New Clade Designation Revealed from Whole Genome Sequencing of *Candida glabrata* Isolates

Brooke Sidney[1] and Aleeza Gerstein[2]

[1] Department of Microbiology, University of Manitoba, Winnipeg, Canada

[2] Departments of Microbiology and Statistics, University of Manitoba, Winnipeg, Canada

Correspondence:

Brooke Sidney

434 Notre Dame Street, Winnipeg, Canada

204-298-1935

sidneyb@myumanitoba.ca

Abstract

Candida glabrata is an increasingly common fungal species that causes mucosal and systemic infections in humans. Despite global distribution of *C. glabrata*, there is not a complete understanding of the phylogenetic structure due to limited sampling and a focus on other *Candida* spp., such as *Candida albicans.* By analyzing the phylogeny of *C. glabrata* isolates from Manitoba, Canada and global isolates we can obtain a better understanding of the distribution of Manitoba isolates. Previously, seven clades were found with whole-genome sequencing (WGS) data of 53 *C. glabrata* isolates from seven countries and three anatomical sites. We expanded on this with WGS data of 126 *C. glabrata* isolates from nine countries and 15 anatomical sites. We whole-genome sequenced 18 clinical *C. glabrata* isolates acquired in 2012-2013 from the major regional hospital in Winnipeg, Manitoba. These isolates were placed into a global phylogenetic context with 107 additional isolates from the NCBI Sequence Read Archive database. We identified 13 clades and Manitoba isolates clustered with other Canadian isolates and were enriched in four clades. Overall, this study expanded on past studies with a larger dataset of *C. glabrata* isolates from more countries and anatomical sites and suggests expanded new clades designations.

Key words: WGS, MLST, aneuploidy, candidiasis, Canada

Candida spp. are part of the normal microflora of humans, but can cause problems If present in immunocompromised or other susceptible hosts (Dabiri *et al.* 2018). In these cases, *Candida* can act as an opportunistic pathogen and cause a fungal infection known as candidiasis (Papon *et al.* 2013; Dabiri *et al.* 2018). The occurrence of *Candida* infections has consistently increased over the past 30 years (Chakrabarti *et al.* 2009; Dabiri *et al.* 2018). It is estimated that *Candida* infections make up approximately 80% of all hospital-acquired fungal infections (Colombo *et al.* 2013). *Candida albicans* causes the majority of *Candida* spp. infections, though non-*albicans* *Candida* species (primarily *Candida glabrata, Candida parapsilosis* and *Candida tropicalis*), are increasing in prevalence (Papon *et al.* 2013; Kumar *et al.* 2014). *C.* glabrata has emerged as an increasingly common species, which causes mucosal and systemic infections and is the second most common cause of candidiasis in the United States (Grenouillet *et al.* 2007; Amanloo *et al.* 2018). Previously, it was found that the incidence of *C. glabrata* increased 3.6% globally from the time periods 1997-2001 to 2015-2016 (Pfaller *et al.* 2019). *C. glabrata* also has decreased susceptibility to all antifungal agents, with resistance to azoles most common (Sanguinetti *et al.* 2005). This makes *C. glabrata* infections more difficult to treat.

Examining the phylogeny of *C. glabrata* isolates allows us to visually observe the distribution and see which isolates are closely related. This is useful if an isolate is known to have antifungal resistance, as other members in the same clade can be examined for a similar antifungal resistance (McManus and Coleman 2014). Geographic

43 enrichment can also be observed from the phylogeny, which shows the distribution of
44 samples globally. This can illustrate any recent dispersal of the *C. glabrata* isolates, and
45 the dispersal cause can be examined.
46 To analyze the phylogenetic structure of microorganisms such as *Candida* spp.*,*
47 the primary methods are multi-locus sequence typing (MLST) and whole-genome
48 sequencing (WGS) (Byun *et al.* 2018; Biswas *et al.* 2018). MLST analysis characterizes
49 chosen isolates based on the DNA sequences of the internal fragments of multiple
50 housekeeping genes. MLST has high discriminatory power and is a fast and easy
51 method for genotyping (Gits-Muselli *et al.* 2020). MLST schemes for *C. glabrata* and
52 other *Candida* species are available through the online PubMLST database (the Public
53 databases for molecular typing and microbial genome diversity, https://pubmlst.org)
54 (Garcia-Hermoso *et al.* 2016). As of April 2021, the PubMLST database has 4464
55 isolates for *C. albicans* and 1376 isolates for *C. glabrata*, with no Canadian *C. glabrata*
56 isolates. This shows a current downside to using MLST, as more time is needed to
57 increase the number and geographical location of isolates in the PubMLST database.
58 Researchers across the world need to add their sequences to PubMLST to properly
59 analyze the global distribution of isolates using MLST. Comparatively, WGS is used to
60 obtain the DNA sequences of entire genomes. Barriers such as time limitations and high
61 costs have limited its used for phylogenetic analysis, but it is better suited to
62 distinguishing differences between isolates and see genome-wide diversity (Ene *et al.*

63 2021.; Kovanen *et al.* 2014). Over the years WGS has become more affordable,
64 potentially increasing its use moving forward (Kovanen *et al.* 2014).
65 Previously, WGS data of 53 *C. glabrata* clinical isolates from seven countries and
66 various anatomical sites was analyzed (Carreté *et al.* 2018). *C. glabrata* is a haploid
67 organism but can have aneuploidy, which can have an effect on mutations that occur
68 during evolution (Tsai and Nelliat 2019). Aneuploidy is an imbalance in the genome due
69 to a loss or gain of chromosomes (Tsai and Nelliat 2019). In a previous study,
70 aneuploidy was found in chromosomes C, E, G and J (Carreté *et al.* 2018, 2019). In the
71 same study, seven major clades were found by constructing a phylogeny using the
72 maximum-likelihood method and using multiple correspondence analysis and model-
73 based clustering (Carreté *et al.* 2018). The seven clades were found to be
74 geographically structured (Carreté *et al.* 2018). In a previous study, it was also found
75 that *C. glabrata* clades tended to group together based on geographic location
76 (Dodgson *et al.* 2005). *C. glabrata* has been found to be globally distributed, and is
77 more commonly isolated in certain geographic locations such as Canada (Colombo *et*
78 *al.* 2003). It has been found that *C. glabrata* isolates are more frequently isolated in
79 Canada than other geographic regions such as Latin America, Asia and South Africa
80 (McTaggart *et al.* 2020). The species distribution of *Candida* spp. including *C. glabrata*
81 in Ontario, Canada was found to remain stable over a four year period (McTaggart *et al.*
82 2020).

Overall, we analyzed the phylogenetic structure of *C. glabrata* isolated from Manitoba, Canada. We whole-genome sequenced 18 clinical *C. glabrata* isolates acquired in 2012-2013 from the major regional hospital in Winnipeg, Manitoba. We placed these isolates into a global phylogenetic context with 107 additional isolates from the Sequence Read Archive database. By placing these isolates into the global phylogeny (Carreté *et al.* 2018), we can determine whether Canadian strains clustered together or were spread throughout.

In addition to the 18 clinical *C. glabrata* isolates from Manitoba, the fastq data of 107 additional isolates from the Sequence Read Archive database were included in this study (Håvelsrud and Gaustad 2017; Biswas *et al.* 2017; Carreté *et al.* 2018; McTaggart *et al.* 2020), along with *C. glabrata* CBS138 as the reference genome and *Candida nivariensis* as the outgroup, both obtained from the National Center for Biotechnology information (https://www.ncbi.nlm.nih.gov). In total, 126 *C. glabrata isolates* were examined (Table S1). DNA extraction using a standard phenol:chloroform protocol was previously performed. WGS libraries were built at the MiGS Sequencing Center (Pittsburgh, USA) and run on the Illumina NextSeq 550 platform. Isolates were sequenced to a depth of 40X.

