Genotypic and phenotypic homogeneity of vaginal and rectal yeast isolates from recurrent vulvovaginal candidiasis

Abdul-Rahman Adamu Bukari[a], Rebekah Kukurudz[a], Alexia de Graaf[a*], Devin Habon[a*], Beamlak Manyaz[a*], Yana Syvolos[a], Aruni Sumanarathne[a], Vanessa Poliquin[b], Aleeza Gerstein[a,b,#]

[a]Department of Microbiology, Faculty of Science, University of Manitoba, Winnipeg, Manitoba, Canada

[b]Obstetrics, Gynecology and Reproductive Sciences, University of Manitoba, Winnipeg, Manitoba, Canada

[c]Department of Statistics, Faculty of Science, University of Manitoba, Winnipeg, Manitoba, Canada

Running head: Low variation in recurrent vulvovaginal candidiasis

* Alexia de Graaf, Devin Habon, and Beamlak Manyaz contributed equally to this work and are listed alphabetically

Author Correspondence to Aleeza Gerstein, aleeza.gerstein@umanitoba.ca

Abstract

Vulvovaginal candidiasis is one of the most common vaginal and fungal infections. Many are successfully treated with antifungal drugs, but ~9% of cases will recur even with treatment (RVVC). We quantified the genotypic and phenotypic diversity of vaginal and rectal yeast isolates from four individuals with a history of recurrent vulvovaginal candidiasis during a symptomatic relapse. One participant had a *Nakaseomyces glabrata* infection while the other three had *Candida albicans*. We used whole genome sequencing to place the isolates into a global phylogenic context and precisely quantify standing genetic variation within RVVC yeast populations. The *C. albicans* isolates statistically clustered close together and closer than expected to other vaginal isolates in a subgroup of clade 1. In all participants, we found that vaginal and rectal isolates are monophyletic and phylogenetically overlapping, consistent with frequent migration between sites. We found very low levels of genotypic and phenotypic diversity and few phenotypic outliers, consistent with rapid population expansion. This provides a view of the within-host isolate variation that is inconsistent with a rectal source population for vaginal reinfection and a generally small effective population size over time in RVVC.

Importance

Recurrent vaginal yeast infections are relatively common, yet many open questions remain about the infecting fungal population. We examined the genetic and phenotypic diversity within vaginal and fungal populations from four individuals with a history of recurrent yeast infections experiencing symptoms. Three of the participants had a Candida albicans infection (the most common causative species) while the fourth had a Nakaseomyces glabrata infection (the second most common and increasingly implicated). This is the first study to use whole genome sequencing to capture genotypic diversity within recurrent yeast infections precisely. We found that vaginal and rectal isolates were overlapping, indicating frequent migration of individuals between the two sites. The overall level of genetic variation within the populations was very low and nearly all isolates had very similar phenotypes. These results are consistent with rapid population expansion during symptomatic infection and inconsistent with a rectal source population leading to vaginal reinfection.

Introduction

Vulvovaginal candidiasis (VVC, colloquially "yeast infection") affects approximately 50-75% of people with vaginas at some point in their lives (1–3). Yeast infections cause significant financial, physical, and emotional strain (4–7). People in the United States spend over half a billion dollars on prescription and over-the-counter medications to treat yeast infections each year (8), and the yearly lost productivity in high-income countries is estimated at $14 billion USD (9). Although the majority of infections are successfully cleared, an estimated ~5-9% of people who have had a yeast infection experience recurrent infections (RVVC), defined as three or more episodes a year (7, 10). The health system cost is estimated at more than $130 million USD per year (9), which is predicted to increase by more than $20 million by 2030 (9). Even with weekly drug treatment, 57% of people with RVVC experience a recurrence within a year, with an 8.4-month median time to symptomatic recurrence, compared to 1.9 months in an untreated control group (11). Many diverse factors are known to trigger infection or predispose individuals to VVC, including hygiene practices, the vaginal microbiota, and host genetics and immunology (12, and references within). However, approximately half of all people with RVVC have no identifiable risk factors (13), signifying the need for additional studies on the biological basis of this chronic condition.

Candida albicans is the most prevalent species isolated from VVC and RVVC infections (collectively, R/VVC), identified from swab cultures in 50-90% of cases (14–18). Other fungal yeast species are also isolated, with *Nakaseomyces glabrata* (formerly *Candida glabrata*) and *Candida tropicalis* increasingly implicated (19). Here we collectively refer to this group of species using the colloquial term "yeast," which reflects a shared morphology, rather than the historical and still common "*Candida* spp." given our current understanding of their phylogenetic relationships and recent official renaming of *N. glabrata* (20, 21). To be consistent with the literature on these conditions, we will continue to use the R/VVC abbreviations, noting that "candidiasis" does not reflect the updated genera names of the diverse species involved.

Understanding the etiology of R/VVC is complicated in part since yeast exists as a common commensal member of the vaginal microbiota without causing symptomatic VVC. Identification of vaginal yeast in the absence of symptoms does not warrant treatment. Approximately 70% of healthy women were found to have vaginal yeast at a single time point through an amplicon-based (i.e., culture-free) study (22). Note that this is much higher than the 20-30% asymptomatic commensal colonization estimate frequently cited from culture-based methods (e.g., 23). Furthermore, self-diagnosis due to symptoms has also been shown to be problematic, as other gynecological infections and dermatologic conditions have overlapping

81 symptoms (24). Genetic typing and phylogenetic strain analyses in *C. albicans* have repeatedly
82 found no strict phylogenetic differentiation between the commensal strains and those that cause
83 symptomatic disease, although some phylogentic clades are potentially over-represented by
84 commensal strains and strains causing superficial disease (25, 26).
85 There are limited treatments available for fungal infections in general, including R/VVC.
86 The two primary options for induction or maintenance therapy in Canada are azole drugs or
87 boric acid (17), which seem to yield similar recurrence rates (27, 28), and references within).
88 Isolates from many non-*C. albicans* species have intrinsic resistance to azoles (i.e., have
89 natural resistance even in the absence of prior drug exposure) (29), and although still relatively
90 rare, azole-resistant isolates of *C. albicans* have also been found (30, 31). This suggests that
91 conventional and first-line drugs typically used for VVC are becoming less effective (32). In
92 addition to drug resistance, drug tolerance, the ability of drug-susceptible populations to grow
93 slowly in the presence of high levels of fungistatic drugs, has also emerged as a trait that varies
94 among different fungal species and isolates (33, 34). Tolerance may be implicated in the
95 propensity to cause fungal disease in other infection contexts (35) but has not previously been
96 examined in the context of R/VVC.
97 Relapse in RVVC (i.e., return of symptoms after previous successful treatment) could
98 theoretically be due to either incomplete vaginal eradication of the yeast population between
99 symptomatic infections, or complete vaginal eradication followed by re-colonization from an
100 endogenous or exogenous population (36). There are decades of studies that have sought to
101 understand the etiology of RVVC, as the answer has potential implications for improving
102 treatment and reducing or eliminating recurrence. Re-colonization from a sexual partner has
103 been proposed as a possible mechanism, yet recurrence also occurs in non-sexually active
104 individuals (37), and a review of eight randomized clinical trials found that treating asymptomatic
105 sexual partners did not influence cure or recurrence rates (38). The GI tract has also been
106 suggested as a potential endogenous source population. A study in 1979 that treated RVVC
107 patients with oral nystatin with the goal of reducing a resident GI population found that it did not
108 decrease the time to recurrence (39). Furthermore, studies that have looked at yeast
109 colonization of the GI tract through feces or rectal swabs during an active infection find that not
110 all participants are culture-positive (39–45). However, this does not necessarily preclude that a
111 small GI population is present in all individuals (below the culture detection limit in some), which
112 could act as an endogenous reintroduction source under the right host conditions.
113 Examining the diversity of strains at different body sites and among recurrent infections
114 can potentially differentiate between relapse scenarios. If vaginal isolates have different

115 genotypes at different time points, yet the diversity is present in GI tract isolates, that would be
116 consistent with reintroduction. If vaginal isolates sampled at different time points consistently
117 have the same genotype that is not present in the GI tract, this would be consistent with
118 incomplete eradication. Looking at the level of diversity ("standing genetic variation") within a
119 single time point during an active symptomatic infection prior to drug treatment can also help us
120 to understand the evolutionary dynamics of the yeast populations, in terms of migration rate
121 between sites, population size during and between active infections, and the efficacy of
122 selection.

123 The earliest studies examined diversity by looking at 'resistogram' or phenotypic profiles
124 of strains, by typing strains based on their ability to grow in a panel of diverse chemical
125 challenges (45–47). This was followed by studies using many different methods of genomic
126 profiling to assess relatedness among strains, such as rDNA restriction fragment length (48),
127 DNA/PCR fingerprinting (40, 49–51), CHEF gel analysis (52), microsatellite loci (53–56), and
128 randomly amplified polymorphic DNA (RAPD) (51, 57–59). The highest genomic resolution and
129 an ability to accurately map phylogenetic relationships among divergent strains come from
130 recent studies that have used multilocus sequence typing (MLST) (18, 60, 61) and one very
131 small study that employed whole genome sequencing (WGS) (62). Considering this whole body
132 of work, the overall picture does not change drastically with the method employed. The
133 overarching results are consistent with a role for the maintenance of genotypes ("relapse," i.e.,
134 the same or very similar genotype is recovered from different body sites and between
135 successive symptomatic episodes), alongside an opportunity for novel genotypes to arise
136 ("reinfection," sometimes referred to as strain turnover; i.e., there is diversity in genotypes
137 sampled between subsequent symptomatic episodes). However, the source of genetic novelty
138 remains unanswered, and previous studies do not necessarily rule out a role for standing
139 genetic variation to have been consistently present yet undetected, i.e., relapse masquerading
140 as reinfection.

141 Here we build on previous work by using high-throughput phenotyping and WGS. We
142 quantify and compare the diversity of many vaginal and rectal isolates from four individuals with
143 a history of RVVC in Winnipeg, Canada. Three of the participants had *C. albicans* infections and
144 one had *N. glabrata*. We collectively refer to these isolates as the "THRIVE-yeast" isolates
145 following the name of our local umbrella research program to study The Host-microbial
146 Relationships and Immune function in different Vaginal Environments ("THRIVE",
147 http://www.mthrive.ca). Colonies were acquired during active symptomatic infections prior to
148 participants taking antifungal medication for their current symptoms. From all individuals, we

149 found a complete phylogenetic overlap of vaginal and rectal isolates and low genotypic and
150 phenotypic diversity with very few phenotypic outlier isolates. Our results indicate low levels of
151 standing genetic variation during active symptomatic infection. They are consistent with a recent
152 population bottleneck followed by rapid population expansion and phylogenetic relationships
153 that demonstrate a high level of migration between the vaginal and rectal populations.

Results

THRIVE-yeast isolates in the global species phylogenies

Seventeen women with a history of RVVC were recruited from a yeast clinic in Winnipeg, Canada. Vaginal and rectal swabs were acquired at the clinic visit and used in colony isolation. Vaginal and rectal isolates were obtained from six participant swabs. For this study, we focused on the four participants who yielded at least 24 vaginal and 24 rectal yeast isolates. Two additional participants had a smaller Five additional participants had vaginal but not rectal isolates. The phylogenetic relationship of 32 THRIVE-yeast isolates (four vaginal and four rectal isolates from each of four participants) was evaluated in the context of previously published whole-genome sequenced (WGS) isolates from each species. Isolates were chosen blind with respect to phenotypes. We constructed a global *N. glabrata* phylogeny from an additional 526 predominantly bloodstream isolates using FastTree (63) . The phylogeny includes isolates from two previously published studies (64, 65) and an additional 416 isolates that had FASTQ files available to download on NCBI (Table S1). Excluding the THRIVE-yeast isolates, in the global phylogeny there are only thirteen isolates that are annotated as originating from the vagina, rectum, or stool, and they are dispersed throughout the phylogeny (Figure 1A). The isolates from the participant with an *N. glabrata* infection (YST6) are monophyletic (Figure 1A). They cluster with 68 additional blood isolates and three of unknown provenance from three different countries (66 from the United States, two from Canada, and three from Australia). It was recently shown that *N. glabrata* phylogenic clades from WGS data are consistent with MLST (ST) clusters (64). Consistent with this, an *in silico* MLST analysis found that 64 isolates in the cluster with YST6 isolates are ST16. Two of the other isolates in that cluster are ST60 and two are ST187. None of the other global isolates are ST16.

