Analysis of clinical *Candida parapsilosis* isolates reveals copy number variation in key fluconazole resistance genes

Sean Bergin[a], Laura A. Doorley[b], Jeffrey M. Rybak[b], Kenneth H. Wolfe[c], Geraldine Butler[a#], Christina A. Cuomo[d, e#], and P. David Rogers[b#]

[a]School of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Dublin, Ireland.

[b]Department of Pharmacy and Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN, USA.

[c]School of Medicine, Conway Institute, University College Dublin, Belfield, Dublin, Ireland.

[d]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[e]Molecular Microbiology and Immunology Department, Brown University, Providence, RI, USA.

[#]Joint corresponding authors

Running head: CNVs in fluconazole resistance genes

Abstract

We used whole-genome sequencing to analyse a collection of 35 fluconazole resistant and 7 susceptible *Candida parapsilosis* isolates together with coverage analysis and GWAS techniques to identify new mechanisms of fluconazole resistance. Phylogenetic analysis shows that although the collection is diverse, two probable outbreak groups were identified. We identified copy number variation of two genes, *ERG11* and *CDR1B*, in resistant isolates. Two strains have a CNV at the *ERG11* locus; the entire ORF is amplified in one, and only the promoter region is amplified in the other. We show the annotated telomeric gene *CDR1B* is actually an artefactual *in silico* fusion of two highly similar neighbouring *CDR* genes due to an assembly error in the *C. parapsilosis* CDC317 reference genome. We report highly variable copy numbers of the *CDR1B* region across the collection. Several strains have increased expansion of the two genes into a tandem array of new chimeric genes. Other strains have experienced a deletion between the two genes creating a single gene with a reciprocal chimerism. We find translocations, duplications, and gene conversion across the *CDR* gene family in the *C. parapsilosis* species complex, showing that it is a highly dynamic family.

Introduction

Candida parapsilosis is a human fungal pathogen that is globally one of the most common sources of non-*albicans Candida* infections (1, 2). In the decade 2006-2016, *C. parapsilosis* accounted for ~16% of all candidemia cases (3). Traditionally, *C. parapsilosis* was predominantly found in immunocompromised patients such as transplant recipients or preterm neonates (4, 5). More recently, however, cases have seen a rise in adult patients in non-surgical wards (6, 7). *C. parapsilosis*, and its sister species *Candida orthopsilosis* and *Candida metapsilosis*, belong to the CUG-Ser1 clade along with other major fungal pathogens *Candida albicans*, *Candida dubliniensis*, and *Candida tropicalis* (8). Unlike its sister species and other members of this clade, *C. parapsilosis* is assumed to be completely asexual due to its high homozygosity, pseudogenisation of *MAT***a**, and the lack of a *MAT*α idiomorph (9, 10).

Outbreaks of *C. parapsilosis* have been associated with variants conferring resistance to common antifungal drugs, including fluconazole, a triazole (11). Fluconazole binds to the enzyme lanosterol 14alpha-demethylase, encoded by the gene *ERG11*. This enzyme plays a key role in the ergosterol biosynthesis pathway, which is inhibited by the binding of fluconazole (12-14). Ergosterol is a key component of the fungal cell membrane and in its absence, and with accumulation of alternate sterols, cell growth is arrested (15, 16). Resistance to fluconazole treatment is a growing trend in clinical *Candida spp.* isolates (17). In *C. parapsilosis* resistance is particularly associated with the Y132F substitution in *ERG11* that contributes directly to resistance (18) and has been implicated in many fluconazole-resistant outbreak events across the world (19-24). Equivalent mutations also contribute to fluconazole resistance in *C. albicans*, *C. tropicalis,* and *Candida auris* (25-29).

Most of our understanding of other mechanisms of fluconazole resistance in *Candida* species, including the role of other substitutions in *ERG11*, comes from studies in *C. albicans* (13, 14, 26, 29). Overexpression of *ERG11*, often by gain-of-function mutations in the transcriptional regulator *UPC2*, has been implicated in resistance in *C. albicans* (30-32). In addition, azole resistance is due in part to overexpression of drug efflux pumps (33-35). In *C. albicans*, the contribution of two drug efflux pumps encoded by *CDR1* and *CDR2* that both belong to the ABC transporter (CDR) family to fluconazole resistance has been well studied (33). In the absence of drugs *CDR1* is expressed while *CDR2* is not (36). However, expression of both genes is upregulated in some resistant strains due to activating mutations in *TAC1*, a gene encoding a transcriptional regulator (18, 37, 38). Likewise, overexpression of *MDR1*, which encodes a transporter of the Major Facilitator Superfamily, is overexpressed in some isolates due to activating mutations in *MRR1*, encoding another transcriptional regulator (39). Overexpression of homologs of *CDR1* and *MDR1* have also been found to contribute to resistance in some *C. parapsilosis* clinical isolates that contain similar activating mutations in *TAC1* and *MRR1* (18, 40, 41). Often, multiple resistance mechanisms are found to act in concert in the same isolate leading to high level resistance (42, 43).