This research was completed in part by using Compute Canada (www.computecanada.ca). A quality check of all the reads was completed before and after trimming using FASTQC and MultiQC (Simon 2010; Ewels *et al.* 2016). For all programs, the default settings were used unless otherwise specified. Reads were pre-processed before assembly using Trimmomatic (v0.36) using parameters previously

105 used by Todd *et al.* (2019) (LEADING: 10 TRAILING: 3 SLIDINGWINDOW:4:15
106 MINLEN: 31 TOPHRED33) (Bolger *et al.* 2014). The *Candida glabrata* reference
107 genome CBS138 (ASM254v2) was downloaded from NCBI
108 (https://www.ncbi.nlm.nih.gov) on January 5th 2021. The CBS138 reference genome
109 was indexed using Samtools faidx and Burrows-Wheeler Aligner Index (BWA) (v0.7.17)
110 (Li *et al.* 2009; Li 2013). The reads were aligned to the CBS138 reference genome
111 using the BWA-MEM algorithm (v0.7.17) (Li 2013).
112 The alignment files were processed using Picard tools (v2.23.3) (Picard Toolkit
113 2019). Briefly, read groups were added based on the different batches of strains from
114 other studies obtained from NCBI using AddorReplaceReadGroups (RGID=cell1
115 RGLB=Lib1 RGPL=Illumina RGPU=unit1). ValidateSamFile was completed and
116 MarkDuplicates was used to deduplicate reads. FixMateInformation was used to verify
117 mate-pair information.
118 Coverage analysis was performed using BEDtools genomecov (Quinlan 2014).
119 The depth for each genome position with one-based coordinates was reported, and
120 BAM files were set as the input. Mitochondrial DNA was removed from the coverage
121 analysis output file. Using a custom script, the coverage (ploidy) analysis was done
122 using R (v1.3.1093) (R Core Team 2020). Briefly, the script determines the average
123 coverage over a 5 kb sliding window per chromosome.
124 Variants were analyzed using SAMtools (Li *et al.* 2009). All the alignment files
125 were merged using SAMtools merge, creating a multi-sample BAM file. BCFtools
126 mpileup (-Ou --gvcf 5 --max-dept 10000 -I) and BCFtools call (ploidy set to one as *C.*

glabrata is haploid) were then used to call the variants generating a multi-sample VCF file (Letunic and Bork 2019; Danecek *et al.* 2021). The SNPs per isolate of *Candida glabrata* based on CBS138 reference genome was visualized using Microsoft Excel (v16.43) (Microsoft Corporation 2020).

A phylogenetic tree was constructed using 126 *C. glabrata* isolates and *C. nivariensis* as the outgroup. The multi-sample VCF file was converted to a fasta alignment using an available python script (vcf2phylip.py) from github (https://github.com/edgardomortiz/vcf2phylip/blob/master/vcf2phylip.py) (v2.7.16). FastTree (v2.1.10) was then used to obtain a maximum-likelihood phylogenetic tree using the general time reversible model (Price *et al.* 2010). TreeCluster (v1.0.3) was used to cluster the sequences based on similarity to predict the number of clades (-t 0.0115, -m length_clade) (Balaban *et al.* 2019). TreeCluster is optimized for HIV, so the threshold was modified to identify the window where the cluster number began to stabilize. The clustering method length_clade was chosen which follows conditions such that the cluster does not have edges above the threshold, the leaves can't be linked to branches with support equal to or less than the support (left as default) and the leaves in the cluster must define a clade. The phylogenetic tree was visualized and annotated using the Interactive Tree Of Life (iTOL, https://itol.embl.de) (Letunic and Bork 2019).

Following bioinformatic analysis of the WGS data to identify unique variants in each genome, the phylogeny of *C. glabrata* isolates, karyotypic variations and SNP counts were examined. To identify genetic relatedness among *C. glabrata* isolates, a maximum-likelihood phylogenetic reconstruction was completed based on genome-wide

SNPs. Our work on an expanded dataset illustrates the importance of using larger isolate sets with more globally distributed isolates. From the phylogeny constructed, 13 clades were identified (Figure 1). This shows that our larger dataset expanded and modified the previous seven-clade phylogenetic structure identified by Carreté *et al.* (2018). In the new phylogeny, seven clades contained isolates from a single continent with clades I, IV, V, XI, XII from Australia and clades IX, X from North America. This is influenced by that fact that 47.6% of the isolates are from Australia. Isolates from Manitoba were not widely distributed, as they were present in only four of the 13 clades (clades III, VI, IX and XIII). Interestingly, the Canadian isolates from Manitoba (12) and from Ontario (9) tended to cluster near each other throughout the phylogenetic tree despite the geographical distance of approximately 950 km. This could be due to numerous factors, such as human migration or travel patterns. SNPs for some isolates varied by more than 20,000, and isolates within the same clade had similar SNP counts (Figure 2). The SNPs are based on relatedness to the CBS138 reference genome and the results are similar to a previous study (Carreté *et al.* 2018). SNPs for all of the isolates ranged from 4.34-6.16 SNPs/kb and Carreté *et al.* (2018) found SNPs ranged from 4.66-6.56 SNPs/kb. These results also indicate a relatively deep split between clades I to IV and clades V to XIII. Because only one year of data was collected from Manitoba, additional data will be obtained and compared to examine whether the clade distribution is stable in the future.