The global *C. albicans* phylogeny is composed of 182 isolates from a wide breadth of geographic and anatomical sites (25). Isolates from each of the three participants (YST7, TVY4, and TVY10) form monophyletic groups and fall within a phylogenetic subgroup that contains fifteen previously characterized isolates in clade 1, the most common clade (Figure 1B). Unlike in *N. glabrata*, MLST analysis in *C. albicans* yielding diploid sequence types (DST) is too variable to map to clade assignments (e.g., despite forming a subclade, the fifteen isolates each have a different DST) (25). Interestingly, TVY4 and TVY10 isolates cluster beside each other, while the YST7 isolates are most closely related to three vaginal isolates from other parts of the world (one each from Brazil, Morocco, and China). Five of the remaining twelve isolates in this clade 1 subgroup with YST7 were also isolated from the vagina. There is a statistical enrichment for vaginal isolates in this subgroup compared to the rest of the isolates in clade 1

188 (THRIVE-yeast isolates from each participant were counted as a single isolate; Fisher exact test
189 comparing 11 vaginal isolates out of 18 to 5 vaginal isolates out of 24, p = 0.011). If we discount
190 the 35 predominantly vaginal isolates in clade 13, which is now recognized as likely a separate
191 species (*C. africanus) (66, 67)*, clade 1 as a whole is also statistically overrepresented for
192 vaginal isolates compared to the *C. albicans* tree in general (16 vaginal isolates in clade 1 out of
193 42 total isolates, compared to 18 vaginal isolates out of 105 total isolates; Fisher exact test, p =
194 0.009). Thus, although vaginal isolates are located throughout the *C. albicans* phylogeny, they
195 are over-clustered relative to a neutral expectation that vaginal isolates are equally likely to be
196 found anywhere in the existing tree.
197
198 **Vaginal and rectal isolates are phylogenetically overlapping and closely-related**
199 We then assessed the fine-scale phylogenic relatedness of vaginal and rectal isolates
200 from each participant. The *N. glabrata* vaginal and rectal isolates from YST6 are
201 phylogenetically overlapping and have a low average nucleotide diversity (π = 0.00007) (Figure
202 2A). FastTree has previously been shown to work well at the species level (68), yet the
203 FastTree phylogenies provided insufficient resolution for the very closely related THRIVE-yeast
204 isolate populations. Thus, for each participant, we built a RAxML phylogeny. The YST6
205 phylogeny was based on 1334 informative sites from 198 730 alignment sites. Two branch
206 points had bootstrap supports greater than 0.8; in both cases grouping vaginal and rectal
207 isolates together. The vaginal and rectal isolates from the three participants with *C. albicans*
208 infections are also overlapping (Figure 2B-D) and with low nucleotide diversities (TYV4, TVY10:
209 π = 0.0031; YST7: π = 0.0033). The TVY4 phylogeny was based on 3555 informative sites
210 from 628 501 aligned sites; only one branch leading to four isolates (three vaginal and one
211 rectal) was well supported (Figure 2B). TVY10 (3779 informative sites from 628 679 alignment
212 sites; Figure 2C) and YST7 (3309 informative sites based on 628 958 alignment sites; Figure
213 2D) similarly had a low number of well-supporting branches that nevertheless grouped vaginal
214 and rectal isolates. The relationship between vaginal and rectal isolates in all four participants is
215 thus consistent with ongoing and frequent migration between the two populations, at least
216 during symptomatic infection when the vaginal yeast population is high.
217 To further quantify the level of genotypic variation among isolates from the same
218 participant, we examined the intersection of SNPs among the eight sequenced isolates using
219 the intervene R package (69). The overwhelming majority of SNPs that differentiated the
220 THRIVE-yeast isolates from the reference genomes were in all eight participant isolates (Figure
221 2, right panels). There were few unique SNPs in each isolate, with an average of 74 SNPs

222 among YST6 isolates, 375 SNPs in TVY4 isolates and 345 in YST7 isolates. Interestingly, four
223 TVY10 isolates (V10, R14, R12, R13) had no unique SNPs, with an average of 706 SNPs in the
224 others TVY10 isolates. Combined, the relatedness among isolates and low levels of standing
225 genetic variation seems consistent with a recent and rapid population expansion with ongoing
226 migration between both vaginal and rectal sites during the symptomatic infection phase.
227
228 **Little variation in copy number or loss of heterozygosity within isolates**
229 We compared copy number variation (CNV) and loss-of-heterozygosity (LOH) events among
230 THRIVE-yeast isolates and their closest relatives compared to the reference genomes using
231 Y_{MAP} (70). No CNVs were identified in any YST6 isolates (Figure 3A). A single ~50 kb CNV on
232 the right arm of chr3 from 1407495 - 1453045 bp was identified in all eight YST7 isolates
233 (Figure 3B). This CNV is also present in the closest relative to the YST7 isolates, strain 9518,
234 but is absent in the next closest-related strains (Niger8, B116 and M17) and the reference
235 SC5314. As Y_{MAP} visualizations are based on averages across 5000 bp sliding windows, we
236 sought to examine the region in finer detail. Coverage was measured from the BAM files using
237 the bedtools function genomecov with the -d option to compute the depth at each position of the
238 genome. Mapped coverage in this region was inconsistent with the profile of a typical CNV; the
239 majority of the region had only a slightly elevated copy number relative to the rest of the
240 genome (Figure 3C). There were two small (< 200 bp) regions that spiked up to ~6-fold and
241 ~14-fold coverage, the first internal to *ALS6* and the second to *ALS7* (Figure 3C). A third region
242 comprised of elevated coverage maps to another gene with close homology to other genes in
243 the genome, *CYP5*, a putative peptidyl-prolyl cis-trans isomerase (71). The region identified in
244 Y_{MAP} is thus likely to primarily reflect an error in mapping rather than a true CNV with potential
245 biological effect. No other CNVs or aneuploidies were identified in the other THRIVE-yeast
246 isolates.
247 Nearly all telomeres had at least one LOH event identified among the THRIVE-yeast *C.*
248 *albicans* isolates compared to the SC5314 reference (Figure 3B, Figure S1). No large interstitial
249 LOH events were identified. We observed a general trend where all isolates from the same
250 participant shared an LOH profile. TVY4 isolates share a complicated LOH region on the left
251 arm of chr2 that flips back and forth from ancestral haplotypes, as well as small LOH regions on
252 both the left and right arms of chr4. YST7 shares LOH blocks with its' three closely related
253 vaginal isolates (i.e., 9518, B116, M17) but not Niger8, an oral isolate: on the right arm of chr2
254 between the centromere to the telomere, on the left arm of chr5, and on the left arm of chrR.
255 The exception of shared LOH blocks among participant isolates was a single LOH event on the

256 left arm of chr1, present in three TVY10 isolates (R14, V11 and V13) that cluster together on the
257 phylogeny. Some LOH regions were also shared between isolates from different participants. All
258 isolates have an LOH block on the left arm of chr3 and the right arm of chr6. TVY10 and TVY4
259 isolates share a ~300 kb region on the right arm of chr1. An LOH region on chr7 beside the
260 ancestral chr7 right arm LOH block is also present in both TVY10 and TVY4, though it is larger
261 in TVY10. TVY10 and YST7 both have LOH blocks on the left arm of chr5, though the allelic
262 profile between them is different. The overarching picture that emerges is that LOH seems to be
263 a common occurrence, with evidence for both variation among isolates of the same participant
264 (as in TVY10) as well as the potential for parallelism among isolate populations, i.e., when the
265 region of LOH is shared but the allelic profile differs.

266 **Phenotypic variation**

267 To determine whether the low variation observed in the genotypic results was representative of
268 the larger isolate collection we acquired from each participant, we compared different
269 phenotypes among a larger number of isolates. We focused our initial effort on YST6 (*N.*
270 *glabrata*) and YST7 (*C. albicans*), with the intention of expanding to the other two *C. albicans*
271 populations if phenotypic results deviated from the genotypic analyses.

272 The average growth rate for YST6 isolates was higher than YST7 for vaginal and rectal
273 isolates in Roswell Park Memorial Institute (RPMI) medium (Figure 4A). In contrast, the average
274 growth rate for YST7 was higher than YST6 for vaginal and rectal isolates in vaginal simulative
275 medium (VSM, Figure 4B). There was no difference in growth rate between vaginal and rectal
276 isolates from either YST6 or YST7 when grown in either medium (Welch Two Sample t-test,
277 RPMI medium, YST6: $t_{45} = -0.60$, $P = 0.55$, YST7: $t_{34} = -1.73$, $P = 0.09$; VSM, YST6: $t_{22} = -0.51$,
278 $P = 0.62$, YST7: $t_{22} = 0.24$, $P = 0.81$). A single statistical vaginal outlier with a reduced growth
279 rate relative to other isolates in the YST6 population was identified, consistent with a visual
280 inspection of the growth rates (Figure 4, Rosner's test for outliers). No other outliers were
281 identified in the other populations grown in RPMI or VSM.

282 We then compared drug resistance and drug tolerance from vaginal and rectal isolates
283 from YST6 and YST7 for two common drugs prescribed in our local clinic, fluconazole (FLC)
284 and boric acid (BA). Drug response was assessed through disk diffusion assays at both pH 7
285 and pH 4.2. pH significantly influenced the response phenotype for both resistance and
286 tolerance for isolates from both participants examined in both drugs (Figure 5). This was most
287 apparent in YST6 isolates in FLC, where all isolates were susceptible at pH 7 but resistant at pH
288 4.2. Albeit to a lesser effect, the same pattern of increased resistance (reduced susceptibility)
289 and increased tolerance at lower pH was also seen for all other populations in both drugs

290 except FLC tolerance in YST7. Neither the site of isolation nor the interaction between pH and
291 site of isolation were significant in any environment for either phenotype except YST6 in BA,
292 where there was a significant interaction (Figure 5, statistical results in Table 1). A visual
293 examination of the data also shows that the majority of isolates were very similar to each other.
294 Formal outlier statistical tests were largely consistent with the qualitative visual assessment,
295 identifying only five outlier isolates for resistance and three for tolerance (resistance: a YST6
296 vaginal isolate and two YST7 vaginal isolates with reduced susceptibility to FLC measured at
297 pH 7, one YST6 vaginal isolate with reduced susceptibility to BA at pH 4, and one YST6 rectal
298 isolate with increased susceptibility to BA at pH 4; tolerance: one YST6 vaginal isolate in FLC
299 with reduced tolerance at pH 7, one YST6 rectal isolate in BA at pH 4 with increased tolerance,
300 and one YST7 rectal isolate in FLC at pH 7 with increased tolerance).
301 There was considerable variation among participants and isolates for invasive growth
302 (Figure 6). However, there was no difference between YST6 vaginal and rectal isolates
303 (Wilcoxon rank sum test with continuity correction; W = 252, $P = 0.279$), while YST7 vaginal
304 isolates had higher invasive growth than rectal isolates (W = 115.5, $P = 0.0002$). To determine
305 whether this was a consistent pattern for *C. albicans* populations, we then quantified invasive
306 growth in isolates from the other two participants. We found no difference between sites in
307 TVY10 (W = 372, p-value = 0.74), while rectal isolates exhibited higher invasive growth on
308 average than vaginal isolates in TVY4 (W = 372, p-value = 0.032). The overall picture is thus
309 that invasive growth seems to vary more between participants than between site of isolation,
310 and that statistical differences between sites are likely due to neutral processes rather than
311 selection for invasive growth in the vaginal environment.

312 **Discussion**

313 Vulvovaginal candidiasis is a common vaginal infection. Although there is a large body of
314 literature on VVC and RVVC (73), many open research questions that have potential diagnostic,
315 prognostic and treatment implications remain. Our work here extends past research that
316 examines the diversity and relatedness of vaginal and rectal isolates. We recovered rectal yeast
317 isolates from six of ten participants with a history of RVVC who were vaginal yeast positive
318 during a clinic visit. This is a proportion similar to previous culture-based studies (rectal
319 colonization in tandem with vaginal colonization between 45-86%, (40–43, 74). We whole-
320 genome sequenced four randomly chosen vaginal and four rectal isolates from the four
321 participants with large vaginal and rectal yeast populations. In all cases, isolates from the same
322 participant formed monophyletic groups with phylogenetically overlapping vaginal and rectal
323 isolates.

324 Our results are consistent with previous studies that generally found high genetic
325 relatedness between rectal and vaginal isolates (57, 75, 76) However, whole genome
326 sequencing vastly improves upon previously employed methods (such as microsatellites, PCR
327 fingerprinting, or MLST analyses) in resolution, enabling us to determine not only that different
328 isolates are closely related, but there is an extremely low level of standing genetic variation
329 within both the vaginal and rectal populations. Striking, we only saw loss of heterozygosity
330 (LOH) variation among isolates from one participant (TVY10), with three isolates sharing one
331 LOH block that was not present in the other five isolates. All other LOH blocks were shared
332 among participant isolates. LOH has a relatively high mutation rate relative to other types of
333 mutations (77). Accordingly, variation in LOH patterns have previously been found to be
334 extremely prevalent among related isolates in different species and experimental conditions in
335 *C. albicans* (e.g., serial isolates from HIV patients with oropharyngeal candidiasis, 78, among
336 three pairs of *C. albicans* commensal strains isolated from the digestive tract of healthy
337 individuals, 79, and among experimentally evolved populations, 80, 81). Given that recent work
338 in *Caenorhabditis elegans* found that populations evolving in the presence of host defence
339 mechanisms may elevate the LOH rate (82), it is even more surprising to see the overall
340 uniformity in LOH patterns. The WGS genotypic results are thus consistent with frequent
341 migration between vaginal and rectal populations, at least during active symptomatic infection,
342 and a rapid population expansion yielding a low effective population size and little time for
343 variation to arise. They are inconsistent with the rectal environment acting as a source
344 population for vaginal reinfection, as genetic variation would be expected to be higher.

