across the samples by the average read coverage for each sample. The
849 value for each bin ($v_{b,s}$) for each sample was calculated using the following formula:
850

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

851
852 Additionally, we performed structural variant calling using lumpy-sv [87] and genotyping
853 with svtyper [88] using smoove (version 0.2.5). We visualized the variant calls by
854 breaking the reference genome into 500bp bins and identified the number of variant
855 calls of a give type which overlapped with each bin.
856 To identify potential translocations from the TELL-seq data, we used the Tell-Link
857 pipeline to assemble the read libraries into contigs and aligned the assembled contigs to

the SC5314 reference genome using MiniMap2 [89]. 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 two 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 Table S5.

ZMS1 strain construction

To make the *ZMS1* allele swap strains, we used the NEBuilder HiFi assembly kit to clone both alleles of *ZMS1* from the 814-168 strain background into the pUC19 cloning vector, along with the nourseothricin resistance cassette, to generate plasmid 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 814-183 background using a transient CRISPR approach [90]. 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 [91] using primers oTO1215 and oTO1216, and transformed into

the 814-168 and 814-183 strain backgrounds using a transient CRISPR approach [90]. Integration was tested using oTO5 and oTO1218 and loss of the wild-type *ZMS1* gene was tested using oTO736 and oTO1218.

ACKNOWLEDGEMENTS:

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.

FUNDING:

Funding for this project included mCubed collaborative grant to TRO and TYJ, NIH grant NIH KAI137299 (NIAID) to TRO, NIAID T32 AI007528 to FMA, NIH 1F31HG010569-01 to AMW, NIH T32GM007544 to KRA, University of Michigan Postdoctoral Pioneer Program to MJM. TYJ is a fellow of the Canadian Institute for Advanced Research program Fungal Kingdom: Threats & Opportunities.

Supplemental Tables:

Supplemental table 1: Media and growth conditions.

Supplemental table 2: Full collection growth curves (Tab 1) and invasion scores (Tab 2).

Supplemental table 3: SRA accession numbers for strains included in phylogenetic analysis.

Supplemental table 4: Condensed set strain information and quantitative phenotypes.

Supplemental table 5: Primers used in this study.

Supplemental Figure Legends:

Supplemental Figure 1: 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) Comparison between oral and fecal strains for agar invasion.

Supplemental Figure 2: 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 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.

Supplemental Figure 2: Biofilm formation. Each condensed set isolate was tested for its ability to form a biofilm on a plastic surface. Asterisks indicate $P < 0.05$ (*) compared with SC5314, one-way ANOVA compared with SC5314, with Dunnett's post-hoc test for multiple corrections.

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

Supplemental Figure 5: UMAP embedding.

Non-linear 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 Table S4.

Supplemental Figure 6: 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 $2x10^6$ cells/mL before inoculating 20 *G. mellonella* larvae per strain with 50 µL 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.

971 **References:**

972

973

974 1. Mukherjee PK, Sendid B, Hoarau G, Colombel J-F, Poulain D, Ghannoum MA.
975 Mycobiota in gastrointestinal diseases. Nat Rev Gastroenterol Hepatol. 2015;12:
976 77–87.

977 2. Polvi EJ, Li X, O'Meara TR, Leach MD, Cowen LE. Opportunistic yeast pathogens:
978 reservoirs, virulence mechanisms, and therapeutic strategies. Cell Mol Life Sci.
979 2015;72: 2261–2287.

980 3. Li J, Chen D, Yu B, He J, Zheng P, Mao X, et al. Fungi in Gastrointestinal Tracts of
981 Human and Mice: from Community to Functions. Microb Ecol. 2018;75: 821–829.

982 4. Hoffmann C, Dollive S, Grunberg S, Chen J, Li H, Wu GD, et al. Archaea and fungi
983 of the human gut microbiome: correlations with diet and bacterial residents. PLoS
984 One. 2013;8: e66019.

985 5. Gunsalus KTW, Tornberg-Belanger SN, Matthan NR, Lichtenstein AH, Kumamoto
986 CA. Manipulation of Host Diet To Reduce Gastrointestinal Colonization by the
987 Opportunistic Pathogen Candida albicans. mSphere. 2016;1.
988 doi:10.1128/mSphere.00020-15

989 6. Noble SM, Gianetti BA, Witchley JN. Candida albicans cell-type switching and
990 functional plasticity in the mammalian host. Nat Rev Microbiol. 2017;15: 96–108.

991 7. Alves R, Barata-Antunes C, Casal M, Brown AJP, Van Dijck P, Paiva S. Adapting
992 to survive: How Candida overcomes host-imposed constraints during human
993 colonization. PLoS Pathog. 2020;16: e1008478.

994 8. Ene IV, Brunke S, Brown AJP, Hube B. Metabolism in fungal pathogenesis. Cold
995 Spring Harb Perspect Med. 2014;4: a019695.

996 9. Polke M, Hube B, Jacobsen ID. Candida survival strategies. Adv Appl Microbiol.
997 2015;91: 139–235.

998 10. Zhai B, Ola M, Rolling T, Tosini NL, Joshowitz S, Littmann ER, et al. High-
999 resolution mycobiota analysis reveals dynamic intestinal translocation preceding
1000 invasive candidiasis. Nat Med. 2020. doi:10.1038/s41591-019-0709-7

1001 11. Yan L, Yang C, Tang J. Disruption of the intestinal mucosal barrier in Candida
1002 albicans infections. Microbiol Res. 2013;168: 389–395.