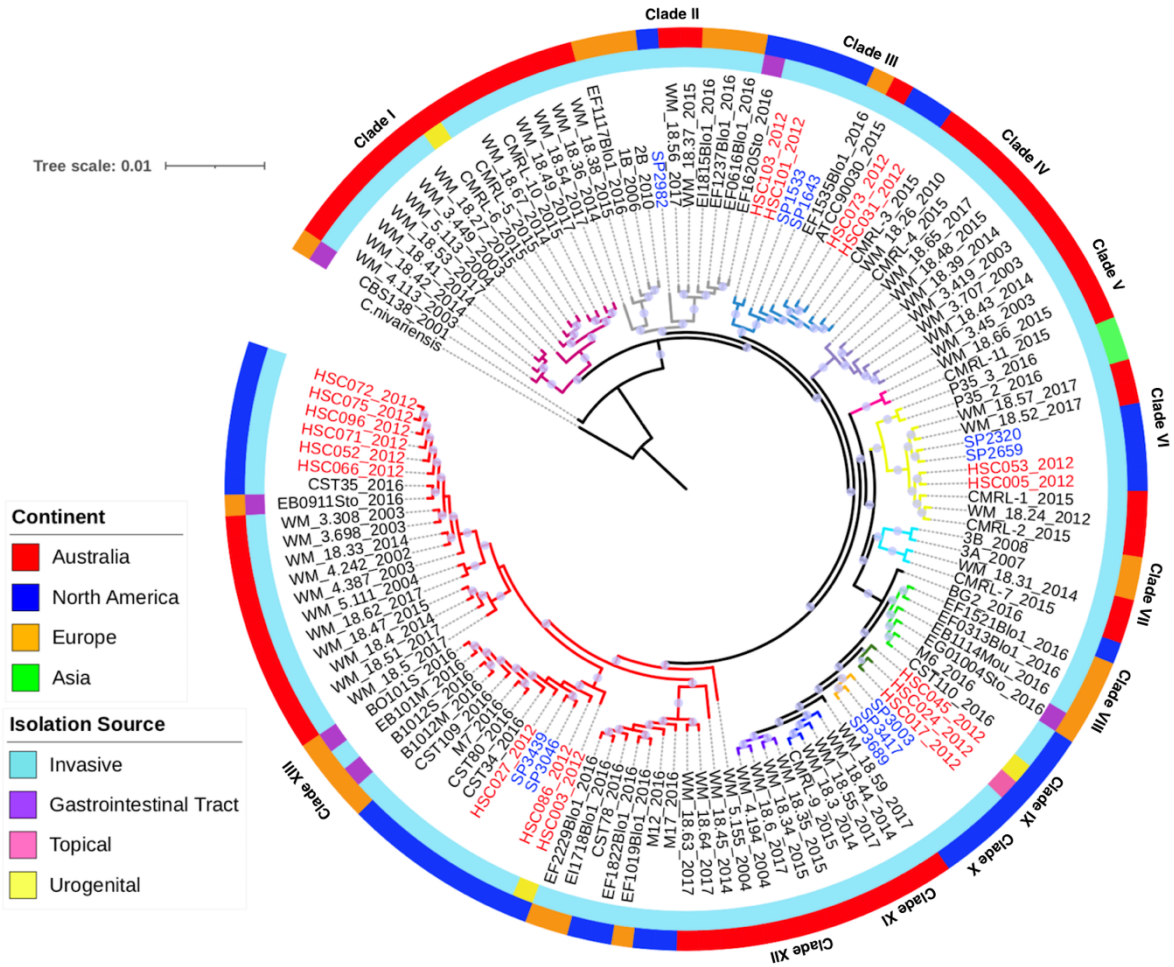


Figure 1. Maximum-likelihood phylogenetic tree for *Candida glabrata* isolates. The phylogenetic tree was constructed using 126 *Candida glabrata* isolates plus *Candida nivariensis*, as the outgroup and was based on genome-wide SNPs. Isolates in red are from Manitoba, and those in blue are from Ontario. Different colours in the center of the phylogenetic tree designate the 13 clades identified using TreeCluster.

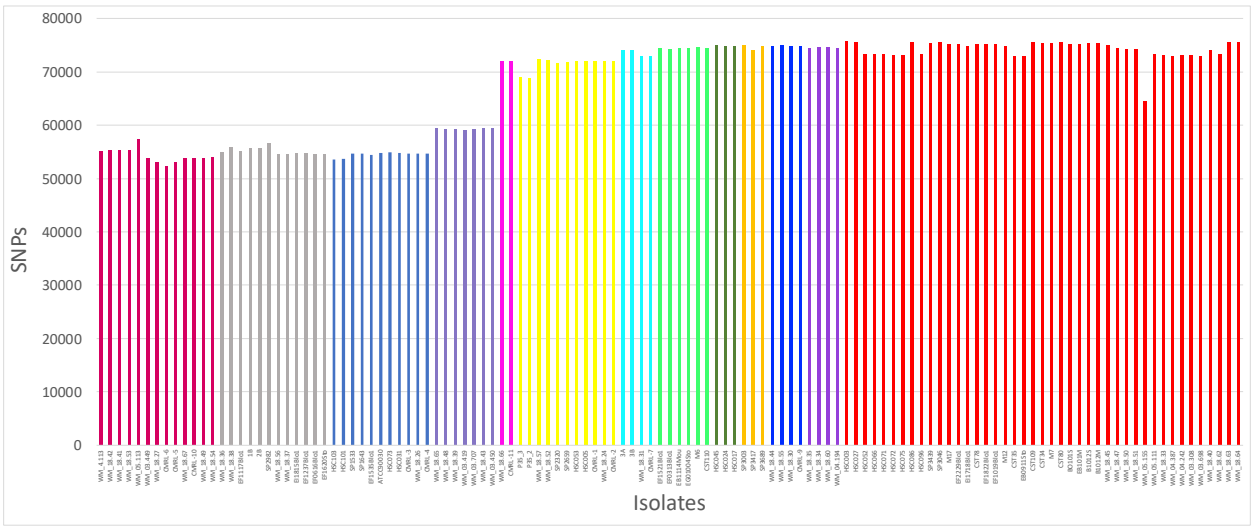


Figure 2. The number of single nucleotide polymorphisms (SNPs) per isolate of *Candida glabrata* based on the CBS138 reference genome.

Each coloured bar is a single sequenced isolate; the colours indicate the clade, with clade I to XIII from left to right.

The majority of isolates analyzed were haploid, but aneuploidy was present in 16.1% of the isolates. Representative aneuploidies are seen in Figure 3. The most common aneuploidy was chrE which was identified in 18 isolates (Table 1). In six isolates, chrE and additional chromosomes were aneuploid (chrC: two isolates, chr M: three isolates, chrG: one isolate). Previously, aneuploidy was examined by only one study (Carreté et al. 2018). A whole duplication of chromosome E and a partial aneuploidy of chromosome G was identified (Carreté *et al.* 2018). 36.4% of isolates from Biswas *et al.* (2017) were aneuploid (CMRL-9 and CMRL-7 with chrE, CMRL-2 with chrC and chrE and CMRL-10 with chrE and chrM) (Figure S1 A). 12.5% of isolates from Carreté *et al.* (2018) were aneuploid (EI1718Blo1, EF1822Blo1 and EF0616Blo1 with chrE and EF1521Blo1 with chrG) (Figure S1 B/C). 26.5% of isolates from Biswas *et*

al. (2018) were aneuploid (WM_18.45, WM_18.5, WM_18.3, WM_18.31, WM_18.51, WM_18.42 and WM_18.34 and WM_18.6 with chrE, WM_18.24 with chrC and chrE, WM_18.44 with chrE and chrM and WM_18.67 with chrE, chrG and chrM) (Figure S1 D/E/F). Håvelsrud *et al.* (2017) (Figure S1 G) and Mctaggart *et al.* (2020) (Figure S1 H) had no aneuploidy. Of the total aneuploidy identified, 80% was from Australia, 10% from France, 5% from Italy and 5% from the USA. Additionally, from the total aneuploidy identified, 90% was isolated from blood, 5% from tissue and 5% from the pelvis. All of the Manitoba isolates were haploid with no aneuploidy present (Figure S1 I). This absence of aneuploidy in the Manitoba isolates could be a result of the low number of isolates (Ene *et al.* 2021).

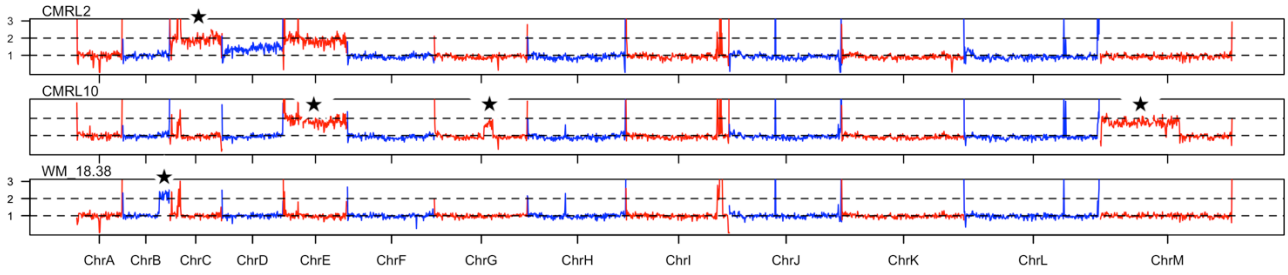


Figure 3. Unique aneuploidies among *C. glabrata* isolates. These are representative plots of the coverage. The coverage was identified using BEDtools genomecov and a custom R script that determined the average coverage over a 5 kb sliding window per chromosome.

212 **Table 1.** Aneuploidy identified among *C. glabrata* isolates.