345 Our work provides a baseline genetic diversity that will be useful in the future to better
346 understand what happens to the vaginal yeast population between symptomatic episodes and
347 to generally compare genetic diversity among different body sites and study participant groups.
348 We chose isolates for sequencing blind to the phenotypic data to provide an unbiased estimate
349 of diversity. Although it is tempting in experimental work when resources are limited to always
350 pick the most diverse isolates for characterization, it makes it more difficult to compare results
351 among studies. The nucleotide diversity in all three *C. albicans* populations was 0.003, two
352 orders of magnitude lower than the species-wide diversity of 0.36 previously reported (based on
353 isolates from the major clades 1, 2, 3, 4 and 11 (25). The nucleotide diversity within isolates
354 from the participant with an *N. glabrata* infection was similarly two orders of magnitude lower
355 than the previously published species-wide diversity (0.00007 compared to 0.00665 based on
356 151 isolates (64). Our future work will extend this study to examine diversity between and
357 across successive symptomatic infections, and to compare the diversity here to within-
358 population yeast diversity in other healthy and disease contexts.
359 WGS analyses of the THRIVE-yeast isolates compared to a global collection of isolates
360 also permitted us to identify a subgroup in clade 1 where vaginal isolates, including the isolates
361 from our participants with *C. albicans* infections, are over-represented. Compared to other
362 clades, a greater proportion of clade 1 isolates have previously been shown to be significantly
363 associated with superficial infections (83), including in the context of VVC (18, 60, 61, 84).
364 Although vaginal isolates can be found throughout the phylogenetic tree, there may be
365 something unique to the common ancestor of the subgroup that makes them more amenable to
366 colonizing and invading epithelial surfaces in general and hence able to cause vaginal diseases
367 (85, 86). The majority of phenotypes of clinical interest that have previously been assessed are
368 found to vary considerably among isolates within the same *C. albicans* clade, precluding clear
369 phenotype × clade associations (e.g., 87, 88). However, it may be that more fine-scale
370 phylogenetic resolution is required to tease apart statistical relationships. If only a subgroup of
371 clade 1 isolates are enriched for a particular phenotype, that would not be seen if all clade 1
372 isolates are grouped together. Intriguingly, Sala *et al.* recently found that VVC isolates induced
373 greater fungal shedding from epithelial cells and differently stimulated epithelial signalling
374 pathways, compared to isolates from healthy women (89). This is the clearest *in vitro* assay able
375 to phenotypically differentiate VVC and healthy isolates, and hence a strong target for a GWAS
376 analysis able to potentially pinpoint the genetic basis of this seemingly important trait. It would
377 be of great interest to determine whether there is a genotypic association between common

378 variants in the subgroup of clade 1 isolates (including isolates from other body sites) and their
379 interaction with vaginal epithelial cells.
380 Significantly less work has been done to examine *N. glabrata* in the context of R/VVC
381 despite the increasing incidence of *N. glabrata* globally as an etiological agent of R/VVC (19,
382 90). Despite the fact that *N. glabrata* exhibits different pathogenesis and immunological
383 responses compared to *C. albicans (90–92),* it has received less attention. This is evident in the
384 very low number of vaginal whole-genome sequences available in the public archive. We
385 exhaustively searched NCBI for available *N. glabrata* sequences. The 526 isolates we found
386 were overwhelmingly blood isolates; we found only a single vaginal isolate. This precludes
387 conducting any statistical examination of the phylogenetic tree to determine whether vaginal
388 isolates cluster more than expected, as we found in *C. albicans.*
389 Our phenotypic analyses were largely consistent with our genetic results. We primarily
390 identified differences species-level differences in growth rate and drug response phenotypes.
391 This was not unexpected, given the common and well-established variation among isolates from
392 different species. We also found differences in drug resistance and tolerance when we
393 conducted disk diffusion assays at pH 7 compared to pH 4.2. This is consistent with previous
394 findings comparing CLSI micro broth testing at pH 7 and more acidic pH values in both *C.*
395 *albicans* and *N. glabrata* (93–95). Consistent with the genotypic results, the overarching pattern
396 was generally low variation and non-statistically different phenotypic distributions between the
397 vaginal and rectal isolates. This is in concordance with a previous *in vitro* study that found that
398 rectal isolates were equally able as vaginal isolates in expressing virulence factors (96). The
399 exception was invasive growth, where the YST7 vaginal population exhibited higher invasive
400 growth than the rectal population, which could possibly be indicative of selection in the vaginal
401 environment. We thus phenotyped isolates from the two additional participants with *C. albicans*
402 infections, finding no consistency in the results: in TVY4 the rectal population had statistically
403 higher invasive growth, while TVY10 the two populations were not statistically different.
404 Overwhelmingly, phenotypic variation was thus between participants (and *in vitro* environments)
405 rather than within isolates from the same individual.
406 Here we have conducted the largest study to date that employed whole genome
407 sequencing and high throughput phenotypic analyses to compare vaginal and rectal isolates
408 from symptomatic RVVC infections. We find no evidence that rectal isolates are more diverse
409 than vaginal isolates which is inconsistent with the GI tract being a source population for vaginal
410 reinfection. The extremely low diversity among isolates, including to the level of loss-of-
411 heterozygosity banding patterns, is indicative of a low level of standing genetic variation across

412 sites, consistent with frequent population bottlenecks and ongoing migration during symptomatic
413 infection. Combined, this may help explain why antifungal resistance rarely emerges in RVVC
414 infections, even in the face of consistent drug exposure, as a low effective population size
415 reduces the efficacy of selection for resistance.

416 **Methods**

417

418 **Clinical isolates**

419 Clinical samples were collected from four non-pregnant female participants between the ages of
420 18-50 with a history of recurrent vulvovaginal candidiasis (RVVC). Seventeen consenting female
421 participants who attended a clinic for individuals with a history of RVVC in Winnipeg, Canada
422 were sampled at the clinic for possible inclusion in this study. Vaginal and rectal swabs were
423 acquired from all participants. Swabs were kept at -4 ℃ and on the ice during transport, and
424 processed within 5 h of acquisition. Swabs were then agitated for ~30 s in 1 mL PBS and a
425 dilution series (1, 1:10, 1:100) was conducted. 100 µL from each of the three different dilutions
426 was spread onto SDA and chromogenic *Candida* agar plates. Plates were incubated for 48 h at
427 30 °C, at which time up to 24 vaginal and 24 rectal isolates were haphazardly isolated from
428 dilution plates that had clear margins between colonies. Colonies were suspended in 1 mL of
429 20% glycerol and kept at -70 °C. We collectively refer to all colonies collected as part of this
430 study as "THRIVE-yeast" isolates. Of the seventeen participants, vaginal colonies were
431 acquired from ten. Of those, six were also positive for rectal isolates. We recovered less than 10
432 rectal colonies from the two participants who were rectal swab positive but not included here.
433 This study has been approved by the University of Manitoba Biomedical Research Ethics Board
434 (HS24769 (B2021:026)) and Shared Health (SH2021:038).

435 **DNA extraction and sequencing**

436

437 Genomic DNA was extracted from eight single colonies (four each from the vagina and rectum)
438 from one individual with an *N. glabrata* infection (YST6) and three individuals with *C. albicans*
439 infections (YST7, TVY4, TVY10). We followed a standard phenol-chloroform protocol as
440 previously described (97). DNA quality and concentration were assessed on a Thermo
441 Scientific™ NanoDrop 2000 and Qubit® 2.0 Fluorometer respectively. Genomic DNA was sent
442 to either Microbial Genome Sequencing Center ("MIGS", Pittsburgh, USA; YST6 and YST7) or
443 SeqCoast Genomics (New Hampshire Ave., USA; TVY4 and TVY10) for sequencing. At MIGS,
444 sample libraries were prepared using the Illumina DNA Prep kit and IDT 10bp UDI indices, and
445 sequenced on NextSeq 2000 using a 300-cycle flow cell kit, producing 2×151bp reads. The bcl-
446 convert v3.9.3 software was used to assess read quality, demultiplex and trim adapter
447 sequences. At SeqCoast Genomics, samples were prepared using an Illumina DNA Prep
448 tagmentation kit and unique dual indexes. Sequencing was performed on the Illumina

449 NextSeq2000 platform using a 300-cycle flow cell kit to produce 2×150bp paired reads.
450 DRAGEN v3.10.11 was used to assess read quality, demultiplex and trim adapter sequences.
451 The reads have been deposited at the National Center for Biotechnology Information (NCBI)
452 Sequence Read Archive under BioProject ID PRJNA991137.
453
454 In addition to the 32 genomes we sequenced, we downloaded an additional 182 *C. albicans*
455 FASTQ files from the NCBI Sequence Read Archive database (98) from BioProject Accession
456 PRJNA432884 (25) and 526 *N. glabrata* FASTQ files from nineteen BioProjects (see Table S1).
457
458 **Variant calling**
459 The sequence reads were trimmed with Trimmomatic (v0.39) (99) with standard parameters
460 (LEADING: 10, TRAILING: 3, SLIDINGWINDOW:4:15, MINLEN: 31, TOPHRED33, following
461 (100)). Quality was assessed with FASTQC
462 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and MultiQC (101). *C. albicans*
463 trimmed paired-end reads were mapped using bwa-mem (102) to the SC5314 haplotype A
464 reference genome (A22-s07-m01-r160) downloaded from the Candida Genome Database
465 (103). The resulting SAM file was coordinate-sorted and converted to a Binary Alignment Map
466 (BAM) file using samtools v1.9 (104). *N. glabrata* isolates were mapped to the CBS138
467 reference genome (GCA000002545v2) downloaded from the Ensembl Genome Database
468 (105). Alignment quality was assessed with CollectAlignmentSummaryMetrics from picard
469 v2.26.3 (http://broadinstitute.github.io/picard) and consolidated across all samples with MultiQC
470 (101). All files had a >95% mapping quality. BAM files were further processed with picard by
471 adding a read group annotation so that samples with the same Bioproject ID had the same read
472 group, removing duplicate PCR amplicons, and fixing mate pairs. Base quality scores for the *C.*
473 *albicans* aligned reads were recalibrated with known single-nucleotide polymorphisms obtained
474 from the Candida Genome Database website
475 (http://www.candidagenome.org/download/gff/C_albicans_SC5314/Assembly21/A21_Jones_PM
476 ID_15123810_Polymorphisms.vcf; downloaded on July 29 2020) (106) using the
477 BaseRecalibrator and ApplyBQSR from the Genome Analysis Toolkit 4.2.4.0. The average
478 coverage for each isolate was estimated using samtools v1.9 (104).
479 The GATK Best Practices were adapted for variant calling. In sequence,
480 HaplotypeCaller, CombineGVCFs, GenotypeVCFs, VariantFiltration, and SelectVariants (107–
481 109) was used to identify single nucleotide variants (SNPs) among all sequenced isolates in
482 diploid and haploid mode for *C. albicans* and *N. glabrata* respectively. The results SNP table

483 was hard filtered using the suggested parameters and to match (25) (QualByDepth <2.0,
484 FisherStrand >60.0, root mean square mapping quality <30.0, MappingQualityRankSumTest <
485 −12.5, ReadPosRankSumTest < −8.0).
486
487 **Phylogeny construction**
488 Phylogenetic trees were constructed for *C. albicans* and *N. glabrata* isolates. For each species,
489 the multi-sample VCF file consisting of only genomic DNA SNPs was converted to a FASTA
490 alignment using a publicly available python script that creates an alignment matrix for
491 phylogenetic analysis (vcf2phylip.py v2.8, downloaded from
492 https://github.com/edgardomortiz/vcf2phylip) (110). For heterozygous SNPs in *C. albicans* the
493 consensus sequence is preferentially made based on the reference (haplotype A) base.
494 Ambiguous bases are written following IUPAC nucleotide ambiguity codes in the matrix. Ploidy
495 is not specified, as the most parsimonious ploidy is detected by the script. The FASTA
496 alignment was parsed in FastTree (2.1.11) (63) in the double precision mode to construct an
497 approximate maximum-likelihood phylogenetic tree using the general time reversible model and
498 the -gamma option to rescale the branch lengths. Compared to the time-consuming ML-based
499 phylogeny predictors such as RAxML (111), FastTree has been found to produce large dataset
500 trees with similar accuracy within a significantly shorter amount of time (112). The phylogeny
501 was visualized and annotated with the Interactive Tree Of Life (iTOL, v5) (113). Isolates from *C.*
502 *albicans* Clade 13, which is believed to be a closely related but different species termed *C.*
503 *africanus* (66, 67, 114) was used to root the phylogeny. The *N. glabrata* phylogeny was rooted
504 at the midpoint. While constructing the phylogeny with FastTree with this large data set yielded
505 a phylogeny comparable to *C. albicans* whole-genome phylogenies previously generated by
506 other groups using RAxML (25, 115), an attempt to use FastTree to resolve the relatedness of
507 the THRIVE-yeast isolates yielded poor resolution, as most of the SNP sites were ignored
508 during the tree construction. Hence to generate a more-refined phylogeny depicting the
509 relatedness among the isolates from each participant, we constructed a maximum likelihood
510 phylogeny with RAxML v8.2.12 (111) following standard practice using the GTR+G model with
511 20ML inferences on the alignment, inferring bootstrap replicate trees, applying MRE-based
512 bootstrapping test, and drawing support values using TBE (Transfer Bootstrap Expectation) on
513 the best-scoring tree.

514 We also looked at the variation in SNPs among isolates from the same participant. Using
515 RTG tools with the vcfsplit option (116), the VCF files for each isolate were extracted from the
516 multi-sample VCF file. Using the bedr package in R (117), the VCF files were converted to bed
517 files. These were used to construct upset plots using Intervene, a tool for the intersection and
518 visualization of multiple gene or genomic region sets (69).
519
520 ***In silico* MLST typing**
521 As clades in *N. glabrata* are commonly identified and named through MLST analysis ((64, 65)),
522 we conducted *in silico* MLST analyses of the YST6 isolates as well as the cluster in which the
523 YST6 isolated are associated with using stringMLST (118) which uses a predefined MLST *N.*
524 *glabrata* allele library (FKS, LEU2, NMT1, TRP1, UGP1, URA3) in the PubMLST database
525 (119). The diploid sequence types (DSTs) of the global *C. albicans* isolates were determined in
526 a previous study using a PCR-based method (25). *In silico* identification of DST from the WGS
527 was error-prone and unable to replicate known DSTs.
528
529 **Nucleotide diversity among THRIVE-yeast isolates**
530 Nucleotide diversity was estimated between the eight THRIVE-yeast isolates from the same
531 participant as π, the average pairwise difference between all possible isolate pairs. Briefly, as
532 above, variants were called using GenotypeGVCFs with --all-sites option activated. Vcftools
533 (v0.1.16) (120) was used to filter the variants (with --max-meanDP 500, --min-meanDP 20, --
534 max-missing 0.8). Indels and mitochondrial DNA were excluded (--remove-indels, --not-chr). 14
535 027 543 sites remained among the *C. albicans* isolates (TVY4, YST7, TVY10) and
536 12 048 365 among the *N. glabrata* (YST6) isolates. Pixy (v1.2.6.beta1) (121) was used to
537 estimate π using the four isolates from each site (vaginal or rectal) as a population. Compared
538 to other tools, pixy has been shown to better deal with bias that results from the implicit
539 assumption that missing genotypes, including those at sites not represented in the VCF, are
540 homozygous for the reference allele (121).
541
542 **Loss of heterozygosity and copy number analyses**
543 We conducted genomewide loss of heterozygosity (LOH) and copy number variant (CNV)
544 analyses using the web-based yeast analyses pipeline (Y_{MAP}) (70). Mitochondrial DNA is by
545 default excluded from the pipeline. Paired-end reads of the *N. glabrata* isolates were uploaded
546 and analyzed to the CBS138 reference genome (CGD: s05-m01-r09). As *N. glabrata* is haploid,
547 the ploidy was set as 1, which by default excludes LOH analyses. For the *C. albicans* isolates