In this study we investigate the genetic mechanisms underlying fluconazole resistance in 42 *C. parapsilosis* isolates. Fluconazole resistance has previously been

studied in 34 of these isolates using targeted gene sequencing and gene expression analysis (18, 40, 41). Mutations in *ERG11* and over-expression of drug transporters were identified in some isolates. However, some isolates that share the same azole resistance-associated mutation exhibit a range of MIC values, and for other isolates, no obvious resistance mechanisms were identified. Here we use whole genome sequencing, coverage analysis and GWAS methods to identify point mutations and copy number variants (CNVs) associated with novel mechanisms of fluconazole resistance. Using phylogenomic methods, we also identify two probable outbreak clades, from Bloemfontein and Johannesburg, South Africa.

Results

Phylogeny and two outbreak clades

Azole resistance mechanisms have previously been studied in 34 fluconazole resistant (MIC ≥ 8 µg/ml) and three fluconazole sensitive (MIC ≤ 2 µg/ml) isolates of *C. parapsilosis* included in this study (18, 40, 41). To improve the power of the analysis (especially for GWAS), we sequenced all 37 genomes, and included one more resistant isolate CDC317 (the reference strain for *C. parapsilosis*) and four susceptible isolates 73/037, 73/114, FM16 and MSK809 (44). The isolates originate from several geographical locations, including several collected from two cities in South Africa between 2001 and 2009 (Johannesburg and Bloemfontein, Table 1). Phylogenomic analysis shows that the isolates represent a broad range of the *C. parapsilosis* phylogeny, as seen when integrated into a tree containing >200 other strains (Fig. S1) (44). Resistant isolates fall into each of the five global clades of *C. parapsilosis* that we have previously identified (44), and susceptible isolates belong to four out of five clades. Despite this breadth, two groups of isolates have very shallow branches, indicating that they have a very close relationship (Fig. 1). The clade marked by a single asterisk contains isolates all originating from the same clinic in Bloemfontein (Table 1). For the clade marked with double asterisks, one isolate comes from Ann Arbor, Michigan whereas the rest originate from a clinic in Johannesburg, collected over a period of eight years. These clades indicate putative outbreaks in the South African clinics.

For 34 of the fluconazole-resistant isolates, multiple potential resistance mechanisms were previously identified using gene expression analysis (RT-qPCR) and targeted gene sequencing (18, 40, 41) (Table 1). To improve the resolution we measured the expression of target genes using RNA-seq, by comparing to the expression of the genes in the azole susceptible isolate Cp13 (Table 1). The RNA-seq data supports the previous analysis (40, 41). For example, overexpression of drug transporter *CPAR2_603010* (*MDR1B*) in strains from Bloemfontein (log_2FC ranging from 1.8 to 5.3, Table 1), mediated at least in part by the A854V activating mutation in the regulator gene *CPAR2_807270* (*MRR1*), contributes directly to resistance to fluconazole (41). Some Bloemfontein strains have increased expression of *MDR1B* but do not have a corresponding *MRR1* mutation (e.g. Cp11, Table 1) (41). Strikingly, the isolate homozygous for A854V, Cp29, has much higher expression of both *MDR1B* (log_2FC = 5.28) and *CDR1B* (log_2FC = 2.18) compared to the other Bloemfontein strains. In addition, mutations in the ergosterol-biosynthesis gene *CPAR2_303740* (*ERG11*) and the *CDR*-family regulator *CPAR2_303510* (*TAC1*) were shown to contribute to fluconazole resistance in other isolates (40).

Fourteen strains (including the reference strain *C. parapsilosis* CDC317) harbour the *ERG11* Y132F substitution which is a well-documented resistance mutation (11, 17, 18, 25, 45). The Y132F substitution is heterozygous in CDC317 and in Cp4, and homozygous in the other 12. The isolates from Johannesburg all have the Y132F mutation (including the heterozygous Cp4), but the isolates from Bloemfontein do not (Table 1).