Isolate Name	Aneuploidy	Citation
CMRL-2	chrC, chrE	Biswas *et al.* 2017
CMRL-9, CMRL-7	chrE	Biswas *et al.* 2017
CMRL-10	chrE, chrM	Biswas *et al.* 2017
EF1521Blo1	chrG	Carreté *et al.* 2018
EI1718Blo1, EF0616Blo1, EF1822Blo1	chrE	Carreté *et al.* 2018
WM_18.45, WM_18.5, WM_18.3, WM_18.31, WM_18.51, WM_18.42, WM_18.34, WM_18.6	chrE	Biswas *et al.* 2018
WM_18.24	chrC, chrE	Biswas *et al.* 2018
WM_18.44	chrE, chr M	Biswas *et al.* 2018
WM_18.67	chrE, chrG, chrM	Biswas *et al.* 2018
WM_18.38	chrB	Biswas *et al.* 2018
WM_18.41	chrE, chrM	Biswas *et al.* 2018

213

214 Overall, a maximum-likelihood phylogenetic tree of 126 *C. glabrata* isolates plus
215 *C. nivariensis* as the outgroup was constructed. We identified 13 clades expanding from
216 seven clades identified in a previous study (Carreté *et al.* 2018). Manitoba isolates
217 clustered with previously sequenced isolates from Ontario and were found in only four
218 clades. This work on an expanded dataset shows the importance of using larger isolate
219 sets with more globally distributed isolates. In the future, with access to additional
220 Manitoba isolates from different years, we will determine if the clade distribution is
221 stable.

References

Amanloo, S., Shams-Ghahfarokhi, M., Ghahri, M., and Razzaghi-Abyaneh, M. 2018. Genotyping of clinical isolates of *Candida glabrata* from Iran by multilocus sequence typing and determination of population structure and drug resistance profile. Med. Mycol. **56**(2): 207–215. doi:10.1093/mmy/myx030.

Balaban, M., Moshiri, N., Mai, U., Jia, X., and Mirarab, S. 2019. TreeCluster: Clustering biological sequences using phylogenetic trees. PLoS One **14**(8): e0221068. doi:10.1371/journal.pone.0221068.

Biswas, C., Chen, S.C.-A., Halliday, C., Kennedy, K., Playford, E.G., Marriott, D.J., *et al.* 2017. Identification of genetic markers of resistance to echinocandins, azoles and 5-fluorocytosine in *Candida glabrata* by next-generation sequencing: a feasibility study. Clin. Microbiol. Infect. **23**(9): 676.e7-676.e10. doi:10.1016/j.cmi.2017.03.014.

Biswas, C., Marcelino, V.R., Van Hal, S., Halliday, C., Martinez, E., Wang, Q., *et al*. 2018. Whole genome sequencing of Australian *Candida glabrata* isolates reveals genetic diversity and novel sequence types. Front. Microbiol. **9**: 2946. doi:10.3389/fmicb.2018.02946.

Bolger, A.M., Lohse, M., and Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30**(15): 2114–2120. doi:10.1093/bioinformatics/btu170.

Byun, S.A., Won, E.J., Kim, M.-N., Lee, W.G., Lee, K., Lee, H.S., *et al.* 2018. Multilocus sequence typing (MLST) genotypes of *Candida glabrata* bloodstream isolates in

Korea: association With antifungal resistance, mutations in mismatch repair gene (Msh2), and clinical outcomes. Front. Microbiol. **9**: 1523. doi:10.3389/fmicb.2018.01523.

Carreté, L., Ksiezopolska, E., Gómez-Molero, E., Angoulvant, A., Bader, O., Fairhead, C., *et al*. 2019. Genome comparisons of *Candida glabrata* serial clinical isolates reveal patterns of genetic variation in infecting clonal populations. Front. Microbiol. **10**: 112. doi:10.3389/fmicb.2019.00112.

Carreté, L., Ksiezopolska, E., Pegueroles, C., Gómez-Molero, E., Saus, E., Iraola-Guzmán, S., *et al*. 2018. Patterns of genomic variation in the opportunistic pathogen *Candida glabrata* suggest the existence of mating and a secondary association with humans. Curr. Biol. **28**(1): 15-27.e7. doi:10.1016/j.cub.2017.11.027.

Chakrabarti, A., Chatterjee, S.S., Rao, K.L.N., Zameer, M.M., Shivaprakash, M.R., Singhi, S., *et al*. 2009. Recent experience with fungaemia: change in species distribution and azole resistance. Scand. J. Infect. Dis. **41**(4): 275–284. doi:10.1080/00365540902777105.

Colombo, A.L., Guimarães, T., Camargo, L.F.A., Richtmann, R., Queiroz-Telles, Fd., Salles, M.J.C., *et al.* 2013. Brazilian guidelines for the management of candidiasis - a joint meeting report of three medical societies: Sociedade Brasileira de Infectologia, Sociedade Paulista de Infectologia and Sociedade Brasileira de Medicina Tropical. Braz. J. Infect. Dis. **17**(3): 283–312. doi:10.1016/j.bjid.2013.02.001.

Colombo, A.L., Perfect, J., DiNubile, M., Bartizal, K., Motyl, M., Hicks, P., *et al*. 2003. Global distribution and outcomes for *Candida* species causing invasive candidiasis: results from an international randomized double-blind study of caspofungin versus amphotericin B for the treatment of invasive candidiasis. Eur. J. Clin. Microbiol. Infect. Dis. **22**(8): 470–474. doi:10.1007/s10096-003-0973-8.

Dabiri, S., Shams-Ghahfarokhi, M., and Razzaghi-Abyaneh, M. 2018. Comparative analysis of proteinase, phospholipase, hydrophobicity and biofilm forming ability in *Candida* species isolated from clinical specimens. J. Mycol. Med. **28**(3): 437–442. doi:10.1016/j.mycmed.2018.04.009.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., *et al*. 2021. Twelve years of SAMtools and BCFtools. Gigascience **10**(2). doi:10.1093/gigascience/giab008.

Dodgson, A.R., Pujol, C., Pfaller, M.A., Denning, D.W., and Soll, D.R. 2005. Evidence for recombination in *Candida glabrata.* Fungal Genet. Biol. **42**(3): 233–243. doi:10.1016/j.fgb.2004.11.010.

Ene, I.V., Hickman, M.A., and Gerstein, A.C. (2021). The interplay between neutral and adaptive processes shapes genetic variation during *Candida* species evolution. Manuscript submitted for publication.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics **32**(19): 3047–3048. doi:10.1093/bioinformatics/btw354.

Garcia-Hermoso, D., Desnos-Ollivier, M., and Bretagne, S. 2016. Typing *Candida*

species Using microsatellite length polymorphism and multilocus sequence typing. *In Candida* Species: Methods and Protocols. *Edited by* R. Calderone and R. Cihlar. Springer New York, New York, NY. pp. 199–214. doi:10.1007/978-1-4939-3052-4_15.

Gits-Muselli, M., Campagne, P., Desnos-Ollivier, M., Le Pape, P., Bretagne, S., Morio, F., *et al.* 2020. Comparison of multiLocus sequence typing (MLST) and microsatellite length polymorphism (MLP) for pneumocystis jirovecii genotyping. Comput. Struct. Biotechnol. J. **18**: 2890–2896. doi:10.1016/j.csbj.2020.10.005.

Grenouillet, F., Millon, L., Bart, J.-M., Roussel, S., Biot, I., Didier, E., *et al.* 2007. Multiple-locus variable-number tandem-repeat analysis for rapid typing of *Candida glabrata.* J. Clin. Microbiol. **45**(11): 3781–3784. doi:10.1128/JCM.01603-07.

Håvelsrud, O.E., and Gaustad, P. 2017. Draft Genome Sequences of *Candida glabrata* Isolates 1A, 1B, 2A, 2B, 3A, and 3B. Genome Announc. **5**(10). doi:10.1128/genomeA.00328-16.

Kovanen, S.M., Kivistö, R.I., Rossi, M., Schott, T., Kärkkäinen, U.-M., Tuuminen, T., *et al*. 2014. Multilocus sequence typing (MLST) and whole-genome MLST of Campylobacter jejuni isolates from human infections in three districts during a seasonal peak in Finland. J. Clin. Microbiol. **52**(12): 4147–4154. doi:10.1128/JCM.01959-14.

Kumar, A., Sharma, P.C., Kumar, A., and Negi, V. 2014. A study on phenotypic traits of *Candida* species isolated from blood stream infections and their in vitro

susceptibility to fluconazole. J. Med. Sci. **7**(1): 83–91.