1003 12. Alenazy H, Alghamdi A, Pinto R, Daneman N. Candida colonization as a predictor
1004 of invasive candidiasis in non-neutropenic ICU patients with sepsis: A systematic
1005 review and meta-analysis. Int J Infect Dis. 2021;102: 357–362.

13. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. Sci Transl Med. 2012;4: 165rv13-165rv13.

14. 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.

15. 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.

16. Hittinger CT. Saccharomyces diversity and evolution: a budding model genus. Trends Genet. 2013;29: 309–317.

17. Huang MY, Woolford CA, May G, McManus CJ, Mitchell AP. Circuit diversification in a biofilm regulatory network. PLoS Pathog. 2019;15: e1007787.

18. 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.

19. 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. doi:10.1128/mBio.00586-21

20. McDonough L, Mishra AA, Tosini N, Kakade P, Penumutchu S, Liang S-H, et al. Candida albicans Isolates 529L and CHN1 Exhibit Stable Colonization of the Murine Gastrointestinal Tract. bioRxiv. 2021. p. 2021.06.27.450080. doi:10.1101/2021.06.27.450080

21. Soll DR, Pujol C. Candida albicans clades. FEMS Immunol Med Microbiol. 2003;39: 1–7.

22. 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.

23. Odds FC. Molecular phylogenetics and epidemiology ofCandida albicans. Future Microbiology. 2010. pp. 67–79. doi:10.2217/fmb.09.113

24. 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.

25. 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.

26. Bougnoux 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.

27. Odds FC, Bougnoux M-E, Shaw DJ, Bain JM, Davidson AD, Diogo D, et al. Molecular phylogenetics of Candida albicans. Eukaryot Cell. 2007;6: 1041–1052.

28. 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.

29. Li X, Yan Z, Xu J. Quantitative variation of biofilms among strains in natural populations of Candida albicans. Microbiology. 2003;149: 353–362.

30. Slutsky B, Staebell M, Anderson J, Risen L. " White-opaque transition": a second high-frequency switching system in Candida albicans. Journal of …. 1987. doi:papers3://publication/uuid/02E4C532-97D3-4987-B50F-D8EAF9F2A6CF

31. 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.

32. Calderone RA, Fonzi WA. Virulence factors of Candida albicans. Trends Microbiol. 2001;9: 327–335.

33. Pappas PG, Lionakis MS, Arendrup MC, Ostrosky-Zeichner L, Kullberg BJ. Invasive candidiasis. Nat Rev Dis Primers. 2018;4: 18026.

34. 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.

35. 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.

36. Pierce JV, Kumamoto CA. Variation in Candida albicans EFG1 expression enables host-dependent changes in colonizing fungal populations. MBio. 2012;3: e00117-12.

37. 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.

38. 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. doi:10.1016/j.chom.2021.03.019

39. 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: e184-13.

40. 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.

41. 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.

42. Schönherr FA, Sparber F, Kirchner FR, Guiducci E, Trautwein-Weidner K, Gladiator A, et al. The intraspecies 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.

43. Magee BB, Magee PT. Electrophoretic karyotypes and chromosome numbers in Candida species. J Gen Microbiol. 1987;133: 425–430.

44. 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.

45. Tan CT, Xu X, Qiao Y, Wang Y. A peptidoglycan storm caused by β-lactam antibiotic's action on host microbiota drives Candida albicans infection. Nat Commun. 2021;12: 2560.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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. Journal of Antimicrobial Chemotherapy. 2020. doi:10.1093/jac/dkz537

51. Todd RT, Wikoff TD, Forche A, Selmecki A. Genome plasticity in Candida albicans is driven by long repeat sequences. Elife. 2019;8. doi:10.7554/eLife.45954

52. 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.

53. 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.

54. Forche A, Cromie G, Gerstein AC, Solis NV, Pisithkul T, Srifa W, et al. Rapid Phenotypic and Genotypic Diversification After Exposure to the Oral Host Niche in Candida albicans. Genetics. 2018;209: 725–741.

55. Smith AC, Hickman MA. Host-Induced Genome Instability Rapidly Generates Phenotypic Variation across Candida albicans Strains and Ploidy States. mSphere. 2020;5. doi:10.1128/mSphere.00433-20

56. 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.

57. 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.

58. Santana DJ, Anderson FM, O'Meara TR. Monitoring Inflammasome Priming and Activation in Response to Candida albicans. Curr Protoc Microbiol. 2020;59: e124.

59. 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.

60. 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.

61. Wellington M, Koselny K, Krysan DJ. Candida albicans morphogenesis is not required for macrophage interleukin 1β production. MBio. 2012;4: e00433-12.

62. Shankar M, Lo TL, Traven A. Natural Variation in Clinical Isolates of Candida albicans Modulates Neutrophil Responses. mSphere. 2020;5.

1150 doi:10.1128/mSphere.00501-20

1151 63. Gerwien F, Dunker C, Brandt P, Garbe E, Jacobsen ID, Vylkova S. Clinical
1152 Candida albicans Vaginal Isolates and a Laboratory Strain Show Divergent
1153 Behaviors during Macrophage Interactions. mSphere. 2020;5.
1154 doi:10.1128/mSphere.00393-20

1155 64. O'Meara TR, O'Meara MJ. DeORFanizing Candida albicans Genes using
1156 Coexpression. mSphere. 2021;6. doi:10.1128/mSphere.01245-20

1157 65. Homann OR, Dea J, Noble SM, Johnson AD. A phenotypic profile of the Candida
1158 albicans regulatory network. PLoS Genet. 2009;5: e1000783.