548 paired-end read data for each isolate was uploaded and analyzed against the SC5314 A21-s02-
549 m09-r10 reference genome. Ploidy was left at the default value (two, i.e. diploid), and correction
550 was enabled for GC-content bias and chromosome-end bias. THRIVE-yeast isolates were
551 compared against the SC5314 haplotype map to distinguish between ancestral and newly
552 evolved LOH signatures and determine which allele was retained. CNV profiles were compared
553 to the reference SC5134 isolate as well as closely related isolates in the phylogeny. The
554 genomic elements within observed CNV regions were identified using the "gene/sequence
555 resources" section of the Candida Genome Database (CGD, https://candidagenome.org).
556
557 **Growth Rate Assay**
558 Two separate growth rate assays were conducted to measure growth in RPMI (10.4 % w/v
559 RPMI powder, 1.5 % w/v dextrose, 1.73 % w/v MOPS, adjusted to pH 7 with NaOH tablets) and
560 vaginal simulative medium ("VSM", following (25): 1.16 % v/v 5 mM NaCl, 3.6 % v/v 0.5 M KOH,
561 0.0128 % v/v 99% glycerol, 20 % v/v 0.01 M $Ca(OH)_2$, 1.34 % v/v 0.5 M Urea, 6.6 % v/v 0.5 M
562 glucose, 0.67% w/v solid YNB, 0.85 % v/v 2 M acetic acid, 0.192 % v/v lactic acid, adjusted to
563 pH 4.2 with NaOH tablets). For each experiment, 5 µL of frozen glycerol stock from YST6 (*N.*
564 *glabrata*) and YST7 (*C. albicans*) vaginal and rectal isolates was inoculated in duplicate into 500
565 µL of RPMI or VSM, and incubated for 48 h at 37 °C with agitation at 250 rpm. Cultures were
566 then standardized to an optical density (OD) of 0.01 A600 in RPMI or VSM and 200 µL was
567 transferred into a 96-well round bottom plate and sealed with a Breathe-Easier sealing
568 membrane (Electron Microscopy Sciences, PA, United States). OD_{600} readings were taken by
569 the Epoch plate reader (Biotek) every 15 minutes, with continuous shaking at 37 °C for 48 h.
570 From each well, the maximal growth rate was calculated as the spline with the highest slope
571 using a custom R script written by Dr. Richard Fitzjohn (https://github.com/acgerstein/THRIVE-
572 variation/scripts_real/growthrate_generic.R). The average growth rate between two technical
573 replicates for each isolate in each growth medium was used for visualization and statistical
574 analysis. Statistical outliers were determined through Rosner's test of outliers available through
575 the rosterTest function in the EnvStats R package (122).
576
577 **Drug resistance and tolerance**
578 Disk diffusion assays were carried out to measure resistance and tolerance to fluconazole and
579 boric acid at pH 7 and pH 4.2. Standard CLSI M44 guidelines for antifungal disk diffusion
580 susceptibility testing (123) were generally followed except for pH adjustment as appropriate. To
581 initiate each experiment, frozen stock of isolates from YST6 (*N. glabrata*) and YST7 (*C.*

582 *albicans*) were streaked onto Sabouraud dextrose agar (SDA) plates and incubated for 72 h at
583 room temperature. Isolates were then subcultured by streaking one colony onto a fresh SDA
584 plate and incubating at 37 ºC for 24 h. One colony from each isolate was then suspended in 200
585 µl of 0.85% saline solution and standardized to an OD_{600} of 0.01 in 1 mL of saline solution.
586 Within 15 minutes of standardization, 100 µl of the culture was spread evenly using autoclaved
587 glass beads on Mueller-Hinton (MH) plates that had been adjusted to a pH of 7.0 or a pH of 4.2.
588 Plates were left to dry for 20 minutes before placing a single 5 mg boric acid or 25 mg
589 fluconazole disk in the center of the plate. Plates were incubated upside down at 37 ºC. After 48
590 h, photographs of each plate were taken on a lightbox. Previous work demonstrated consistent
591 drug resistance values at 24 h and 48 h. The entire experiment was conducted twice for each
592 isolate × drug × pH combination, with two technical replicates for each of the two biological
593 replicates.
594 The 48 h images were processed in ImageJ as previously described (33). They were
595 cropped to a uniform size, the colours were inverted, and brightness and contrast were adjusted
596 to maximize the contrast between the white disk and the black background. The adjusted
597 images were run through the diskImageR package (72) for drug resistance (RAD_{20}) and
598 tolerance (FOG_{20}) quantification. Briefly, diskImageR calculates resistance as RAD_{20}, as the
599 radius of the zone of inhibition where growth is reduced by 20% relative to growth on the
600 margins of the plate where there is no drug, and tolerance as FoG_{20}, the fraction of realized
601 growth between RAD_{20} and the disk.
602 For each participant × drug combination, linear mixed-effect models were fit to determine
603 whether body site (vaginal or rectal) and environment (pH 4.2 or pH 7) or their interaction
604 significantly influenced resistance or tolerance. This was done using the lmer function in the
605 lme4 package in the R programming language with replicate as a random effect (124, 125).
606 Statistical significance was assessed from the type III analysis of variance table with
607 Satterthwaite's method.
608
609 **Invasive growth assay**
610 To examine invasive growth, we revised the methods from (126). Freezer stock from 24 isolates
611 from all four participants (YST6, TVY4, TVY10, and YST7) were streaked onto 20 mL yeast
612 peptone dextrose (YPD) plates (2% w/v peptone, 2% w/v yeast extract, 1.8% w/v agar, 1% w/v
613 glucose, 0.00016% w/v adenine sulfate, 0.00008% w/v uridine, 0.1% v/v of chloramphenicol and
614 ampicillin) and grown for 72 h at room temperature. A single colony was then randomly chosen
615 from each isolate and inoculated into 200 µL YPD. If no single colonies were available, a similar

amount of culture from the colony lawn was used. Cultures were standardized to OD_{600} 0.01 in 1 mL of liquid YPD media, then 2 µL of standardized culture was spotted onto the surface of a 20 mL solid YPD plate in a hexagonal pattern for a total of 7 spots per plate (i.e., spotted at each vertex and in the center). Plates were incubated for 96 h at 37 ºC. The surface growth was washed off using distilled water and a photograph was taken in a dark room on a lightbox. Two biological replicates were performed for each isolate.

The qualitative amount of invasive growth for each isolate was determined by visual examination of the post-wash photographs. To develop a five-point scale, two different people independently went through the post-wash pictures from YST6 and YST7 and selected two to six representative pictures that fit into five levels of invasive growth (scored as 1-5). The independent selections were then compared, and one image from each person was chosen as the most representative for each level of the scale. Using these as a reference, each isolate was then categorized into the five levels of the scale (0 - no growth/pipette tip indent, 0.25 - pinprick growth, 0.5 - circular growth evident, 0.75 - circular growth with pinprick, 1 - dense growth throughout). The maximum score between the two bio-replicates of each isolate was used for statistical analysis, though the same statistical conclusions were obtained if the mean score was used instead. For each participant, a Wilcoxon rank sum test was used to compare vaginal and rectal isolates.

Data availability

FASTQ files generated for this project has been deposited at the National Center for Biotechnology Information (NCBI) Sequence Read Archive under BioProject ID PRJNA991137. All phenotypic data and code required to reproduce figures and statistical analyses are available at https://github.com/acgerstein/THRIVE_yeast-VR. Large files (e.g., BAM, VCF) are available upon request.

Acknowledgements & Funding

We are extremely thankful to the women who have participated in THRIVE-yeast. We are grateful to Dr. Alicia Berard, Dr. Adam Burgener, Kenzie Birse and other THRIVE investigators for sharing their knowledge and resources which were integral in the initiation of this study and thank other members of the THRIVE-yeast research team including study clinicians, laboratory and administrative staff for their contributions. The study was funded through the Manitoba Medical Service Foundation, a University of Manitoba Faculty of Science Collaborative grant and the Canadian Institute for Advanced Research (CIFAR). ACG acknowledges the support of the CIFAR Azrieli Global Scholars Program and start-up funding from the University of Manitoba. A-RAB and AS were supported by EvoFunPath (NSERC CREATE) fellowships. DAH was funded by a University of Manitoba Research Award and the University of Manitoba Department of Microbiology Fellowship for Education Purposes. AdG was supported by a University of Manitoba Science Students' Association and Faculty of Science Undergraduate Student Research Award (USRA). BM was supported by a Faculty of Science USRA. Some swabs and amies buffer tubes were donated by COPAN Italia.

657 **References:**

658 1. Benedict K, Singleton AL, Jackson BR, Molinari NAM. 2022. Survey of incidence, lifetime
659 prevalence, and treatment of self-reported vulvovaginal candidiasis, United States, 2020.
660 BMC Womens Health 22:147.

661 2. Yano J, Sobel JD, Nyirjesy P, Sobel R, Williams VL, Yu Q, Noverr MC, Fidel PL Jr. 2019.
662 Current patient perspectives of vulvovaginal candidiasis: incidence, symptoms,
663 management and post-treatment outcomes. BMC Womens Health 19:48.

664 3. Rathod SD, Buffler PA. 2014. Highly-cited estimates of the cumulative incidence and
665 recurrence of vulvovaginal candidiasis are inadequately documented. BMC Womens Health
666 14:43.

667 4. Eschenbach DA. 2004. Chronic vulvovaginal candidiasis. N Engl J Med 351:851-852.

668 5. Sobel JD, Wiesenfeld HC, Martens M, Danna P, Hooton TM, Rompalo A, Sperling M,
669 Livengood C 3rd, Horowitz B, Von Thron J, Edwards L, Panzer H, Chu T-C. 2004.
670 Maintenance fluconazole therapy for recurrent vulvovaginal candidiasis. N Engl J Med
671 351:876–883.

672 6. Sojakova M, Liptajova D, Borovsky M, Subik J. 2004. Fluconazole and itraconazole
673 susceptibility of vaginal yeast isolates from Slovakia. Mycopathologia 157:163–169.

674 7. Foxman B, Muraglia R, Dietz J-P, Sobel JD, Wagner J. 2013. Prevalence of recurrent
675 vulvovaginal candidiasis in 5 European countries and the United States: results from an
676 internet panel survey. J Low Genit Tract Dis 17:340–345.

677 8. Marrazzo J. 2003. Vulvovaginal candidiasis. BMJ 326:993–994.

678 9. Denning DW, Kneale M, Sobel JD, Rautemaa-Richardson R. 2018. Global burden of
679 recurrent vulvovaginal candidiasis: a systematic review. Lancet Infect Dis 18:e339–e347.

680 10. Donders GGG, Bellen G, Mendling W. 2010. Management of recurrent vulvo-vaginal
681 candidosis as a chronic illness. Gynecol Obstet Invest 70:306–321.

682 11. Donders G, Sziller IO, Paavonen J, Hay P, de Seta F, Bohbot JM, Kotarski J, Vives JA,
683 Szabo B, Cepulienė R, Mendling W. 2022. Management of recurrent vulvovaginal
684 candidosis: Narrative review of the literature and European expert panel opinion. Front Cell
685 Infect Microbiol 12:934353.

686 12. Rosati D, Bruno M, Jaeger M, Ten Oever J, Netea MG. 2020. Recurrent vulvovaginal
687 candidiasis: An immunological perspective. Microorganisms 8:144.

688 13. Sobel JD. 2003. Management of patients with recurrent vulvovaginal candidiasis. Drugs
689 63:1059–1066.

690 14. Zhang J-Y, Liu J-H, Liu F-D, Xia Y-H, Wang J, Liu X, Zhang Z-Q, Zhu N, Yan-Yan, Ying Y,
691 Huang X-T. 2014. Vulvovaginal candidiasis: species distribution, fluconazole resistance and
692 drug efflux pump gene overexpression. Mycoses 57:584–591.

693 15. Shi X-Y, Yang Y-P, Zhang Y, Li W, Wang J-D, Huang W-M, Fan Y-M. 2015. Molecular
694 identification and antifungal susceptibility of 186 *Candida* isolates from vulvovaginal

695 candidiasis in southern China. J Med Microbiol 64:390–393.

696 16. Guzel AB, Ilkit M, Akar T, Burgut R, Demir SC. 2011. Evaluation of risk factors in patients
697 with vulvovaginal candidiasis and the value of chromID Candida agar versus CHROMagar
698 Candida for recovery and presumptive identification of vaginal yeast species. Med Mycol
699 49:16–25.

700 17. Van Schalkwyk, J., Yudin, M.H., Allen, V., Bouchard, C., Boucher, M., Boucoiran, I., Caddy,
701 S., Castillo, E., Kennedy, V.L., Money, D.M. and Murphy, K., 2015. Vulvovaginitis:
702 screening for and management of trichomoniasis, vulvovaginal candidiasis, and bacterial
703 vaginosis. J Obstet Gynaecol Can 37:266–274

704 18. Song N, Kan S, Pang Q, Mei H, Zheng H, Li D, Cui F, Lv G, An R, Li P, Xiong Z, Fan S,
705 Zhang M, Chen Y, Qiao Q, Liang X, Cui M, Li D, Liao Q, Li X, Liu W. 2022. A prospective
706 study on vulvovaginal candidiasis: multicentre molecular epidemiology of pathogenic yeasts
707 in China. J Eur Acad Dermatol Venereol 36:566–572.

708 19. Makanjuola O, Bongomin F, Fayemiwo SA. 2018. An update on the roles of Non-albicans
709 *Candida* species in vulvovaginitis. J Fungi (Basel) 4:121.