Letunic, I., and Bork, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. **47**(W1): W256–W259. doi:10.1093/nar/gkz239.

Li, H. 2013, March 16. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Available from http://arxiv.org/abs/1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., *et al*. 2009. The sequence alignment/map format and SAMtools. Bioinformatics **25**(16): 2078–2079. doi:10.1093/bioinformatics/btp352.

McManus, B.A., and Coleman, D.C. 2014. Molecular epidemiology, phylogeny and evolution of *Candida albicans*. Infect. Genet. Evol. **21**: 166–178. doi:10.1016/j.meegid.2013.11.008.

McTaggart, L.R., Cabrera, A., Cronin, K., and Kus, J.V. 2020. Antifungal susceptibility of clinical yeast isolates from a large canadian reference laboratory and application of whole-genome sequence analysis to elucidate mechanisms of acquired resistance. Antimicrob. Agents Chemother. **64**(9). doi:10.1128/AAC.00402-20.

Microsoft Corporation. 2020. Microsoft Excel. Available from https://www.microsoft.com/en-ca/microsoft-365/excel?legRedir=true&CorrelationId=07bd3cfd-9054-4671-9bd8-3cff61083339&rtc=1.

Papon, N., Courdavault, V., Clastre, M., and Bennett, R.J. 2013. Emerging and emerged pathogenic *Candida* species: beyond the *Candida albicans* paradigm.

PLoS Pathog. **9**(9): e1003550. doi:10.1371/journal.ppat.1003550.

Pfaller, M.A., Diekema, D.J., Turnidge, J.D., Castanheira, M., and Jones, R.N. 2019. Twenty Years of the SENTRY Antifungal Surveillance Program: Results for *Candida* Species From 1997-2016. Open Forum Infect Dis **6**(Suppl 1): S79–S94. doi:10.1093/ofid/ofy358.

Picard Toolkit. 2019. Broad Institute, GitHub Repository. Available from https://broadinstitute.github.io/picard/ [accessed 10 March 2021].

Price, M.N., Dehal, P.S., and Arkin, A.P. 2010. FastTree 2--approximately maximum-likelihood trees for large alignments. PLoS One **5**(3): e9490. doi:10.1371/journal.pone.0009490.

Quinlan, A.R. 2014. BEDTools: The swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics **47**: 11.12.1-34. doi:10.1002/0471250953.bi1112s47.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/ [accessed 2 March 2021].

Sanguinetti, M., Posteraro, B., Fiori, B., Ranno, S., Torelli, R., and Fadda, G. 2005. Mechanisms of azole resistance in clinical isolates of *Candida glabrata* collected during a hospital survey of antifungal resistance. Antimicrob. Agents Chemother. **49**(2): 668–679. doi:10.1128/AAC.49.2.668-679.2005.

Simon, A. 2010. FastQC: A quality control tool for high throughput sequence data. Babraham Institute, Cambridge, United Kingdom. Available from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [accessed 2 March

2021].

Todd, R.T., Wikoff, T.D., Forche, A., and Selmecki, A. 2019. Genome plasticity in *Candida albicans* is driven by long repeat sequences. Elife **8**: e45954. doi:10.7554/eLife.45954.

Tsai, H.-J., and Nelliat, A. 2019. A Double-Edged Sword: Aneuploidy is a Prevalent Strategy in Fungal Adaptation. Genes **10**(10). doi:10.3390/genes10100787.

Supplementary Information

Table S1. Information on the *Candida glabrata* isolates used in this study.

Sample	Clade	Country of Isolation	Anatomical Site[1]	Citation
WM_4.113	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_18.42	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_18.41	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_18.53	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_05.113	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_03.449	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_18.27	Clade I	Australia	Blood	Biswas *et al.* 2018
CMRL-6	Clade I	Australia	Blood	Biswas *et al.* 2017
CMRL-5	Clade I	Australia	Blood	Biswas *et al.* 2017
WM_18.67	Clade I	Australia	Tissue	Biswas *et al.* 2018
CMRL-10	Clade I	Australia	Pelvis	Biswas *et al.* 2017
WM_18.49	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_18.54	Clade I	Australia	Blood	Biswas *et al.* 2018
WM_18.36	Clade II	Australia	Blood	Biswas *et al.* 2018
WM_18.38	Clade II	Australia	Blood	Biswas *et al.* 2018
EF1117Blo1	Clade II	France	Blood	Carreté *et al.* 2018
1B	Clade II	Norway	Blood	Håvelsrud *et al.* 2017
2B	Clade II	Norway	Blood	Håvelsrud *et al.* 2017
SP2982	Clade II	Canada	Peritoneal fluid	Mctaggart *et al.* 2020
WM_18.56	Clade II	Australia	Blood	Biswas *et al.* 2018
WM_18.37	Clade II	Australia	Blood	Biswas *et al.* 2018
EI1815Blo1	Clade II	Italy	Blood	Carreté *et al.* 2018
EF1237Blo1	Clade II	France	Blood	Carreté *et al.* 2018
EF0616Blo1	Clade II	France	Blood	Carreté *et al.* 2018
EF1620Sto	Clade II	France	Stool	Carreté *et al.* 2018
HSC103	Clade III	Canada	Pelvis	This study
HSC101	Clade III	Canada	Blood	This study
SP1533	Clade III	Canada	Abscess	Mctaggart *et al.* 2020
SP1643	Clade III	Canada	Abdomen	Mctaggart *et al.* 2020
EF1535Blo1	Clade III	France	Blood	Carreté *et al.* 2018
ATCC90030	Clade III	Australia	Blood	Biswas *et al.* 2017
HSC073	Clade III	Canada	Kidney	This study
HSC031	Clade III	Canada	Lung	This study
CMRL-3	Clade III	Australia	Blood	Biswas *et al.* 2017
WM_18.26	Clade III	Australia	Blood	Biswas *et al.* 2018

CMRL-4	Clade III	Australia	Blood	Biswas *et al.* 2017
WM_18.65	Clade IV	Australia	Blood	Biswas *et al.* 2018
WM_18.48	Clade IV	Australia	Blood	Biswas *et al.* 2018
WM_18.39	Clade IV	Australia	Blood	Biswas *et al.* 2018
WM_03.419	Clade IV	Australia	Blood	Biswas *et al.* 2018
WM_03.707	Clade IV	Australia	Blood	Biswas *et al.* 2018
WM_18.43	Clade IV	Australia	Blood	Biswas *et al.* 2018
WM_03.450	Clade IV	Australia	Blood	Biswas *et al.* 2018
WM_18.66	Clade V	Australia	Blood	Biswas *et al.* 2018
CMRL-11	Clade V	Australia	Urine	Biswas *et al.* 2017
P35_3	Clade VI	Taiwan	Mouth	Carreté *et al.* 2018
P35_2	Clade VI	Taiwan	Mouth	Carreté *et al.* 2018
WM_18.57	Clade VI	Australia	Blood	Biswas *et al.* 2018
WM_18.52	Clade VI	Australia	Body fluid	Biswas *et al.* 2018
SP2320	Clade VI	Canada	Blood	Mctaggart *et al.* 2020
SP2659	Clade VI	Canada	Blood	Mctaggart *et al.* 2020
HSC053	Clade VI	Canada	Abdomen	This study
HSC005	Clade VI	Canada	Abdomen	This study
CMRL-1	Clade VI	Australia	Blood	Biswas *et al.* 2017
WM_18.24	Clade VI	Australia	Blood	Biswas *et al.* 2018
CMRL-2	Clade VI	Australia	Blood	Biswas *et al.* 2017
3A	Clade VII	Norway	Blood	Håvelsrud *et al.* 2017
3B	Clade VII	Norway	Blood	Håvelsrud *et al.* 2017
WM_18.31	Clade VII	Australia	Blood	Biswas *et al.* 2018
CMRL-7	Clade VII	Australia	Blood	Biswas *et al.* 2017
EF1521Blo1	Clade VIII	France	Blood	Carreté *et al.* 2018
EF0313Blo1	Clade VIII	France	Blood	Carreté *et al.* 2018
EB1114Mou	Clade VIII	Belgium	Mouth	Carreté *et al.* 2018
EG01004Sto	Clade VIII	Germany	Stool	Carreté *et al.* 2018
M6	Clade VIII	USA	Blood	Carreté *et al.* 2018
CST110	Clade VIII	USA	Blood	Carreté *et al.* 2018
HSC045	Clade IX	Canada	Urine	This study
HSC024	Clade IX	Canada	Neck	This study
HSC017	Clade IX	Canada	Abdomen	This study
SP3003	Clade X	Canada	Blood	Mctaggart *et al.* 2020
SP3417	Clade X	Canada	Blood	Mctaggart *et al.* 2020
SP3689	Clade X	Canada	Blood	Mctaggart *et al.* 2020
WM_18.44	Clade XI	Australia	Blood	Biswas *et al.* 2018
WM_18.55	Clade XI	Australia	Blood	Biswas *et al.* 2018
WM_18.30	Clade XI	Australia	Blood	Biswas *et al.* 2018