1159 66. Brennan M, Thomas DY, Whiteway M, Kavanagh K. Correlation between virulence
1160 of Candida albicans mutants in mice and Galleria mellonella larvae. FEMS Immunol
1161 Med Microbiol. 2002;34: 153–157.

1162 67. Dunn MJ, Woodruff AL, Anderson MZ. The Galleria mellonella Waxworm Infection
1163 Model for Disseminated Candidiasis. J Vis Exp. 2018. doi:10.3791/58914

1164 68. Hill JA, O'Meara TR, Cowen LE. Fitness trade-offs associated with the evolution of
1165 resistance to antifungal drug combinations. Cell Rep. 2015;10: 809–819.

1166 69. Martin RM, Bachman MA. Colonization, Infection, and the Accessory Genome of
1167 Klebsiella pneumoniae. Front Cell Infect Microbiol. 2018;8: 4.

1168 70. Allen JP, Snitkin E, Pincus NB, Hauser AR. Forest and Trees: Exploring Bacterial
1169 Virulence with Genome-wide Association Studies and Machine Learning. Trends
1170 Microbiol. 2021;29: 621–633.

1171 71. Gerstein AC, Jackson KM, McDonald TR, Wang Y, Lueck BD, Bohjanen S, et al.
1172 Identification of Pathogen Genomic Differences That Impact Human Immune
1173 Response and Disease during Cryptococcus neoformans Infection. MBio. 2019;10.
1174 doi:10.1128/mBio.01440-19

1175 72. Beale MA, Sabiiti W, Robertson EJ, Fuentes-Cabrejo KM, O'Hanlon SJ, Jarvis JN,
1176 et al. Genotypic Diversity Is Associated with Clinical Outcome and Phenotype in
1177 Cryptococcal Meningitis across Southern Africa. PLoS Negl Trop Dis. 2015;9:
1178 e0003847.

1179 73. Ene IV, Bennett RJ, Anderson MZ. Mechanisms of genome evolution in Candida
1180 albicans. Curr Opin Microbiol. 2019;52: 47–54.

1181 74. Wakade RS, Huang M, Mitchell AP, Wellington M, Krysan DJ. Intravital imaging of
1182 Candida albicans identifies differential in vitro and in vivo filamentation phenotypes
1183 for transcription factor deletion mutants. bioRxiv. 2021. p. 2021.05.10.443530.
1184 doi:10.1101/2021.05.10.443530

75. 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.

76. 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.

77. 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.

78. 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.

79. Sprouffske K, Wagner A. Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. BMC Bioinformatics. 2016;17: 172.

80. Selmecki A, Bergmann S, Berman J. Comparative genome hybridization reveals widespread aneuploidy in Candida albicans laboratory strains. Mol Microbiol. 2005;55: 1553–1565.

81. Chibana H, Beckerman JL, Magee PT. Fine-resolution physical mapping of genomic diversity in Candida albicans. Genome Res. 2000;10: 1865–1877.

82. 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.

83. Revell LJ. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol. John Wiley & Sons, Ltd (10.1111); 2012.

84. Yu G. Using ggtree to Visualize Data on Tree-Like Structures. Curr Protoc Bioinformatics. 2020;69: e96.

85. 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.

86. 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.

87. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biol. 2014;15: R84.

88. 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.

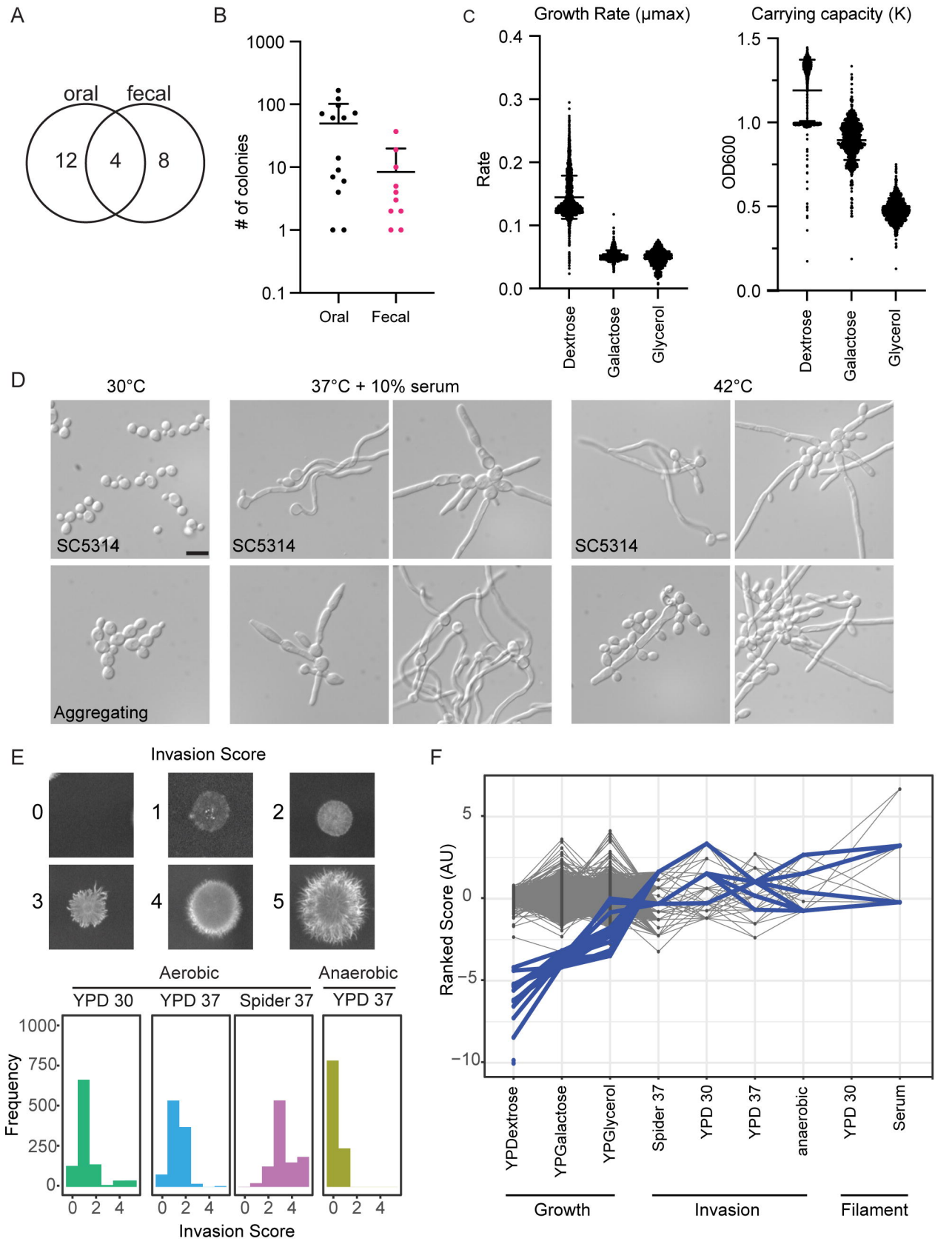
89. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34: 3094–3100.