Eight of the 35 resistant strains do not have any mutations in *ERG11*, *TAC1*, or *MRR1* that have been experimentally determined to affect fluconazole resistance (Table 1). The R398I mutation in *ERG11* has been frequently observed occurring in tandem with Y132F (21, 23), but has also been found without Y132F in susceptible isolates (40). The substitutions A854V, R479K, and I283R in *MRR1* have all been identified as activating mutations leading to the upregulation of genes including *CDR1B,* and *MDR1B* (46). The *TAC1* G650E mutation has shown to increase fluconazole resistance and overexpress *CDR1* and *CDR1B* when introduced into a susceptible isolate (18). In addition, among this collection of isolates, strains that share the same mutation can differ 32-fold in their MIC values (e.g. isolates in the

Johannesburg clade have MICs varying from 8 to 256 μg/ml, Table 1). This suggests that novel resistance mechanisms remain to be identified, and that different mechanisms may have additive effects that have not been captured by these studies.

Copy number variation of *ERG11*

We found that aneuploidy is relatively common in the 42 isolates; 13 have either segmental or whole chromosome aneuploidies (Fig. S2), and several isolates have aneuploidies of multiple chromosomes. However, unlike in *C. albicans* where aneuploidy of the chromosome containing *ERG11* and *TAC1* has been associated with fluconazole resistance (47-49), none of the resistant strains in the collection have extra copies of chromosome 3. Several resistant strains have aneuploidy of chromosome 5, which contains the *ERG4* gene. The potential role of *ERG4* in fluconazole resistance in *Candida spp.* has not been well characterised, but the gene has been found overexpressed alongside *ERG11* in azole-resistant *C. albicans* (50).

We found that two resistant strains, Cp15 and Cp27, have small CNVs at the *ERG11* locus on chromosome 3 (Fig. 2). In Cp27, the entire coding sequence of both *ERG11* and the upstream gene (*HMS1*), and part of the downstream gene (*THR1*), has been amplified by the CNV (1,309,908-1,315,502 bp). Here, the locus has been amplified to nine copies. Short read mate-pair mapping supports the interpretation that this CNV is a tandem array of duplicated sequence. In Cp15, a 341 bp section of the *ERG11* promoter region is amplified to eight copies (1,312,556-1,312,896 bp). Cp15 and Cp27 are the only two isolates in this collection with increased expression of *ERG11* (Table 1). As far as we are aware, this is the first time that an amplification of an *ERG11* promoter has been observed in any *Candida* species.

Genome Wide Association Study

We performed a Genome Wide Association Study (GWAS) using all 42 isolates to identify potential variants associated with fluconazole resistance that had not been found in earlier studies. The GWAS was carried out using GEMMA (Genome-wide Efficient Mixed Model Association) (51), which calculates and incorporates relatedness data between isolates in order to minimise the confounding effect of

population structure on association scores. Because the MIC assays for fluconazole were measured in two different laboratories (Table 1), the phenotypic data was converted to a binary score of either resistant or susceptible to reduce possible bias, with MIC ≤ 2 µg/ml classed as susceptible and MIC ≥ 8 µg/ml classed as resistant, in line with CLSI (52) and EUCAST (53) guidelines. Genotypes were also converted into binary presence/absence of non-reference alleles, with both heterozygous and homozygous variants treated as present. In addition, only variants that were predicted to affect protein function by SIFT (54) were included in the analysis to narrow the search. In total, 7462 variants were used as input to GEMMA. The GWAS analysis did not identify any significant associations below the Bonferroni-corrected p-value threshold of 6.7×10^{-6} (Table S1). However, several of the variants with the lowest p-value scores were in *CPAR2_405290* (*CDR1*) and *CPAR2_304370* (*CDR1B*), members of the ABC family of putative drug transporters. Investigating the alignments leading to these calls showed that the *CPAR2_405290* variants are likely a result of mismapping from a similar gene in the genome, so we did not investigate *CPAR2_405290* further. Overexpression of *CPAR2_304370* (*CDR1B*) has previously been observed in fluconazole-resistant isolates and shown to directly contribute to this phenotype (40, 41, 46). In a study investigating acquired azole resistance in consecutive isolates taken from a patient undergoing fluconazole treatment, one isolate with reduced susceptibility to fluconazole had undergone amplification of the *CDR1B* locus, from 4 to ~15 copies (46). We therefore investigated the *CDR1B* locus in more depth.