710 20. Borman AM, Johnson EM. 2021. Name Changes for Fungi of Medical Importance, 2018 to
711 2019. J Clin Microbiol 59:10-1128.

712 21. Kidd SE, Abdolrasouli A, Hagen F. 2023. Fungal nomenclature: Managing change is the
713 name of the game. Open Forum Infect Dis 10:ofac559.

714 22. Drell T, Lillsaar T, Tummeleht L, Simm J, Aaspõllu A, Väin E, Saarma I, Salumets A,
715 Donders GGG, Metsis M. 2013. Characterization of the vaginal micro- and mycobiome in
716 asymptomatic reproductive-age Estonian women. PLoS One 8:e54379.

717 23. Beigi RH, Meyn LA, Moore DM, Krohn MA, Hillier SL. 2004. Vaginal yeast colonization in
718 nonpregnant women: A longitudinal study. Obstetrics & Gynecology 104:926-926.

719 24. Bullarbo M, Andersch B, Samuelson E, Lindgren A, Kondori N, Baltzer IM. 2017. Self-
720 diagnosis of vulvovaginal candidiasis is poor - A comparison of diagnostic methods
721 introducing β-glucan as a complement. Bioenergetics 6:1000202.

722 25. Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, Sertour N, Mosca K, Permal E,
723 Laval G, Bouchier C, Ma L, Schwartz K, Voelz K, May RC, Poulain J, Battail C, Wincker P,
724 Borman AM, Chowdhary A, Fan S, Kim SH, Le Pape P, Romeo O, Shin JH, Gabaldon T,
725 Sherlock G, Bougnoux M-E, d'Enfert C. 2018. Gene flow contributes to diversification of the
726 major fungal pathogen *Candida albicans*. Nat Commun 9:2253.

727 26. Anderson FM, Visser N, Amses K, Hodgins-Davis A, Weber AM, Metzner KM, McFadden
728 MJ, Mills RE, O'Meara MJ, James TY, O'Meara TR. 2023. Candida albicans selection for
729 human commensalism results in substantial within-host diversity without decreasing fitness
730 for invasive disease. PLoS Biol 21:e3001822.

731 27. Iavazzo C, Gkegkes ID, Zarkada IM, Falagas ME. 2011. Boric acid for recurrent
732 vulvovaginal candidiasis: the clinical evidence. J Womens Health 20:1245–1255.

733 28. Guaschino S, De Seta F, Sartore A, Ricci G, De Santo D, Piccoli M, Alberico S. 2001.

734 Efficacy of maintenance therapy with topical boric acid in comparison with oral itraconazole
735 in the treatment of recurrent vulvovaginal candidiasis. Am J Obstet Gynecol 184:598–602.

736 29. Pfaller MA, Rhomberg PR, Messer SA, Jones RN, Castanheira M. 2015. Isavuconazole,
737 micafungin, and 8 comparator antifungal agents' susceptibility profiles for common and
738 uncommon opportunistic fungi collected in 2013: temporal analysis of antifungal drug
739 resistance using CLSI species-specific clinical breakpoints and proposed epidemiological
740 cutoff values. Diagn Microbiol Infect Dis 82:303–313.

741 30. Fan SR, Liu XP, Li JW. 2008. Clinical characteristics of vulvovaginal candidiasis and
742 antifungal susceptibilities of *Candida* species isolates among patients in southern China
743 from 2003 to 2006. J Obstet Gynaecol Res 34:561–566.

744 31. Berkow EL, Lockhart SR. 2017. Fluconazole resistance in *Candida* species: a current
745 perspective. Infect Drug Resist 10:237–245.

746 32. Richter SS, Galask RP, Messer SA, Hollis RJ, Diekema DJ, Pfaller MA. 2005. Antifungal
747 susceptibilities of *Candida* species causing vulvovaginitis and epidemiology of recurrent
748 cases. J Clin Microbiol 43:2155–2162.

749 33. Salama OE, Gerstein AC. 2022. Differential response of *Candida* species morphologies
750 and isolates to fluconazole and boric acid. Antimicrob Agents Chemother 66:e0240621.

751 34. Rosenberg A, Ene IV, Bibi M, Zakin S, Segal ES, Ziv N, Dahan AM, Colombo AL, Bennett
752 RJ, Berman J. 2018. Antifungal tolerance is a subpopulation effect distinct from resistance
753 and is associated with persistent candidemia. Nat Commun 9:2470.

754 35. Berman J, Krysan DJ. 2020. Drug resistance and tolerance in fungi. Nature Reviews
755 Microbiology, 18:319-331.

756 36. Tasić S, Tasić N, Tasić A, Mitrović S. 2002. Recurrent genital candidosis of women;
757 consequence of reinfection of relapse. Med Biol 9:217-222.

758 37. Lanchares JL, Hernández ML. 2000. Recurrent vaginal candidiasis changes in
759 etiopathogenical patterns. Int J Gynaecol Obstet 71:29-35.

760 38. Giraldo PC, Rodrigues HM, Gosson AM, Amaral R, Eleutério J, Gonçalves AKS. 2014.
761 Vulvovaginitis and the treatment of asymptomatic partners: a systematic review and meta-
762 analysis. J Bras Doenças Sex Transm 26:15–20.

763 39. Milne JD, Warnock DW. 1979. Effect of simultaneous oral and vaginal treatment on the rate
764 of cure and relapse in vaginal candidosis. Br J Vener Dis 55:362–365.

765 40. El-Din SS, Reynolds MT, Ashbee HR, Barton RC, Evans EG. 2001. An investigation into
766 the pathogenesis of vulvo-vaginal candidosis. Sex Transm Infect 77:179–183.

767 41. Fong IW. 1994. The rectal carriage of yeast in patients with vaginal candidiasis. Clin Invest
768 Med 17:426–431.

769 42. Sobel JD. 1986. Recurrent Vulvovaginal Candidiasis. N Engl J Med 315:1455–1458.

770 43. Mårdh P-A, Novikova N, Stukalova E. 2003. Colonisation of extragenital sites by Candida in
771 women with recurrent vulvovaginal candiosis. BJOG 110:934–937.

772 44. Spinillo A, Nicola S, Colonna L, Marangoni E, Cavanna C, Michelone G. 1994. Frequency
773 and significance of drug resistance in vulvovaginal candidiasis. Gynecol Obstet Invest
774 38:130–133.

775 45. O'Connor MI, Sobel JD. 1986. Epidemiology of recurrent vulvovaginal candidiasis:
776 identification and strain differentiation of *Candida albicans*. J Infect Dis 154:358–363.

777 46. Warnock DW, Speller DC, Day JK, Farrell AJ. 1979. Resistogram method for differentiation
778 of strains of *Candida albicans*. J Appl Bacteriol 46:571–578.

779 47. Odds FC, Abbott AB. 1980. A simple system for the presumptive identification of *Candida*
780 *albicans* and differentiation of strains within the species. Sabouraudia 18:301–317.

781 48. Stein GE, Sheridan VL, Magee BB, Magee PT. 1991. Use of rDNA restriction fragment
782 length polymorphisms to differentiate strains of *Candida albicans* in women with
783 vulvovaginal candidiasis. Diagn Microbiol Infect Dis 14:459–464.

784 49. Lockhart SR, Reed BD, Pierson CL, Soll DR. 1996. Most frequent scenario for recurrent
785 *Candida vaginitis* is strain maintenance with "substrain shuffling": demonstration by
786 sequential DNA fingerprinting with probes Ca3, C1, and CARE2. J Clin Microbiol 34:767–
787 777.

788 50. Soll DR, Galask R, Isley S, Rao TV, Stone D, Hicks J, Schmid J, Mac K, Hanna C. 1989.
789 Switching of *Candida albicans* during successive episodes of recurrent vaginitis. J Clin
790 Microbiol 27:681–690.

791 51. Mendling W, Pinto De Andrade M, Gutschmidt J, Gantenberg R, Presber W, Schonian G.
792 2000. Strain specificity of yeasts isolated from different locations of women suffering from
793 vaginal candidosis, and their partners. Mycoses 43:387–392.

794 52. Vazquez JA, Sobel JD, Demitriou R, Vaishampayan J, Lynch M, Zervos MJ. 1994.
795 Karyotyping of *Candida albicans* isolates obtained longitudinally in women with recurrent
796 vulvovaginal candidiasis. J Infect Dis 170:1566–1569.

797 53. Amouri I, Sellami H, Abbes S, Hadrich I, Mahfoudh N, Makni H, Ayadi A. 2012.
798 Microsatellite analysis of *Candida* isolates from recurrent vulvovaginal candidiasis. J Med
799 Microbiol 61:1091–1096.

800 54. Sampaio P, Gusmão L, Correia A, Alves C, Rodrigues AG, Pina-Vaz C, Amorim A, Pais C.
801 2005. New microsatellite multiplex PCR for *Candida albicans* strain typing reveals
802 microevolutionary changes. J Clin Microbiol 43:3869–3876.

803 55. Güzel AB, Döğen A, Aydın M, Serin A, Serin MS, Kalkancı A, Ilkit M. 2013. Genotyping
804 reveals no link between *Candida albicans* genotype and vaginitis severity in Turkish
805 women. Mycopathologia 175:287–294.

806 56. Hong N, Lei Y, Chen H, Chen X, Tsui KM, Hu D, Liao W, Yan L, Zhang H, Zhao R, Wu G,
807 Yu N, Deng S. 2022. Genotyping and Drug Resistance Profile of Clinical Isolates of
808 *Candida albicans* from Vulvovaginal Candidiasis in the Eastern China. Mycopathologia
809 187:217–224.

810 57. Araújo Paulo de Medeiros M, Vieira de Melo AP, Gonçalves SS, Milan EP, Chaves GM.

811 2014. Genetic relatedness among vaginal and anal isolates of *Candida albicans* from
812 women with vulvovaginal candidiasis in north-east Brazil. J Med Microbiol 63:1436–1445.

813 58. Chong PP, Abdul Hadi SR, Lee YL, Phan CL, Tan BC, Ng KP, Seow HF. 2007. Genotyping
814 and drug resistance profile of *Candida* spp. in recurrent and one-off vaginitis, and high
815 association of non-albicans species with non-pregnant status. Infect Genet Evol 7:449–456.

816 59. Faria-Gonçalves P, Gaspar C, Oliveira AS, Palmeira-de-Oliveira R, Gonçalves T, Martinez-
817 de-Oliveira J, Palmeira-de-Oliveira A, Rolo J. 2021. Evaluation of overtime phenotypic
818 variation of yeasts in chronic vulvovaginal candidosis cases. Med Mycol 59:1166–1173.

819 60. Tian J-Y, Yang Y-G, Chen S, Teng Y, Li X-Z. 2021. Genetic diversity and molecular
820 epidemiology of *Candida albicans* from vulvovaginal candidiasis patients. Infect Genet Evol
821 92:104893.

822 61. Zhu Y, Fang C, Shi Y, Shan Y, Liu X, Liang Y, Huang L, Liu X, Liu C, Zhao Y, Fan S, Zhang
823 X. 2022. *Candida albicans* multilocus sequence typing clade I contributes to the clinical
824 phenotype of vulvovaginal candidiasis patients. Front Med 9:837536.

825 62. Bradford LL, Chibucos MC, Ma B, Bruno V, Ravel J. 2017. Vaginal *Candida* spp. genomes
826 from women with vulvovaginal candidiasis. Pathog Dis 75:ftx061.

827 63. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2--approximately maximum-likelihood trees
828 for large alignments. PLoS One 5:e9490.

829 64. Helmstetter N, Chybowska AD, Delaney C, Da Silva Dantas A, Gifford H, Wacker T, Munro
830 C, Warris A, Jones B, Cuomo CA, Wilson D, Ramage G, Farrer RA. 2022. Population
831 genetics and microevolution of clinical *Candida glabrata* reveals recombinant sequence
832 types and hyper-variation within mitochondrial genomes, virulence genes, and drug targets.
833 Genetics 221:iyac031.

834 65. Carreté L, Ksiezopolska E, Pegueroles C, Gómez-Molero E, Saus E, Iraola-Guzmán S,
835 Loska D, Bader O, Fairhead C, Gabaldón T. 2018. Patterns of genomic variation in the
836 opportunistic pathogen *Candida glabrata* suggest the existence of mating and a secondary
837 association with humans. Curr Biol 28:15–27.

838 66. Romeo O, Tietz H-J, Criseo G. 2013. *Candida africana*: Is It a fungal pathogen? Curr
839 Fungal Infect Rep 7:192–197.

840 67. Mixão V, Saus E, Boekhout T, Gabaldón T. 2021. Extreme diversification driven by parallel
841 events of massive loss of heterozygosity in the hybrid lineage of *Candida albicans*.
842 Genetics 217:iyaa004.

843 68. Riley R, Nagy L. 2018. Fungal phylogenomics. Methods Mol Biol 1775:251–266.

844 69. Khan A, Mathelier A. 2017. Intervene: a tool for intersection and visualization of multiple
845 gene or genomic region sets. BMC Bioinformatics 18:1-8.

846 70. Abbey DA, Funt J, Lurie-Weinberger MN, Thompson DA, Regev A, Myers CL, Berman J.
847 2014. YMAP: a pipeline for visualization of copy number variation and loss of
848 heterozygosity in eukaryotic pathogens. Genome Med 6:1-16.

849 71. Pemberton TJ. 2006. Identification and comparative analysis of sixteen fungal peptidyl-prolyl

850 cis/trans isomerase repertoires. BMC Genomics 7:1-30.

851 72. Gerstein AC, Rosenberg A, Hecht I, Berman J. 2016. diskImageR: quantification of
852 resistance and tolerance to antimicrobial drugs using disk diffusion assays. Microbiology
853 162:1059–1068.

854 73. Jafarzadeh L, Ranjbar M, Nazari T, Naeimi Eshkaleti M, Aghaei Gharehbolagh S, Sobel JD,
855 Mahmoudi S. 2022. Vulvovaginal candidiasis: An overview of mycological, clinical, and
856 immunological aspects. J Obstet Gynaecol Res 48:1546–1560.