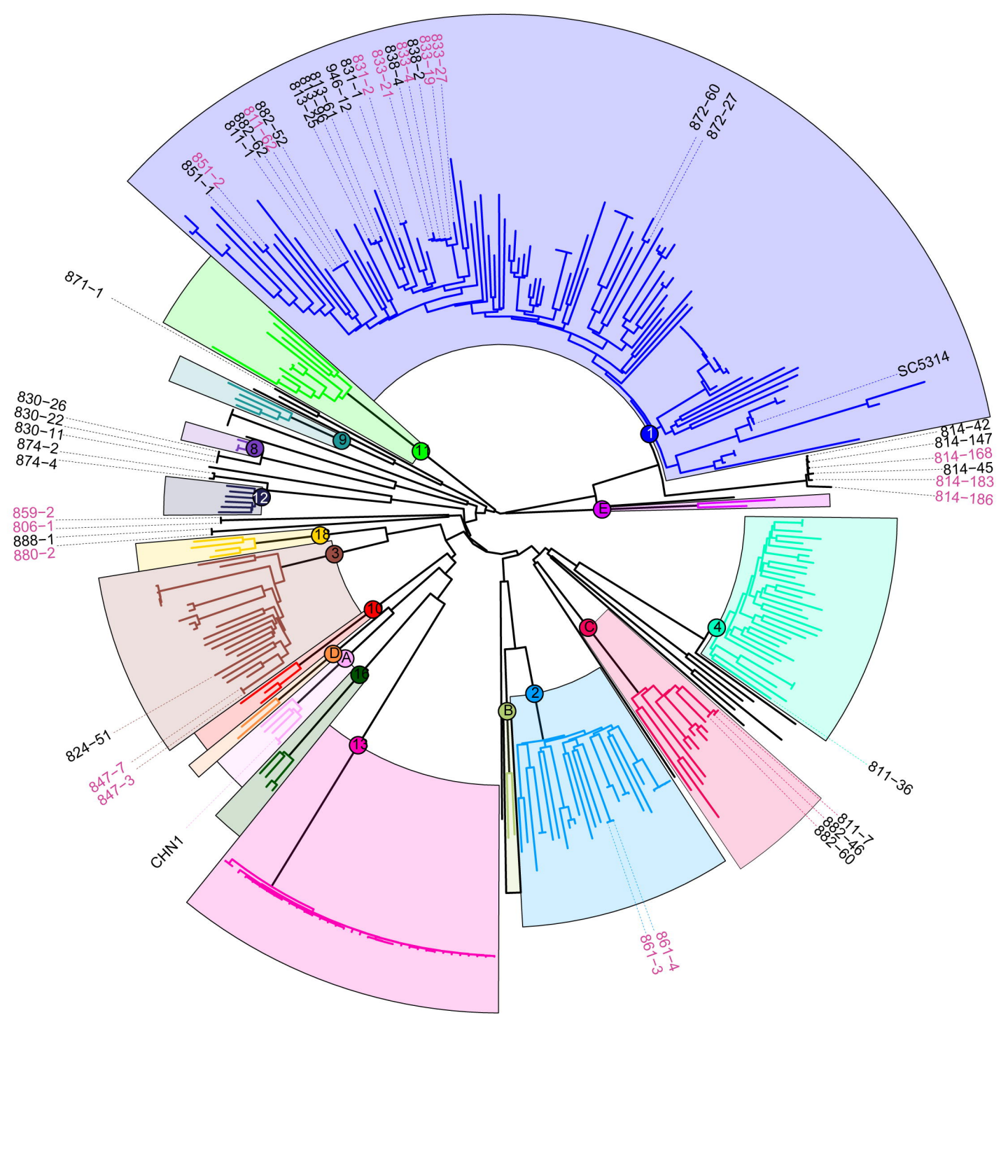
90. Veri AO, Miao Z, Shapiro RS, Tebbji F, O'Meara TR, Kim SH, et al. Tuning Hsf1 levels drives distinct fungal morphogenetic programs with depletion impairing Hsp90 function and overexpression expanding the target space. PLoS Genet. 2018;14: e1007270.