The *CDR1B* locus contains two genes, *CDR1B.1* and *CDR1B.2*, and is amplified in resistant isolates

While further characterising the role of *CPAR2_304370* (the annotated *CDR1B* gene in the reference genome assembly) in azole resistance, we noticed an increased sequence coverage compared to genomic average in most of the isolates. We used the average coverage across the ORF to estimate the copy number of this gene in each of the isolates. Whereas a majority (25/42) of the isolates have a copy number in the range 4-6x, there are several outliers ranging up to 33x, and only three isolates have the expected value of 2x (Fig. 1). Several isolates that have increased expression of *CDR1B* but no corresponding *MRR1* gain-of-function mutation have

increased copy number of *CDR1B*, suggesting that amplification of this locus and activating mutations in upstream regulators can both drive overexpression of this gene. Strikingly two of the three isolates with two copies of the locus are susceptible to fluconazole, and no susceptible isolate has more than five copies.

Of special note is the clade containing *C. parapsilosis* FM16, CDC317, and Cp14 (Fig. 1). FM16 is susceptible to fluconazole and has only two copies of *CPAR2_304370*. CDC317 and Cp14 are both fluconazole resistant. CDC317 has only 5 copies of *CPAR2_304370*. However, CDC317 is heterozygous for a Y132F mutation in *ERG11* that is not present in Cp14 (Table 1). In contrast, Cp14 has 16 copies of *CPAR2_304370*. We propose that the related isolates Cp14 and CDC317 acquired resistance by differing mechanisms, the former by acquiring a mutation in *ERG11*, and the latter through increased copy number of *CPAR2_304370*.

The MIC values of the isolates from the Johannesburg outbreak range from 16-256 µg/ml (18) (Table 1). Two related isolates (Cp38 and Cp35) with MICs of 32 µg/ml have acquired a G650E substitution in *TAC1*, resulting in increased expression of *CDR1* (Table 1) (18). The combination of Y132F in *ERG11* and G650E in *TAC1* likely increases resistance compared to each variant alone (18). The copy number of *CDR1B* is highly variable in the Johannesburg isolates, ranging from 6-33x (Fig. 1). Whereas there is no direct correlation between the copy number of *CDR1B* and MIC in these isolates, it is notable that Cp37 has the highest MIC (256 µg/ml) and the highest number of *CDR1B* copies (~33).

Long-read (Oxford Nanopore) sequencing of CDC317 revealed that the *CPAR2_304370* gene annotated in the Sanger sequencing reference genome assembly of this strain was in fact erroneously assembled by fusing together two highly similar tandem genes, which we now call *CDR1B.1* and *CDR1B.2* (Fig. 3A). As a result, the intergenic space between these two genes, and parts of the genes themselves, are not present in the original reference assembly. It is likely that the presence of *CDR1B.1* and *CDR1B.2* is the ancestral (and most common) state of the locus in *C. parapsilosis* (Fig. 3A), and that the four copies of *CPAR2_304370* indicated in many of the isolates by coverage analysis relative to the reference genome assembly in fact represent diploids with two copies of both *CDR1B.1* and

CDR1B.2 (Fig. 1). We used the long sequencing reads, alongside short reads, to generate a highly accurate, complete chromosome assembly of CDC317. This assembly confirmed that CDC317 has two copies each of *CDR1B.1* and *CDR1B.2*, so the previously estimated copy number of 5x *CPAR2_304370* (Fig. 1) was likely inflated by short reads mismapping from related genes. Using the new accurate assembly of CDC317, we found that *CDR1B.1* and *CDR1B.2* are 98.69% identical at the nucleotide level and the intergenic regions upstream of both genes are 46.10% identical. The two genes differ by only 23 amino acids (out of 1498) when translated.

Significantly, short reads from isolate FM16 (estimated to have two copies of *CPAR2_304370*) map to the original *C. parapsilosis* CDC317 reference genome without an increase in coverage or misaligned read pairs (Fig. 3B). This is evidence of an array contraction of the two genes in FM16 that results in a single chimeric *CDR1B.2/CDR1B.1* gene, biologically mirroring the misassembly in the original reference. Short read alignments of Cp36 and Cp5 suggest that similar array contractions occurred in these isolates.