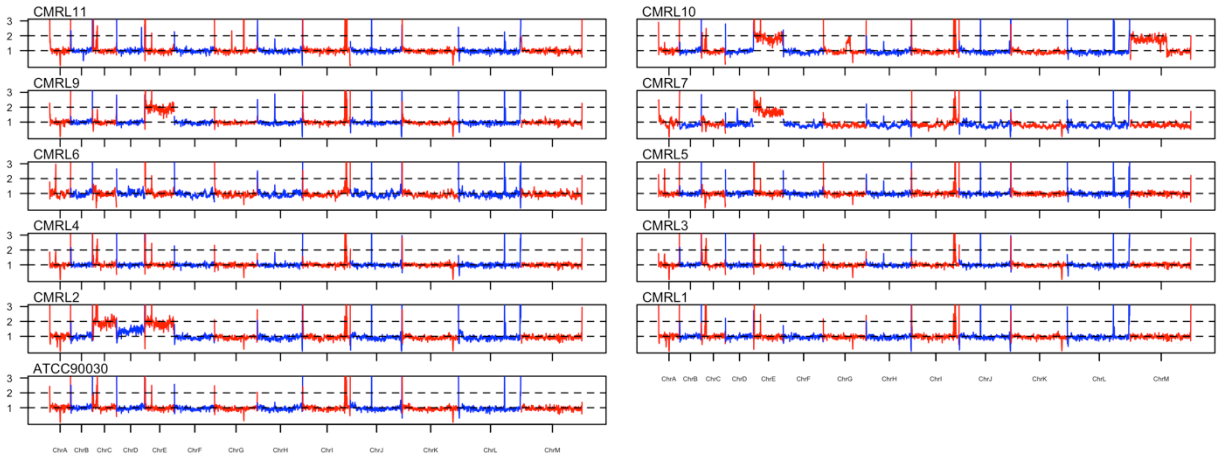
CMRL-9	Clade XI	Australia	Blood	Biswas *et al.* 2017
WM_18.35	Clade XII	Australia	Blood	Biswas *et al.* 2018
WM_18.34	Clade XII	Australia	Blood	Biswas *et al.* 2018
WM_18.60	Clade XII	Australia	Blood	Biswas *et al.* 2018
WM_04.194	Clade XII	Australia	Blood	Biswas *et al.* 2018
HSC003	Clade XIII	Canada	Bladder	This study
HSC027	Clade XIII	Canada	Lung	This study
HSC052	Clade XIII	Canada	Lung	This study
HSC066	Clade XIII	Canada	Lymph node	This study
HSC071	Clade XIII	Canada	Lung	This study
HSC072	Clade XIII	Canada	Lymph node	This study
HSC075	Clade XIII	Canada	Lymph node	This study
HSC086	Clade XIII	Canada	Blood	This study
HSC096	Clade XIII	Canada	Blood	This study
SP3439	Clade XIII	Canada	Blood	Mctaggart *et al.* 2020
SP3046	Clade XIII	Canada	Blood	Mctaggart *et al.* 2020
M17	Clade XIII	USA	Blood	Carreté *et al.* 2018
EF2229Blo1	Clade XIII	France	Blood	Carreté *et al.* 2018
EI1718Blo1	Clade XIII	Italy	Blood	Carreté *et al.* 2018
CST78	Clade XIII	USA	Blood	Carreté *et al.* 2018
EF1822Blo1	Clade XIII	USA	Blood	Carreté *et al.* 2018
EF1019Blo1	Clade XIII	France	Blood	Carreté *et al.* 2018
M12	Clade XIII	USA	Blood	Carreté *et al.* 2018
CST35	Clade XIII	USA	Blood	Carreté *et al.* 2018
EB0911Sto	Clade XIII	Belgium	Stool	Carreté *et al.* 2018
CST109	Clade XIII	USA	Blood	Carreté *et al.* 2018
CST34	Clade XIII	USA	Blood	Carreté *et al.* 2018
M7	Clade XIII	USA	Blood	Carreté *et al.* 2018
CST80	Clade XIII	USA	Blood	Carreté *et al.* 2018
BO101S	Clade XIII	Belgium	Stool	Carreté *et al.* 2018
EB101M	Clade XIII	Belgium	Mouth	Carreté *et al.* 2018
B1012S	Clade XIII	Belgium	Stool	Carreté *et al.* 2018
B1012M	Clade XIII	Belgium	Mouth	Carreté *et al.* 2018
WM_18.45	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.47	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.50	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.51	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_05.155	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_05.111	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.33	Clade XIII	Australia	Blood	Biswas *et al.* 2018

WM_04.387	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_04.242	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_03.308	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_03.698	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.40	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.62	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.63	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.64	Clade XIII	Australia	Blood	Biswas *et al.* 2018
WM_18.59	n/a	Australia	Blood	Biswas *et al.* 2018
BG2	n/a	USA	Blood	Carreté *et al.* 2018
CBS138	n/a	Belgium	Stool	Carreté *et al.* 2018

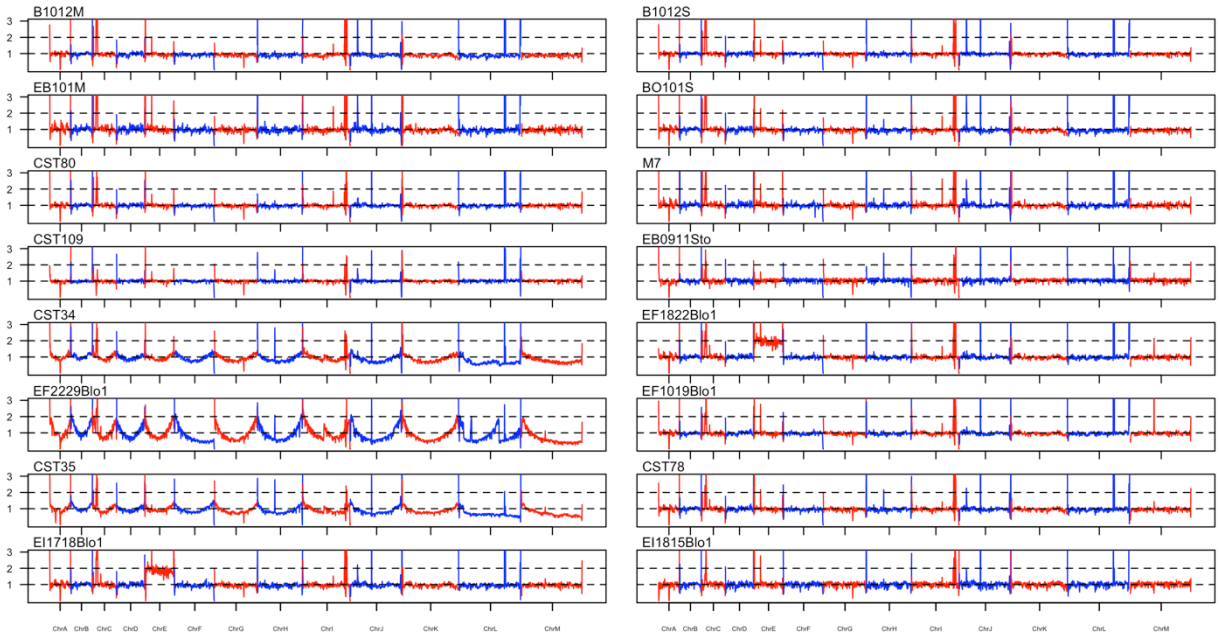
1. Blood, tissue, pelvis, peritoneal fluid, abscess, abdomen, kidney, lung, mouth, lymph node and body fluid are classified as an invasive anatomical site. Stool is classified as a gastrointestinal tract anatomical site. Neck is classified as a topical anatomical site. Bladder and urine are classified as urogenital anatomical site.