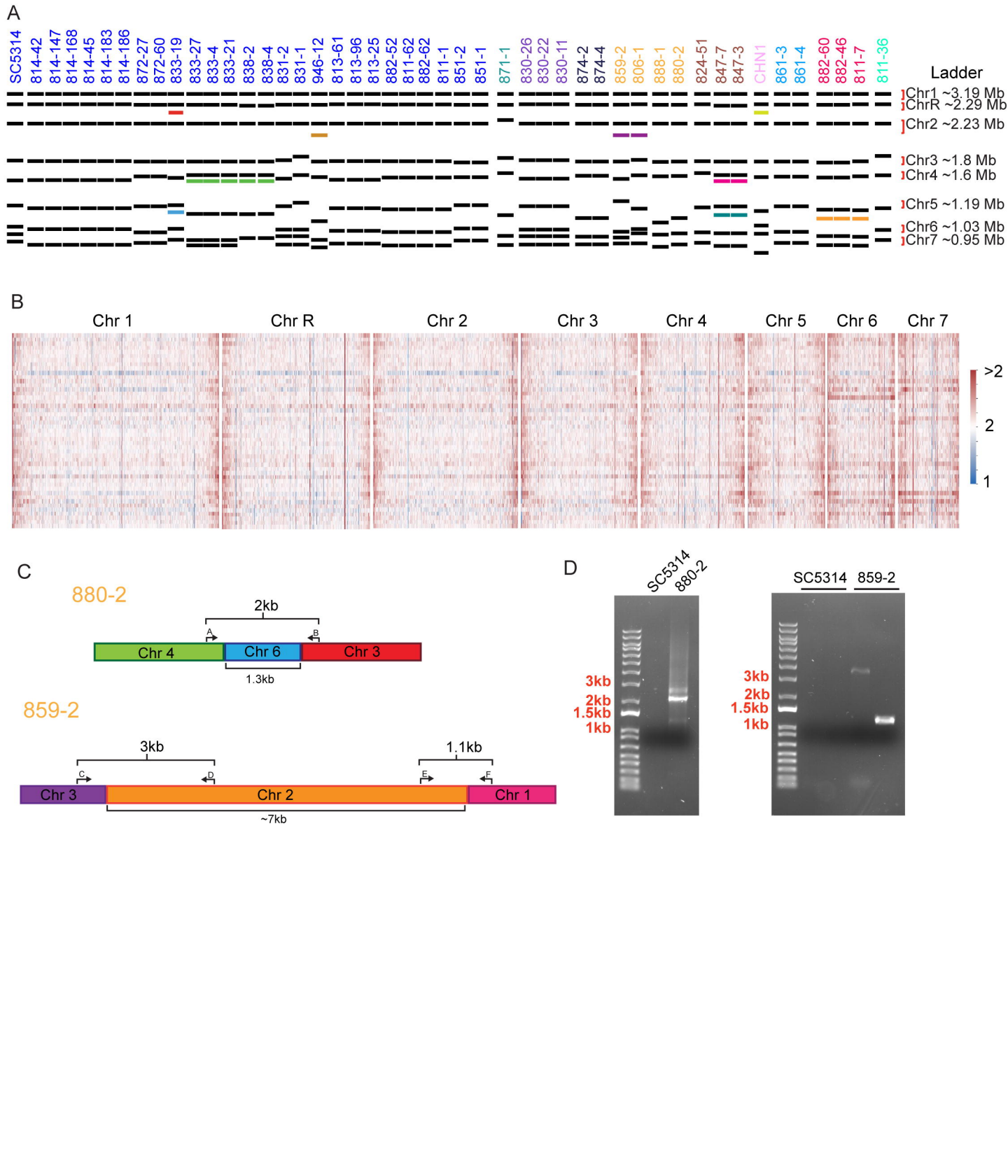
91. Shen J, Guo W, Köhler JR. CaNAT1, a heterologous dominant selectable marker for transformation of Candida albicans and other pathogenic Candida species. Infect Immun. 2005;73: 1239–1242.