857 74. Spinillo A, Carratta L, Pizzoli G, Lombardi G, Cavanna C, Michelone G, Guaschino S. 1992.
858 Recurrent vaginal candidiasis. Results of a cohort study of sexual transmission and
859 intestinal reservoir. J Reprod Med 37:343–347.

860 75. Sampaio P, Gusmão L, Alves C, Pina-Vaz C, Amorim A, Pais C. 2003. Highly polymorphic
861 microsatellite for identification of *Candida albicans* strains. J Clin Microbiol 41:552–557.

862 76. Shi W-M, Mei X-Y, Gao F, Huo K-K, Shen L-L, Qin H-H, Wu Z-W, Zheng J. 2007. Analysis
863 of genital *Candida albicans* infection by rapid microsatellite markers genotyping. Chin Med
864 J (Engl) 120:975–980.

865 77. Smukowski Heil C. 2023. Loss of heterozygosity and its importance in evolution. J Mol Evol
866 91:369–377.

867 78. Ford CB, Funt JM, Abbey D, Issi L, Guiducci C, Martinez DA, Delorey T, Li B yu, White TC,
868 Cuomo C, Rao RP, Berman J, Thompson DA, Regev A. 2015. The evolution of drug
869 resistance in clinical isolates of *Candida albicans*. 4:662.

870 79. Diogo D, Bouchier C, d'Enfert C, Bougnoux M-E. 2009. Loss of heterozygosity in
871 commensal isolates of the asexual diploid yeast *Candida albicans*. Fungal Genet Biol
872 46:159–168.

873 80. James TY, Michelotti LA, Glasco AD, Clemons RA, Powers RA, James ES, Simmons DR,
874 Bai F, Ge S. 2019. Adaptation by loss of heterozygosity in *Saccharomyces cerevisiae*
875 Clones Under Divergent Selection. Genetics 213:665–683.

876 81. Ene IV, Farrer RA, Hirakawa MP, Agwamba K, Cuomo CA, Bennett RJ. 2018. Global
877 analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen.
878 Proc Natl Acad Sci U S A 115:E8688–E8697.

879 82. Smith AC, Morran LT, Hickman MA. 2022. Host defense mechanisms induce genome
880 instability leading to rapid evolution in an opportunistic fungal pathogen. Infect Immun
881 90:e0032821.

882 83. Odds FC, Bougnoux M-E, Shaw DJ, Bain JM, Davidson AD, Diogo D, Jacobsen MD,
883 Lecomte M, Li S-Y, Tavanti A, Maiden MCJ, Gow NAR, d'Enfert C. 2007. Molecular
884 phylogenetics of *Candida albicans*. Eukaryot Cell 6:1041–1052.

885 84. Ge S-H, Xie J, Xu J, Li J, Li D-M, Zong L-L, Zheng Y-C, Bai F-Y. 2012. Prevalence of
886 specific and phylogenetically closely related genotypes in the population of *Candida*
887 *albicans* associated with genital candidiasis in China. Fungal Genet Biol 49:86–93.

888 85. Schmid J, Herd S, Hunter PR, Cannon RD, Yasin MSM, Samad S, Carr M, Parr D,

889 McKinney W, Schousboe M, Harris B, Ikram R, Harris M, Restrepo A, Hoyos G, Singh KP.
890 1999. Evidence for a general-purpose genotype in *Candida albicans*, highly prevalent in
891 multiple geographical regions, patient types and types of infection. Microbiology 145 (Pt
892 9):2405–2413.

893 86. Giblin L, Edelmann A, Zhang N, von Maltzahn NB, Cleland SB, Sullivan PA, Schmid J.
894 2001. A DNA polymorphism specific to *Candida albicans* strains exceptionally successful
895 as human pathogens. Gene 272:157–164.

896 87. Cravener MV, Do E, May G, Zarnowski R, Andes DR, McManus CJ, Mitchell AP. 2023.
897 Reinforcement amid genetic diversity in the *Candida albicans* biofilm regulatory network.
898 PLoS Pathog 19:e1011109.

899 88. MacCallum DM, Castillo L, Nather K, Munro CA, Brown AJP, Gow NAR, Odds FC. 2009.
900 Property differences among the four major *Candida albicans* strain clades. Eukaryot Cell
901 8:373–387.

902 89. Sala A, Ardizzoni A, Spaggiari L, Vaidya N, van der Schaaf J, Rizzato C, Cermelli C,
903 Mogavero S, Krüger T, Himmel M, Kniemeyer O, Brakhage AA, King BL, Lupetti A, Comar
904 M, de Seta F, Tavanti A, Blasi E, Wheeler RT, Pericolini E. 2023. A new phenotype in
905 candida-epithelial cell interaction distinguishes colonization- versus vulvovaginal
906 candidiasis-associated strains. MBio 14:e00107-23.

907 90. Kennedy MA, Sobel JD. 2010. Vulvovaginal candidiasis caused by Non-albicans *Candida*
908 species: New insights. Curr Infect Dis Rep 12:465–470.

909 91. Galocha M, Pais P, Cavalheiro M, Pereira D, Viana R, Teixeira MC. 2019. Divergent
910 approaches to virulence in *C. albicans* and *C. glabrata*: Two sides of the same coin. Int J
911 Mol Sci 20:2345.

912 92. Brunke S, Hube B. 2013. Two unlike cousins: *Candida albicans* and *C. glabrata* infection
913 strategies. Cell Microbiol 15:701–708.

914 93. Liu W, Zhang X, Liu Z, Luo X. 2011. Impact of pH on the antifungal susceptibility of vaginal
915 *Candida albicans*. Int J Gynaecol Obstet 114:278–280.

916 94. Sobel JD, Sebastian S, Boikov DA. 2023. A longitudinal study on Fluconazole resistance in
917 *C.albicans* vaginal isolates. Mycoses 66:563-565.

918 95. Danby CS, Boikov D, Rautemaa-Richardson R, Sobel JD. 2012. Effect of pH on *in vitro*
919 susceptibility of *Candida glabrata* and *Candida albicans* to 11 antifungal agents and
920 implications for clinical use. Antimicrob Agents Chemother 56:1403–1406.

921 96. Araújo Paulo de Medeiros M, Vieira de Melo AP, Maia de Sousa AM, Silva-Rocha WP,
922 Pipolo Milan E, Maranhão Chaves G. 2017. Characterization of virulence factors of vaginal
923 and anal isolates of *Candida albicans* sequentially obtained from patients with vulvovaginal
924 candidiasis in north-east Brazil. J Mycol Med 27:567–572.

925 97. Selmecki A, Forche A, Berman J. 2006. Aneuploidy and isochromosome formation in drug-
926 resistant *Candida albicans*. Science 313:367–370.

927 98. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K,

928 Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen
929 F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD,
930 Sherry ST. 2022. Database resources of the national center for biotechnology information.
931 Nucleic Acids Res 50:D20–D26.

932 99. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
933 sequence data. Bioinformatics 30:2114–2120.

934 100. Todd RT, Wikoff TD, Forche A, Selmecki A. 2019. Genome plasticity in *Candida albicans* is
935 driven by long repeat sequences. Elife 8:e45954.

936 101. Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for
937 multiple tools and samples in a single report. Bioinformatics 32:3047–3048.

938 102. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-
939 MEM. arXiv:1303.3997v2 [q-bio.GN].

940 103. Skrzypek MS, Binkley J, Binkley G, Miyasato SR, Simison M, Sherlock G. 2017. The
941 Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers
942 and visualization of high throughput sequencing data. Nucleic Acids Res 45:D592–D596.

943 104. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
944 R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map
945 format and SAMtools. Bioinformatics 25:2078–2079.

946 105. Yates AD, Allen J, Amode RM, Azov AG, Barba M, Becerra A, Bhai J, Campbell LI,
947 Carbajo Martinez M, Chakiachvili M, Chougule K, Christensen M, Contreras-Moreira B,
948 Cuzick A, Da Rin Fioretto L, Davis P, De Silva NH, Diamantakis S, Dyer S, Elser J, Filippi
949 CV, Gall A, Grigoriadis D, Guijarro-Clarke C, Gupta P, Hammond-Kosack KE, Howe KL,
950 Jaiswal P, Kaikala V, Kumar V, Kumari S, Langridge N, Le T, Luypaert M, Maslen GL,
951 Maurel T, Moore B, Muffato M, Mushtaq A, Naamati G, Naithani S, Olson A, Parker A,
952 Paulini M, Pedro H, Perry E, Preece J, Quinton-Tulloch M, Rodgers F, Rosello M, Ruffier
953 M, Seager J, Sitnik V, Szpak M, Tate J, Tello-Ruiz MK, Trevanion SJ, Urban M, Ware D,
954 Wei S, Williams G, Winterbottom A, Zarowiecki M, Finn RD, Flicek P. 2022. Ensembl
955 Genomes 2022: an expanding genome resource for non-vertebrates. Nucleic Acids Res
956 50:D996–D1003.

957 106. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G,
958 Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S. 2004. The diploid genome
959 sequence of *Candida albicans*. Proc Natl Acad Sci U S A 101:7329–7334.

960 107. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del
961 Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY,
962 Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery
963 and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498.

964 108. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A,
965 Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S,
966 DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis
967 Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–11.10.33.

968 109. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA,

969 Kling DE, Gauthier LD, Levy-Moonshine A, Roazen D, Shakir K, Thibault J, Chandran S,
970 Whelan C, Lek M, Gabriel S, Daly MJ, Neale B, MacArthur DG, Banks E. 2018. Scaling
971 accurate genetic variant discovery to tens of thousands of samples. bioRxiv
972 https://doi.org/10.1101/201178

973 110. Ortiz EM. 2019. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for
974 phylogenetic analysis. Zenodo 10.5281/zenodo.2540861

975 111. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
976 large phylogenies. Bioinformatics 30:1312–1313.

977 112. Liu K, Linder CR, Warnow T. 2011. RAxML and FastTree: comparing two methods for
978 large-scale maximum likelihood phylogeny estimation. PLoS One 6:e27731.

979 113. Letunic I, Bork P. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
980 tree display and annotation. Nucleic Acids Res 49:W293–W296.

981 114. Mixão V, Gabaldón T. 2020. Genomic evidence for a hybrid origin of the yeast
982 opportunistic pathogen Candida albicans. BMC Biol 18:48.

983 115. Szarvas J, Rebelo AR, Bortolaia V, Leekitcharoenphon P, Schrøder Hansen D, Nielsen
984 HL, Nørskov-Lauritsen N, Kemp M, Røder BL, Frimodt-Møller N, Søndergaard TS, Coia JE,
985 Østergaard C, Westh H, Aarestrup FM. 2021. Danish whole-genome-sequenced *Candida*
986 *albicans* and *Candida glabrata* samples fit into globally prevalent clades. J Fungi (Basel) 7:
987 962.

988 116. Cleary JG, Braithwaite R, Gaastra K, Hilbush BS, Inglis S, Irvine SA, Jackson A, Littin R,
989 Rathod M, Ware D, Zook JM, Trigg L, De La Vega FM. 2015. Comparing variant call files
990 for performance benchmarking of xext-generation sequencing variant calling pipelines.
991 bioRxiv doi: https://doi.org/10.1101/023754

992 117. Haider S, Waggott D, C. Boutros P. 2019. bedr: Genomic Region Processing using Tools
993 Such as “BEDTools”, “BEDOPS” and “Tabix.” https://CRAN.R-project.org/package=bedr.

994 118. Gupta A, Jordan IK, Rishishwar L. 2017. stringMLST: a fast k-mer based tool for multilocus
995 sequence typing. Bioinformatics 33:119–121.

996 119. Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics:
997 BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Res
998 3:124.

999 120. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE,
1000 Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis
1001 Group. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

1002 121. Korunes KL, Samuk K. 2021. pixy: Unbiased estimation of nucleotide diversity and
1003 divergence in the presence of missing data. Mol Ecol Resour 21:1359–1368.

1004 122. Millard SP. 2013. EnvStats: An R Package for Environmental Statistics. Springer New
1005 York.

1006 123. NCCLS. Reference method for antifungal disk diffusion susceptibility testing of yeasts;
1007 approved guideline. NCCLS document M44-A National Committee for Clinical Laboratory

1008 Standards Wayne.

1009 124. Kuznetsova A, Brockhoff B, Christensen HB. 2017. lmerTest: Tests in linear mixed effects
1010 models. J Stat Softw 82:1-26.

1011 125. Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using
1012 lme4. J Stat Softw 67:1–48.

1013 126. Zupan J, Raspor P. 2008. Quantitative agar-invasion assay. J Microbiol Methods 73:100–
1014 104.

1015

1016 **Table and figure legends**

1017 **Table 1.** Statistical results of disk diffusion assay linear mixed-effect models.
1018

1019 **FIG 1** Approximate maximum likelihood phylogenies of (A) *N. glabrata* (n = 526 isolates and
1020 eight THRIVE-yeast isolates) and (B) *C. albicans* (n = 182 isolates and 24 THRIVE-yeast
1021 isolates). THRIVE-yeast isolates are indicated with red labels. The YST6 *N. glabrata*
1022 isolates fall in a cluster of predominantly ST16 isolates (ST16 shown by blue branches,
1023 black indicates not ST16). TVY4, TVY10, and YST7 *C. albicans* isolates fall in clade 1 (blue
1024 branches). The physiological sites of isolation are indicated as a colour strip around the
1025 phylogeny indicating vaginal (purple), rectal (orange), and all other sites (grey). The *N.*
1026 *glabrata* phylogeny was rooted at the midpoint, and the *C. albicans* tree was rooted by
1027 clade 13 (grey labels).

1028

1029 **FIG 2** Fine-scale phylogenetic structure among THRIVE-yeast isolates. Vaginal (V) and rectal
1030 (R) isolates were acquired during symptomatic infection from four participants with a history
1031 of RVVC. A) *N. glabrata* isolates from YST6. *C. albicans* isolates from B) TVY4, C) TVY10,
1032 D) YST7. The left panels depict phylogenetic trees generated with RAxML; blue circles
1033 indicate branches with a minimum bootstrap support of 0.7. The right panels show the
1034 number of isolates each variants was found in.