Supplementary Figures

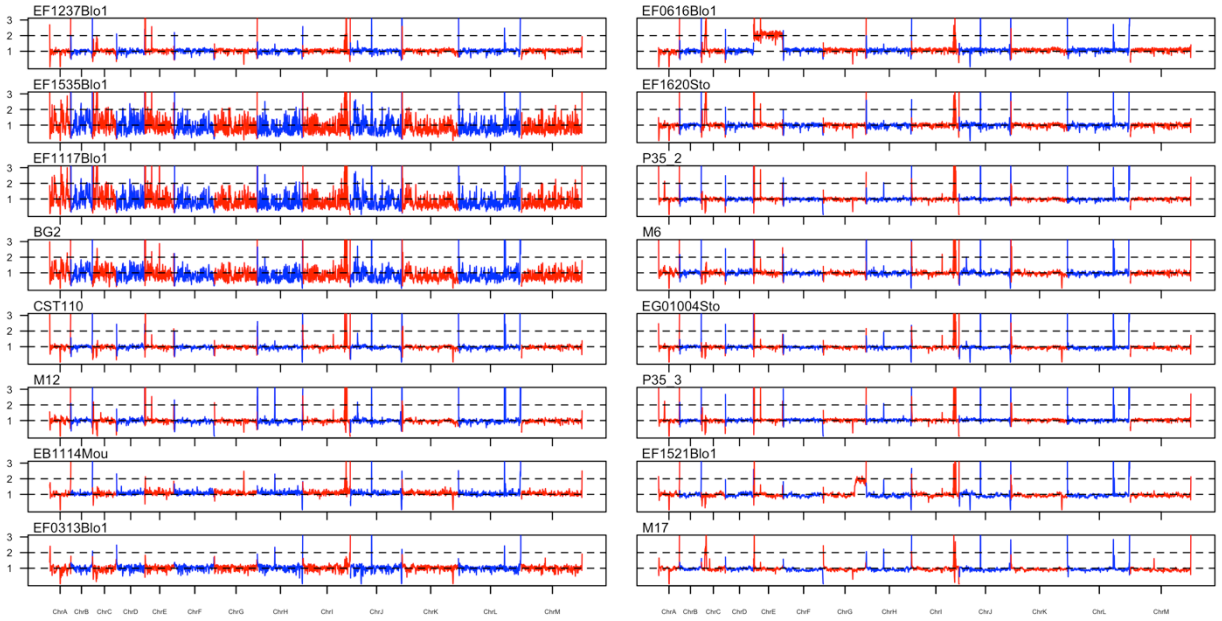
A. Coverage analysis of Biswas *et al.* 2017.



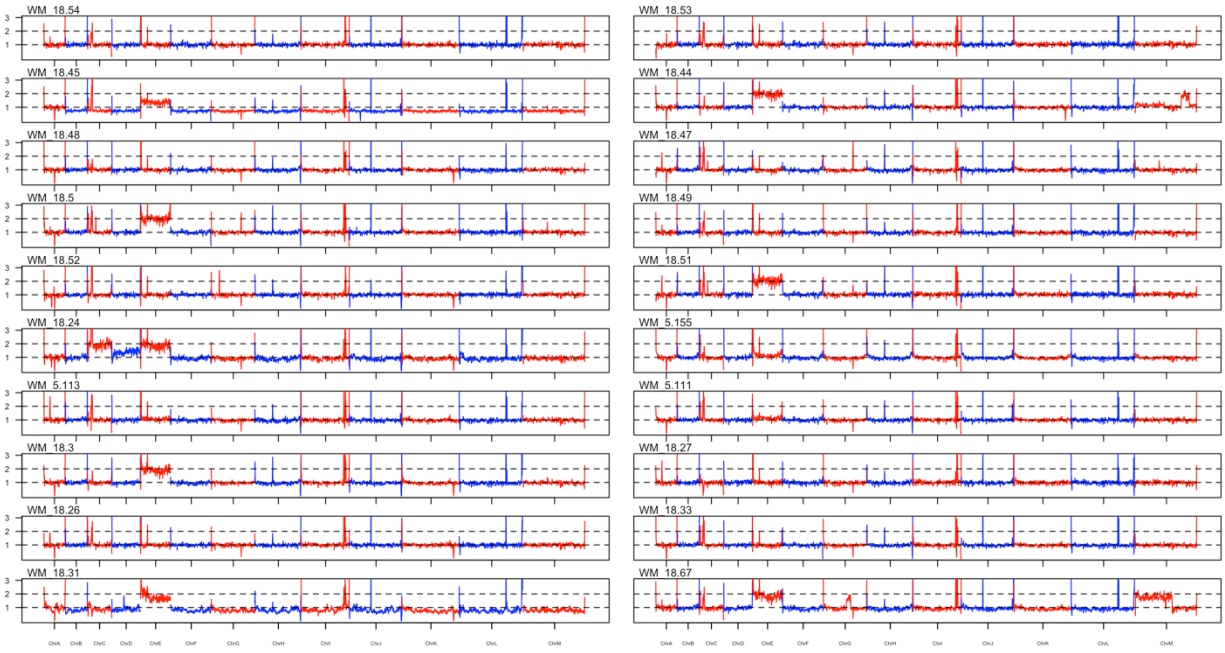
B. Coverage analysis of Carreté *et al.* 2018.



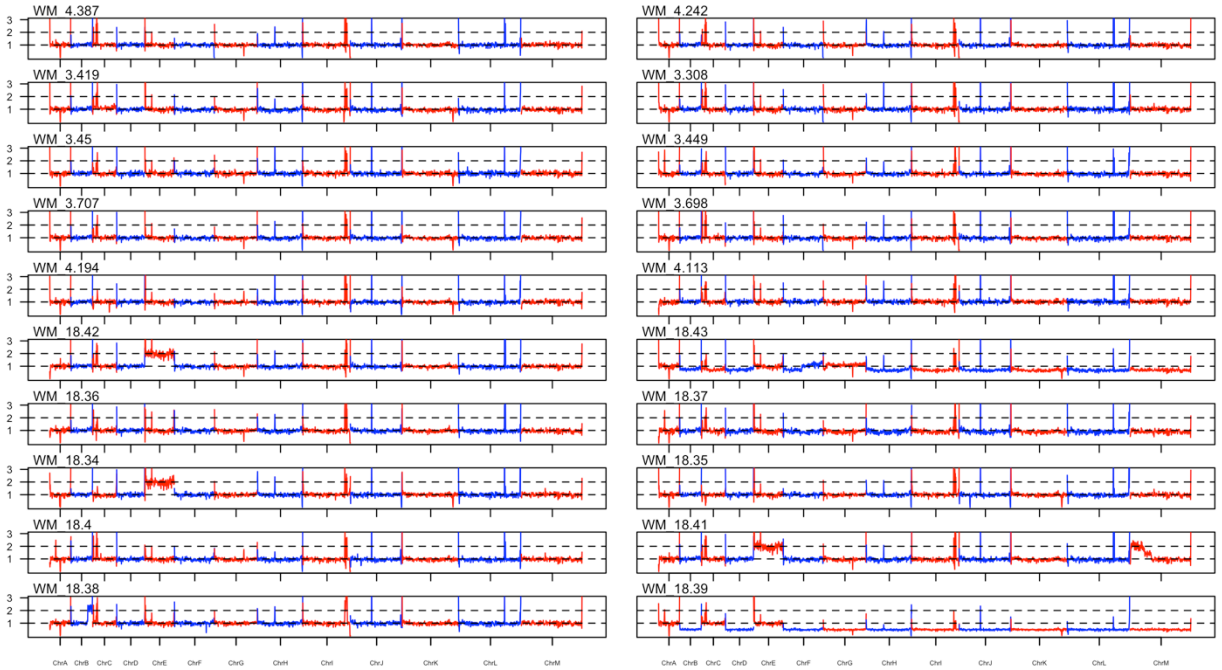
C. Coverage analysis of Carreté *et al.* 2018.