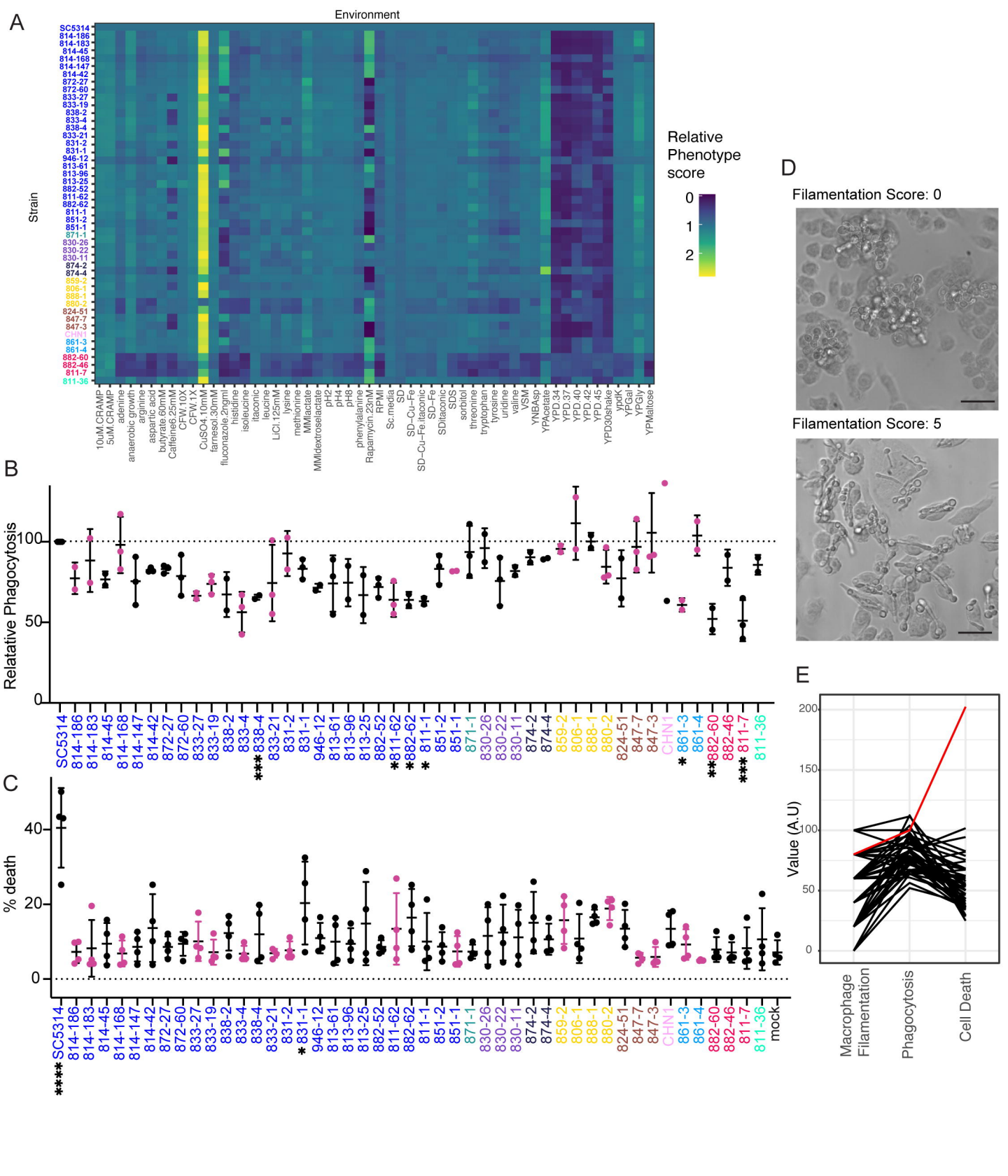
92. Lee DW, Hong CP, Kang HA. An effective and rapid method for RNA preparation from non-conventional yeast species. Anal Biochem. 2019;586: 113408.

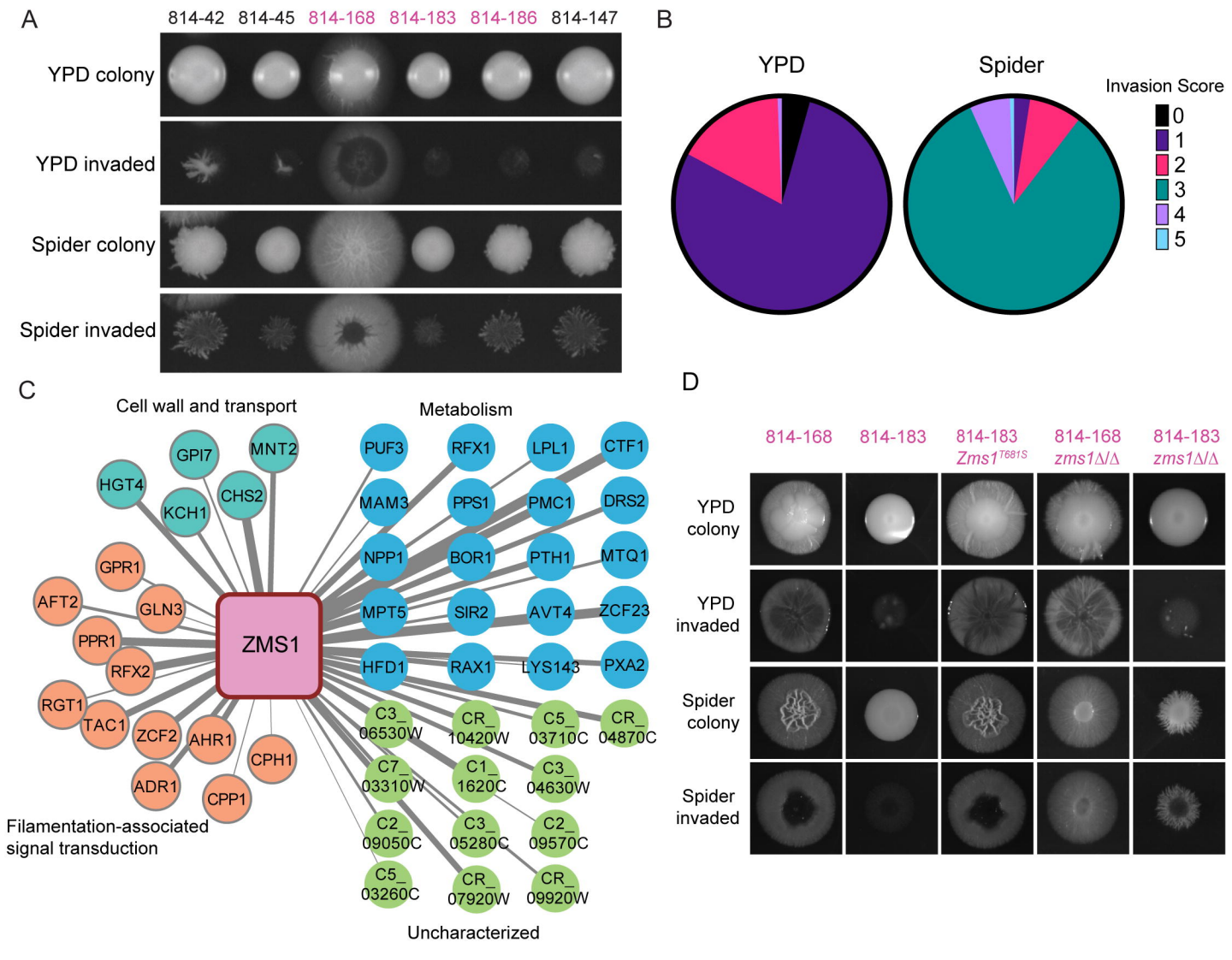
93. Aranda PS, LaJoie DM, Jorcyk CL. Bleach gel: a simple agarose gel for analyzing RNA quality. Electrophoresis. 2012;33: 366–369.