1035

1036 **FIG 3** CNV and LOH profiles of THRIVE-yeast. A) Representative trace of a YST6 isolate (*N.*
1037 *glabrata*) and B) TVY10 isolates, representative traces of TVY4 and YST7 isolates, and
1038 four isolates that are closely related to YST7. Relative copy number compared to the
1039 reference genomes is shown as the horizontal black line with the scale on the y axis. For *C.*
1040 *albicans* isolates, the density of heterozygous SNPs is shown as the vertical lines spanning
1041 the height of each chromosome box, with the intensity representing the number of SNPs in
1042 5 kb bins. Heterozygous SNPs are gray, homozygous SNPs are coloured based on the
1043 retained SC5314 homolog: cyan for “AA” and magenta for “BB.” White indicates an
1044 ancestral LOH in SC5314. For each chromosome the centromere is indicated by an
1045 indentation in the box. The dots on the bottom line below each box indicate the positions of
1046 major repeat sequences. C) Fine-scale coverage mapping of the putative CNV on chr3.
1047 Shown is one representative trace from YST7 R2, all isolates have a similar pattern. Gene
1048 positions above the figure are approximated.

1049

1050 **FIG 4** Growth rate was measured from 24 vaginal and 24 rectal isolates from YST6 and YST7.
1051 Optical density was recorded every 15 minutes in a plate reader with constant shaking and
1052 incubation at 37 ℃. Each point represents the mean of two technical replicates for each
1053 isolate, 24 isolates were measured for each group. Growth rate was calculated as the
1054 spline with the highest slope using a custom R script.

1055

1056 **FIG 5** Drug response phenotypes from disk diffusion assays of isolates from YST6 (*N. glabrata*,

1057 left) and YST7 (*C. albicans*, right). Drug resistance (top panels) and drug tolerance (bottom
1058 panels) was measured from disk diffusion assays for fluconazole (FLC) and boric acid (BA).
1059 Drug response was measured on pH 4 and pH 7 Mueller-Hinton plates using the R
1060 package *diskImageR* (72), which computationally measures response parameters from
1061 images of disk diffusion assays. Each point represents the mean of four replicates (two
1062 technical replicates for two biological replicates), with the coloured line indicating the
1063 median of 24 isolates in each group.

1064

1065 **FIG 6** Invasive growth was qualitatively scored after growth on YPD plates for 96 h. Each point
1066 indicates the maximum scores between two bioreplicates for each isolate.

Table 1. Statistical results of disk diffusion assay linear mixed-effect models.

Participant	Drug	Phenotype	pH	Site of isolation	pH*site of isolation
YST6	FLC	Resistance	$F_{1,55} = 1207.3, p < 0.0001$	$F_{1,55} = 0.007, p = 0.93$	$F_{1,55} = 0.007, p = 0.93$
YST6	BA	Resistance	$F_{1,46} = 305.8, p < 0.0001$	$F_{1,46} = 0.81, p = 0.37$	$F_{1,46} = 5.47, p = 0.024$
YST7	FLC	Resistance	$F_{1,46} = 70.0, p < 0.0001$	$F_{1,46} = 3.3, p = 0.075$	$F_{1,46} = 0.08, p = 0.77$
YST7	BA	Resistance	$F_{1,46} = 296.0, p < 0.0001$	$F_{1,46} = 0.43, p = 0.51$	$F_{1,46} = 2.9, p = 0.10$
YST6	FLC	Tolerance	-	$t_{23} = 0.34, p = 0.74$	-
YST6	BA	Tolerance	$F_{1,92} = 7.9, p = 0.006$	$F_{1,92} = 0.16, p = 0.69$	$F_{1,92} = 0.77, p = 0.38$
YST7	FLC	Tolerance	$F_{1,92} = 8.52, p = 0.0044$	$F_{1,92} = 2.2, p = 0.14$	$F_{1,92} = 0.32, p = 0.57$
YST7	BA	Tolerance	$F_{1,46} = 323.1, p < 0.0001$	$F_{1,46} = 1.63, p = 0.21$	$F_{1,46} = 2.13, p = 0.15$

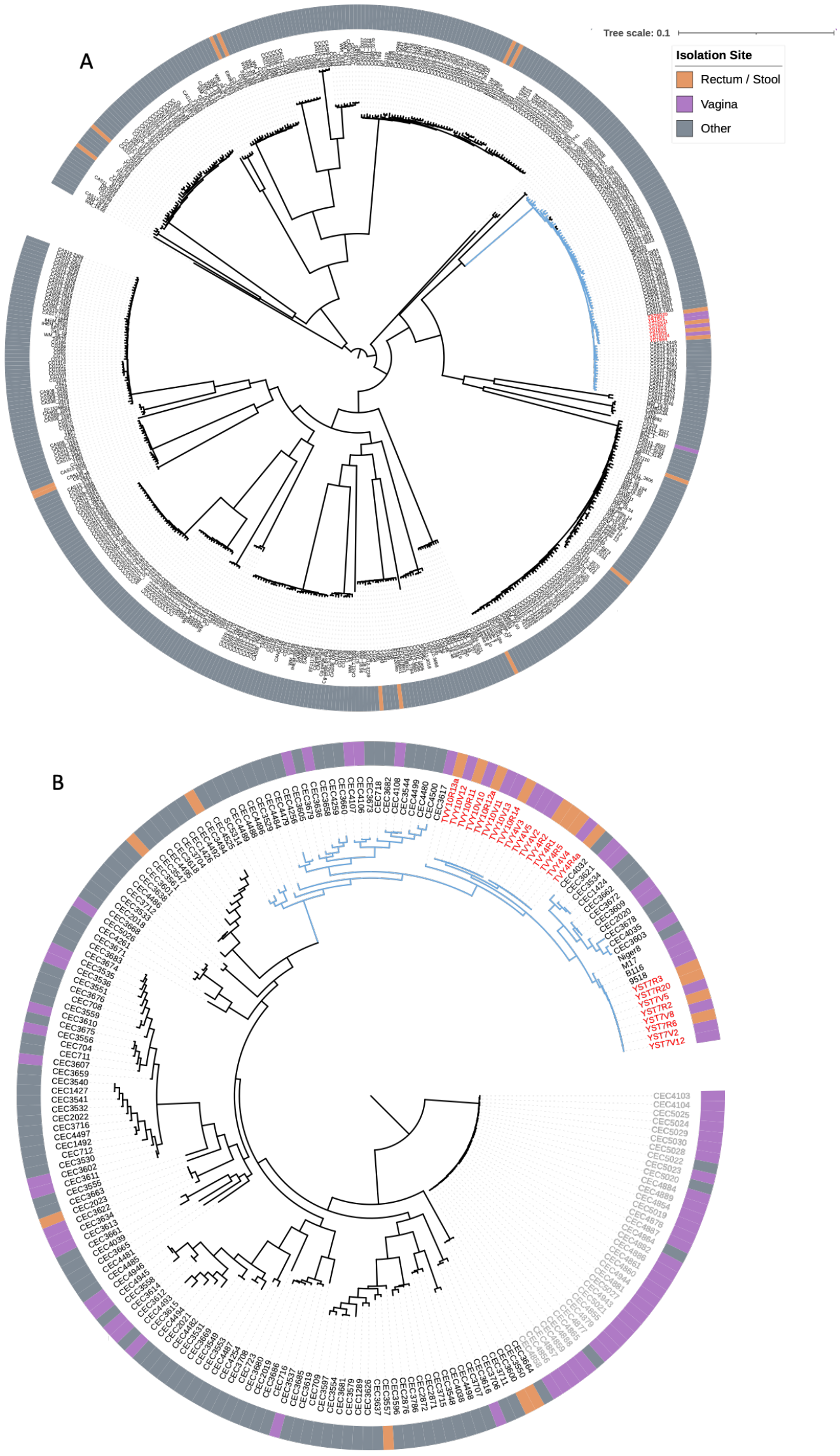

FIG 1 Approximate maximum likelihood phylogenies of (A) *N. glabrata* (n = 526 isolates and eight THRIVE-yeast isolates) and (B) *C. albicans* (n = 182 isolates and 24 THRIVE-yeast isolates). THRIVE-yeast isolates are indicated with red labels. The YST6 *N. glabrata* isolates fall in a cluster of predominantly ST16 isolates (ST16 shown by blue branches, black indicates not ST16). TVY4, TVY10, and YST7 *C. albicans* isolates fall in clade 1 (blue branches). The physiological sites of isolation are indicated as a colour strip around the phylogeny indicating vaginal (purple), rectal (orange), and all other sites (grey). The *N. glabrata* phylogeny was rooted at the midpoint, and the *C. albicans* tree was rooted by clade 13 (grey labels).

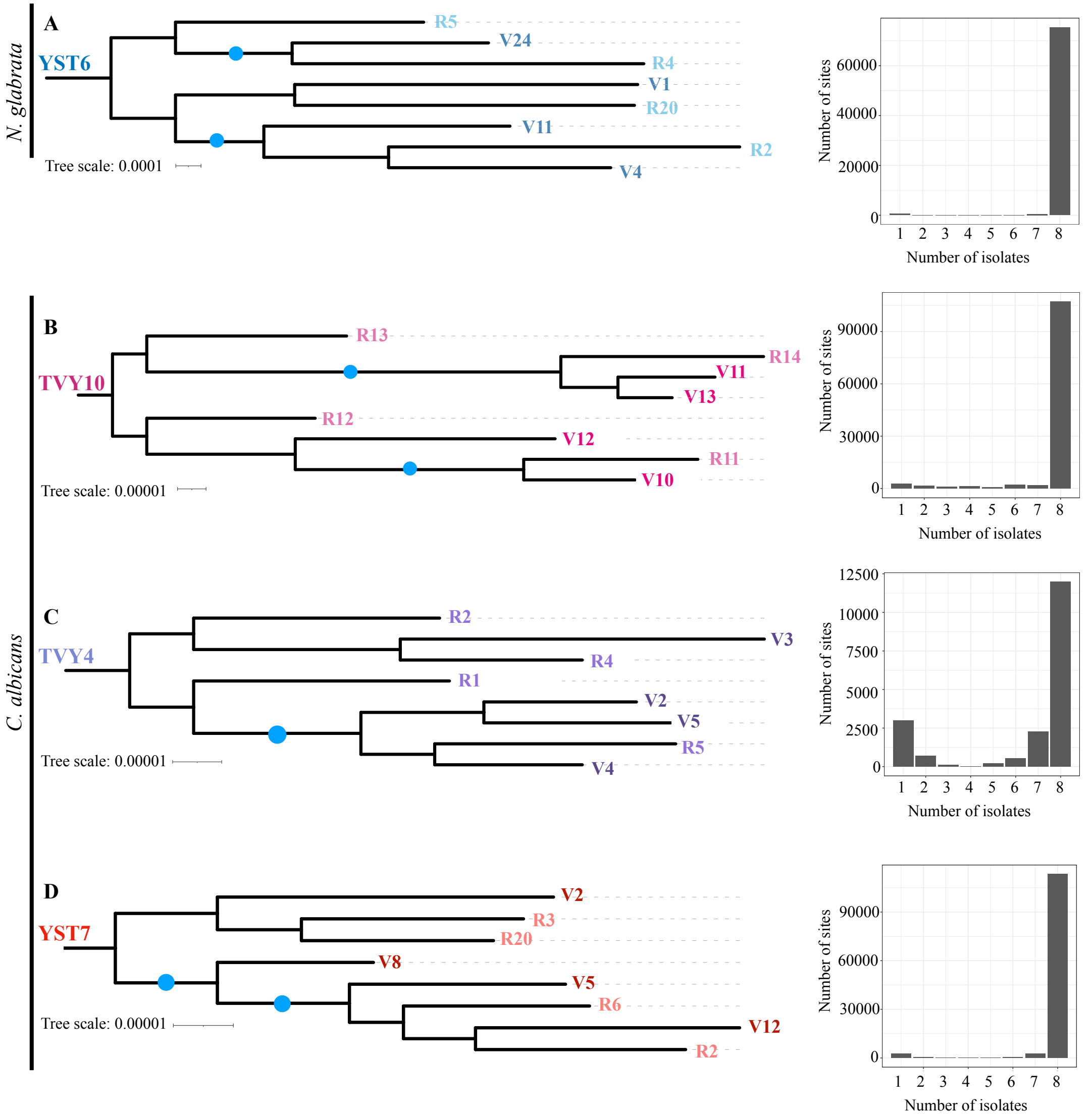

FIG 2 Fine-scale phylogenetic structure among THRIVE-yeast isolates. Vaginal (V) and rectal (R) isolates were acquired during symptomatic infection from four participants with a history of RVVC. A) *N. glabrata* isolates from YST6. *C. albicans* isolates from B) TVY4, C) TVY10, D) YST7. The left panels depict phylogenetic trees generated with RAxML; blue circles indicate branches with a minimum bootstrap support of 0.7. The right panels show the number of isolates each variants was found in.