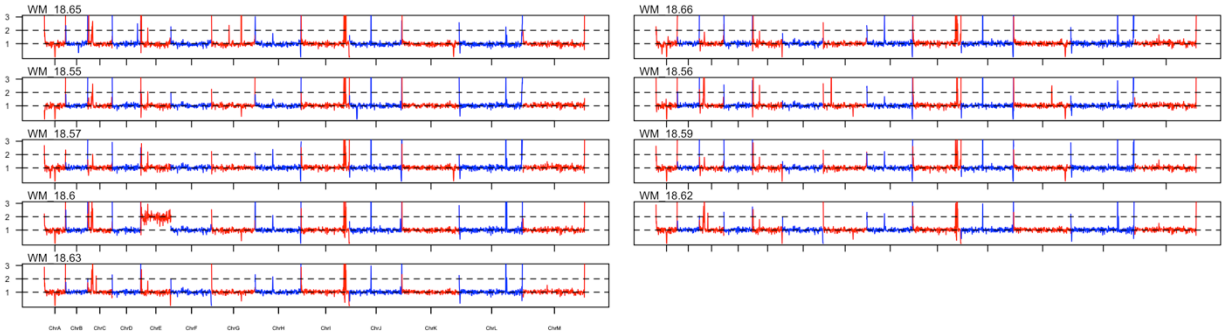
D. Coverage analysis of Biswas *et al.* 2018.



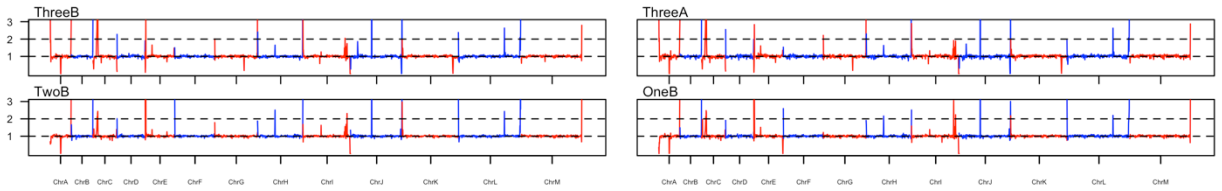
E. Coverage analysis of Biswas *et al.* 2018.



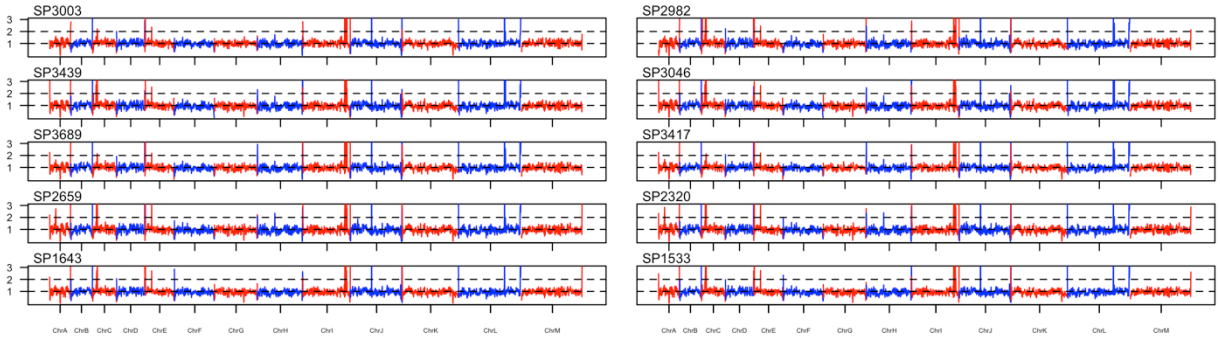
F. Coverage analysis of Biswas *et al.* 2018.



G. Coverage analysis of Håvelsrud *et al.* 2017.



H. Coverage analysis of Mctaggart *et al.* 2020.



I. Coverage analysis of Manitoba isolates.

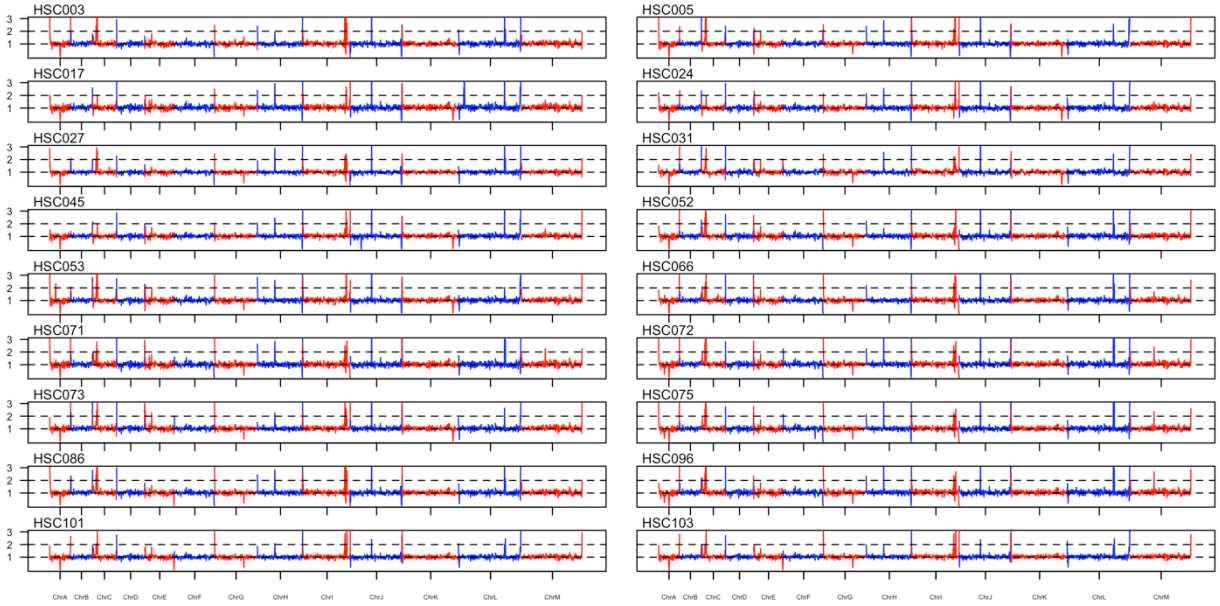


Figure S1. Coverage analysis of *Candida glabrata* isolates.

The coverage was identified using BEDtools genomecov and a custom R script that determined the average coverage over a 5 kb sliding window per chromosome.

(A) Biswas *et al.* 2017. (B/C) Carreté *et al.* 2018. (D/E/F) Biswas *et al.* 2018. (G) Håvelsrud *et al.* 2017. (H) Mctaggart *et al.* 2020. (I) Manitoba isolates.