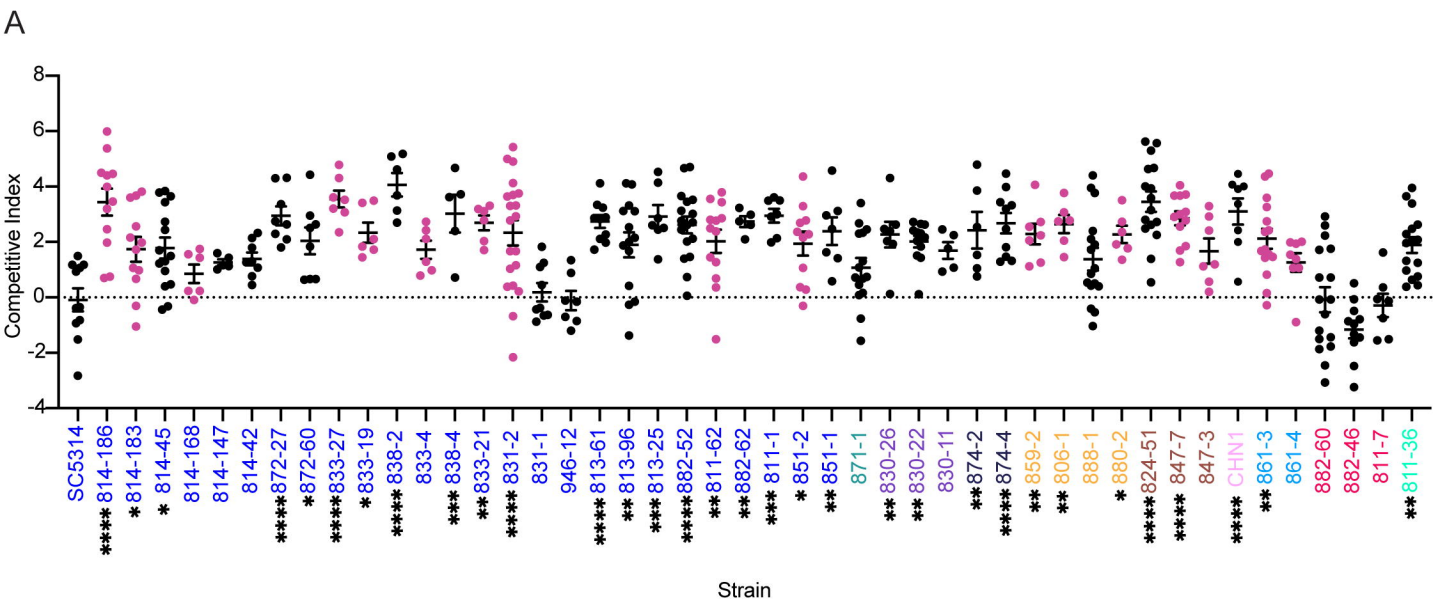

A
oral
fecal
12
4
8
B
# of colonies
1000
100
10
1
0.1
Oral
Fecal
C
Growth Rate (μmax)
Rate
0.4
0.3
0.2
0.1
0.0
Dextrose
Galactose
Glycerol
Carrying capacity (K)
OD600
1.5
1.0
0.5
0.0
Dextrose
Galactose
Glycerol
D
30°C
37°C + 10% serum
42°C
SC5314
SC5314
SC5314
Aggregating
E
Invasion Score
0
1
2
3
4
5
Aerobic
Anaerobic
YPD 30
YPD 37
Spider 37
YPD 37
Frequency
1000
750
500
250
0
0 2 4
0 2 4
0 2 4
0 2 4
Invasion Score
F
Ranked Score (AU)
5
0
−5
−10
YPDextrose
YPGalactose
YPGlycerol
Spider 37
YPD 30
YPD 37
anaerobic
YPD 30
Serum
Growth
Invasion
Filament