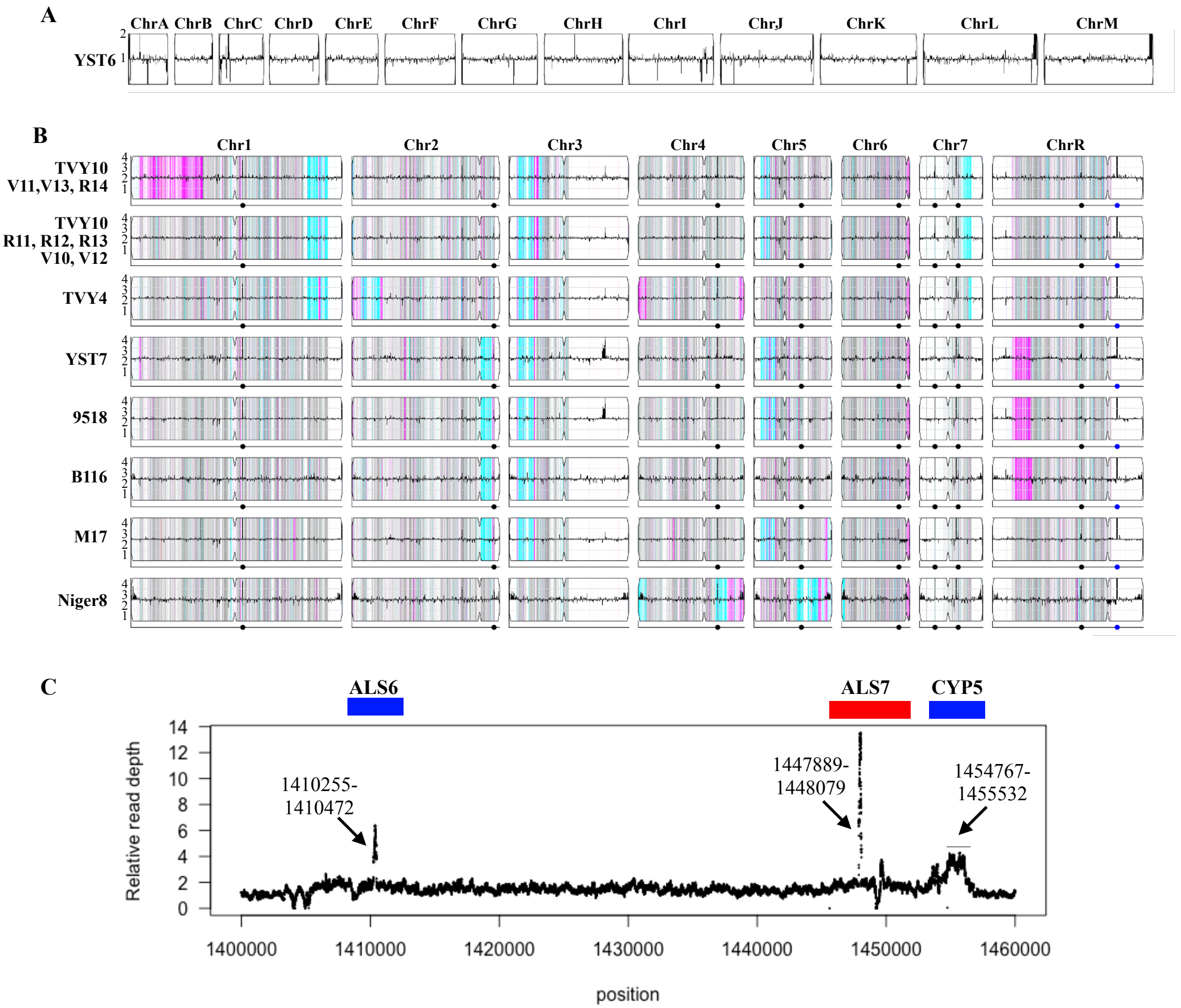

FIG 3 CNV and LOH profiles of THRIVE-yeast. A) Representative trace of a YST6 isolate (*N. glabrata*) and B) TVY10 isolates, representative traces of TVY4 and YST7 isolates, and four isolates that are closely related to YST7. Relative copy number compared to the reference genomes is shown as the horizontal black line with the scale on the y axis. For *C. albicans* isolates, the density of heterozygous SNPs is shown as the vertical lines spanning the height of each chromosome box, with the intensity representing the number of SNPs in 5 kb bins. Heterozygous SNPs are gray, homozygous SNPs are coloured based on the retained SC5314 homolog: cyan for “AA” and magenta for “BB.” White indicates an ancestral LOH in SC5314. For each chromosome the centromere is indicated by an indentation in the box. The dots on the bottom line below each box indicate the positions of major repeat sequences. C) Fine-scale coverage mapping of the putative CNV on chr3. Shown is one representative trace from YST7 R2, all isolates have a similar pattern. Gene positions above the figure are approximated.

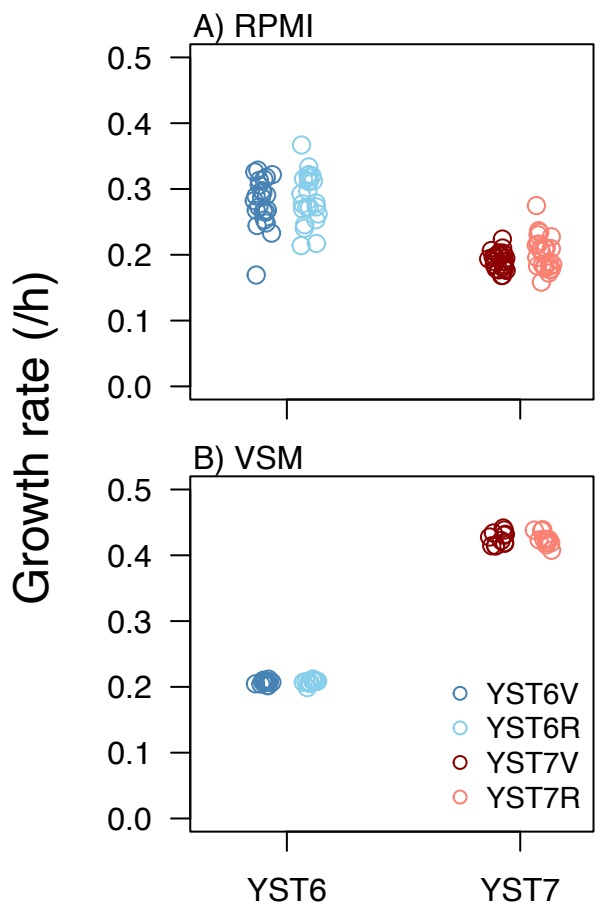

FIG 4 Growth rate was measured from 24 vaginal and 24 rectal isolates from YST6 and YST7. Optical density was recorded every 15 minutes in a plate reader with constant shaking and incubation at 37 °C. Each point represents the mean of two technical replicates for each isolate, 24 isolates were measured for each group. Growth rate was calculated as the spline with the highest slope using a custom R script.

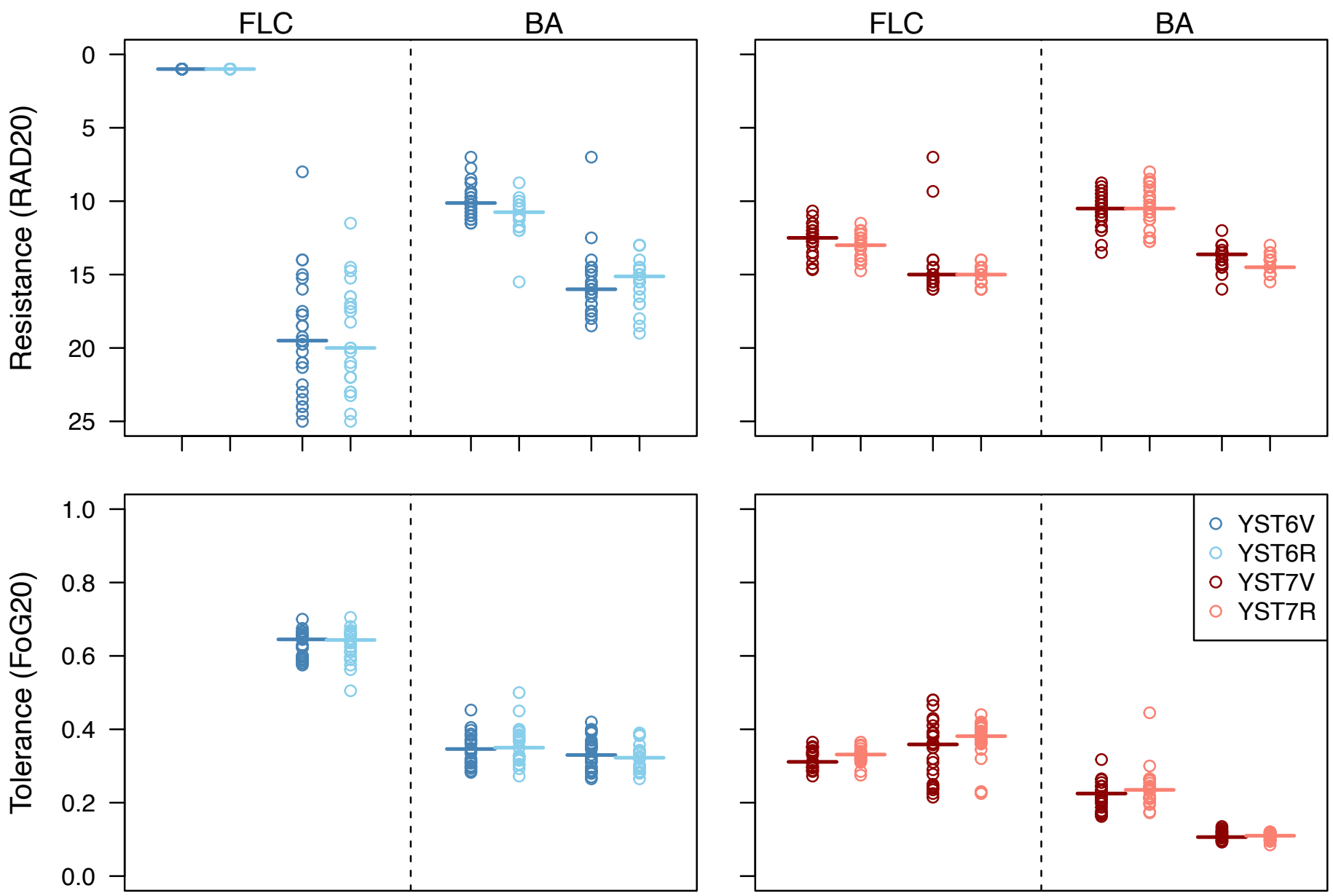

FIG 5 Drug response phenotypes from disk diffusion assays of isolates from YST6 (*N. glabrata*, left) and YST7 (*C. albicans*, right). Drug resistance (top panels) and drug tolerance (bottom panels) was measured from disk diffusion assays for fluconazole (FLC) and boric acid (BA). Drug response was measured on pH 4 and pH 7 Mueller-Hinton plates using the R package *diskImageR* (72), which computationally measures response parameters from images of disk diffusion assays. Each point represents the mean of four replicates (two technical replicates for two biological replicates), with the coloured line indicating the median of 24 isolates in each group.

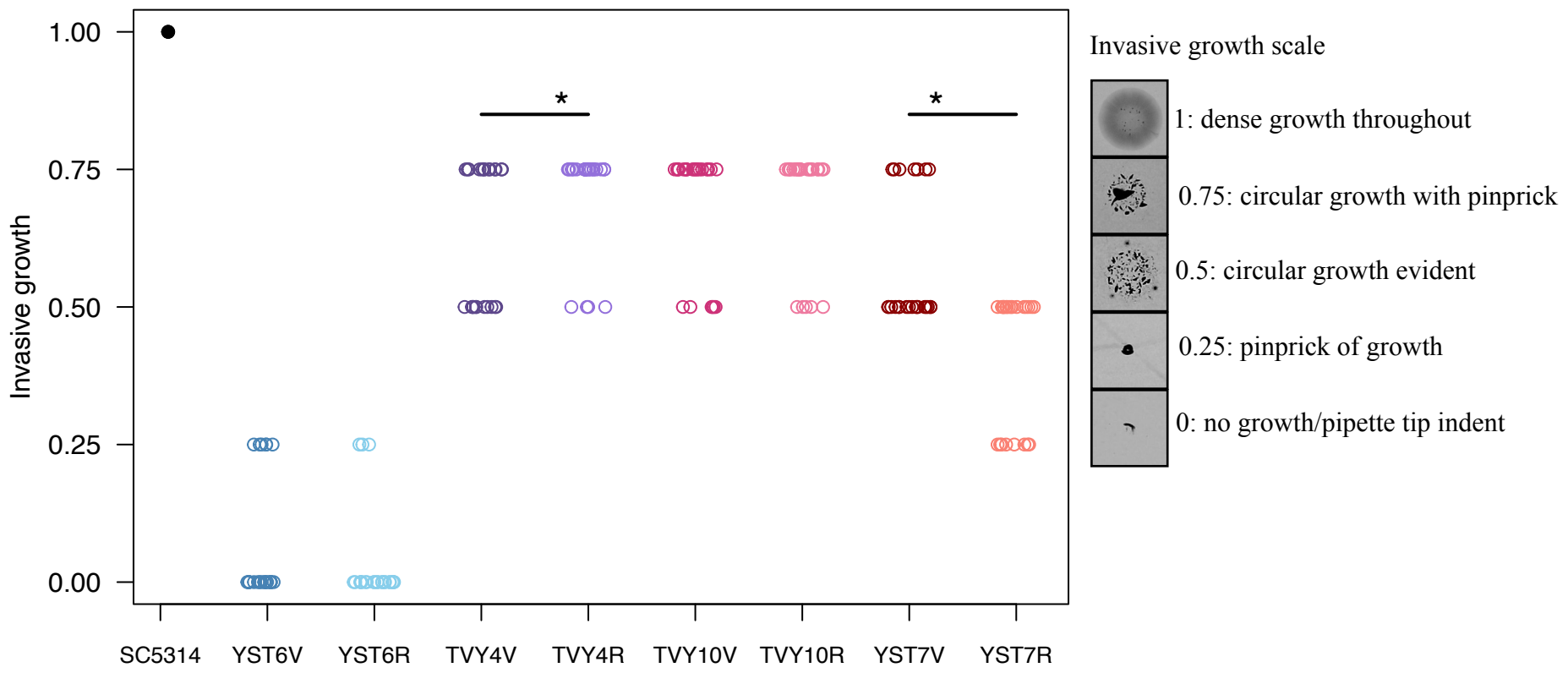

FIG 6 Invasive growth was qualitatively scored after growth on YPD plates for 96 h. Each point indicates the maximum scores between two bioreplicates for each isolate.