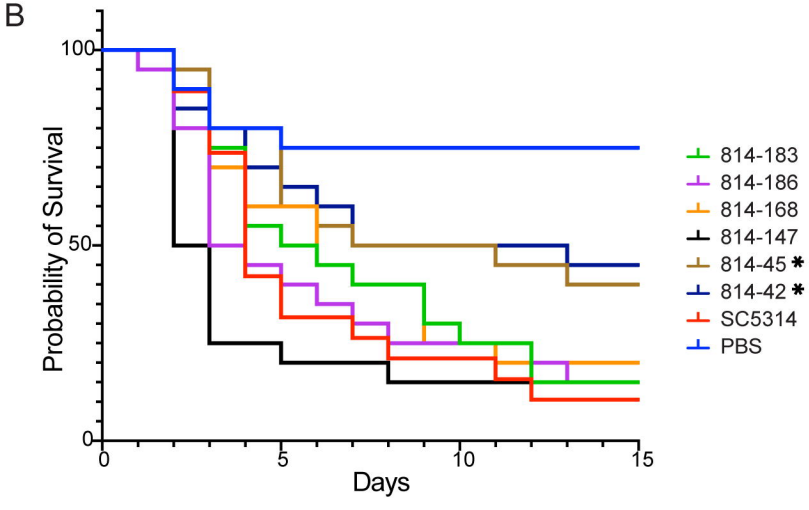

833-27
833-29
838-2
833-4
838-4
833-21
831-2
831-1
946-12
813-961
813-25
882-52
882-62
811-1
851-2
851-1
872-60
872-27
871-1
SC5314
814-42
814-147
814-168
814-45
814-183
814-186
830-26
830-22
830-11
874-2
874-4
859-2
806-1
888-1
880-2
824-51
847-7
847-3
CHN1
811-36
811-7
882-46
882-60
861-4
861-3
1
2
3
4
8
9
10
11
12
13
15
18
A
B
C
D
E


A
SC5314
814-42
814-147
814-168
814-45
814-183
814-186
872-27
872-60
833-19
833-27
833-4
833-21
838-2
838-4
831-2
831-1
946-12
813-61
813-96
813-25
882-52
811-62
882-62
811-1
851-2
851-1
871-1
830-26
830-22
830-11
874-2
874-4
859-2
806-1
888-1
880-2
824-51
847-7
847-3
CHN1
861-3
861-4
882-60
882-46
811-7
811-36
Ladder
Chr1 ~3.19 Mb
ChrR ~2.29 Mb
Chr2 ~2.23 Mb
Chr3 ~1.8 Mb
Chr4 ~1.6 Mb
Chr5 ~1.19 Mb
Chr6 ~1.03 Mb
Chr7 ~0.95 Mb
B
Chr 1
Chr R
Chr 2
Chr 3
Chr 4
Chr 5
Chr 6
Chr 7
>2
2
1
C
880-2
2kb
A
B
Chr 4
Chr 6
Chr 3
1.3kb
859-2
3kb
1.1kb
C
D
E
F
Chr 3
Chr 2
Chr 1
~7kb
D
SC5314
880-2
3kb
2kb
1.5kb
1kb
SC5314
859-2
3kb
2kb
1.5kb
1kb


A
Environment
Strain
Relative Phenotype score
0
1
2
10uM.CRAMP
5uM.CRAMP
adenine
anaerobic growth
arginine
aspartic acid
butyrate.60mM
Caffeine6.25mM
CFW.10X
CFW.1X
CuSO4.10mM
farnesol.30mM
fluconazole.2ngml
histidine
isoleucine
itaconic
leucine
LiCl .125mM
lysine
methionine
MMlactate
MMldextroselactate
pH2
pH4
pH8
phenylalanine
Rapamycin.23nM
RPMI
Sc.media
SD
SD–Cu–Fe
SD–Cu–Fe.Itaconic
SD–Fe
SDItaconic
SDS
sorbitol
threonine
tryptophan
tyrosine
uridine
valine
VSM
YNBASp
YPAcetate
YPD.34
YPD.37
YPD.40
YPD.42
YPD.45
YPD30shake
ypdK
YPGal
YPGly
YPMaltose
B
Relatative Phagocytosis
100
50
0
C
% death
40
20
0
mock
D
Filamentation Score: 0
Filamentation Score: 5
E
Value (A.U)
200
150
100
50
0
Macrophage Filamentation
Phagocytosis
Cell Death


A
814-42
814-45
814-168
814-183
814-186
814-147
YPD colony
YPD invaded
Spider colony
Spider invaded
B
YPD
Spider
Invasion Score
0
1
2
3
4
5
C
Cell wall and transport
Metabolism
GPI7
MNT2
HGT4
CHS2
KCH1
GPR1
AFT2
GLN3
PPR1
RFX2
RGT1
TAC1
ZCF2
AHR1
CPH1
ADR1
CPP1
ZMS1
Filamentation-associated signal transduction
PUF3
RFX1
LPL1
CTF1
MAM3
PPS1
PMC1
DRS2
NPP1
BOR1
PTH1
MTQ1
MPT5
SIR2
AVT4
ZCF23
HFD1
RAX1
LYS143
PXA2
C3_ 06530W
CR_ 10420W
C5_ 03710C
CR_ 04870C
C7_ 03310W
C1_ 1620C
C3_ 04630W
C2_ 09050C
C3_ 05280C
C2_ 09570C
C5_ 03260C
CR_ 07920W
CR_ 09920W
Uncharacterized
D
814-168
814-183
814-183 Zms1T681S
814-168 zms1Δ/Δ
814-183 zms1Δ/Δ
YPD colony
YPD invaded
Spider colony
Spider invaded


A
Competitive Index
8
6
4
2
0
-2
-4
SC5314
**** 814-186
* 814-183
* 814-45
814-168
814-147
814-42
**** 872-27
* 872-60
**** 833-27
* 833-19
**** 838-2
833-4
*** 838-4
** 833-21
**** 831-2
831-1
946-12
**** 813-61
** 813-96
*** 813-25
**** 882-52
** 811-62
** 882-62
*** 811-1
* 851-2
** 851-1
871-1
** 830-26
** 830-22
830-11
** 874-2
**** 874-4
** 859-2
** 806-1
888-1
* 880-2
**** 824-51
**** 847-7
847-3
**** CHN1
** 861-3
861-4
882-60
882-46
811-7
** 811-36
Strain


B
Probability of Survival
100
50
0
0
5
10
15
Days
814-183
814-186
814-168
814-147
814-45 *
814-42 *
SC5314
PBS