Long-read sequencing of Cp14 revealed that the extra copies predicted by coverage analysis (16x *CPAR2_304370*) are the result of a tandem array of identical chimeric *CDR1B.1/CDR1B.2* genes, bounded upstream by non-chimeric *CDR1B.2* and downstream by non-chimeric *CDR1B.1* (Fig. 3A). These chimeric genes inherited their 5' region from *CDR1B.1* and their 3' region from *CDR1B.2*. In this manner they are opposite to the chimeric gene in FM16. None of the Cp14 long sequencing reads reached across the entire tandem array, so the exact copy number of the chimeric genes could not be determined. However, by aligning the reads to the long-read assembly of CDC317, we found reads that contain the beginning, middle, and end of the tandem array (Fig. 3C).

Families of *CDR* orthologs and paralogs in the *C. parapsilosis* clade

CDR1B.1 and *CDR1B.2* are two of nine *CDR* genes in *C. parapsilosis* (Fig. 4). Strikingly, most (5/9) of these genes are located in telomeric regions. Many of the *CDR* genes in *C. parapsilosis*, including all five telomeric ones, have direct orthologs in *C. metapsilosis* and *C. orthopsilosis* but they are more distantly related to *C. albicans CDR* genes. The *CDR* orthologs in *C. metapsilosis* and *C. orthopsilosis*

share synteny of neighbouring genes when compared to the gene order of *C. parapsilosis*. The telomeric *CDR* genes are likely to have originated after *C. parapsilosis* diverged from the *C. albicans* lineage, but before the separation of *C. parapsilosis* from *C. orthopsilosis* and *C. metapsilosis*. In addition, a recent gene duplication in *C. parapsilosis* produced the gene pair *CPAR2_300010/CPAR2_603800* which are duplicated only in this species, while the *CDR1B.1/CDR1B.2* gene pair has a more complicated history.

C. metapsilosis has only a single gene (*CMET_1082*) at the *CDR1B* locus, which is equally related to the two *C. parapsilosis CDR1B* genes (Fig. 4; genes highlighted in pink). Long-read sequences of the reference *C. metapsilosis* strain BP57, recently assembled by Mixão et al. (52), indicate that this is not due to misassembly. There are two *CDR* genes present at the *CDR1B* locus in *C. orthopsilosis*, but interestingly only one of them (*CORT1E06580*) falls phylogenetically into the same clade as *C. parapsilosis CDR1B.1* and *CDR1B.2* and *C. metapsilosis CMET_1082*, while its neighbour (*CORT1E06570*) falls into an adjacent clade with *CMET_1071* and two *C. parapsilosis* genes *CPAR2_300010* and *CPAR2_603800* (Fig. 4). Notably, the *CDR1B* locus, *CPAR2_300010* and *CPAR2_603800* are all telomeric on different chromosomes of *C. parapsilosis*, so we suggest there may have been some *CDR* gene exchange and/or homogenization among telomeric regions within these species.

We did not observe CNVs affecting any of the other *CDR* genes in *C. parapsilosis*, nor in 36 *C. orthopsilosis* or 30 *C. metapsilosis* isolates that we analyzed. The *CDR1B* locus of *C. parapsilosis* is unique among the *CDR* genes of these three species in having two highly similar genes in tandem, which provides a template for amplification of the locus to occur readily.

Discussion

Using a genome-wide approach, we identified two CNVs (*ERG11* and *CDR1B.1*/*CDR1B.2*) that are associated with fluconazole resistance in *C. parapsilosis*. CNVs are a method of gene duplication by which an organism can transiently adapt to its environment (53, 54). Environmental changes, such as introduction of an antifungal drug, can select for specific genes to be duplicated, and

thereby overexpressed (49, 55). After the drug is removed, the CNV can be lost by selective pressure to maintain a compact genome size (55).

Overexpression of *ERG11* by increasing copy number has been observed in *C. albicans* (47, 48). However, in *C. albicans* the gene was amplified along with *TAC1* by means of a partial aneuploidy of chromosome 5 leading to a formation of an isochromosome, i(5L). The i(5L) isochromosome, which typically results in a single extra copy of the chromosomal region, has been identified in multiple clinical isolates where it has a modest but measurable impact on fluconazole resistance in different genetic backgrounds (47, 49, 56). Genomic expansion of *ERG11* in combination with hotspot mutations is also associated with azole resistance of *C. tropicalis* (57). In *C. auris*, a large survey of 304 isolates identified a CNV including *ERG11* in 18 isolates (most from a single clade) which was associated with fluconazole resistance (58). Recently, a laboratory-directed evolution experiment also showed that reduced azole susceptibility is associated with large segmental duplications containing *ERG11* in *C. auris*. One evolved strain had a 191 kb long CNV with 75 protein encoding genes including *ERG11* amplified, while another had a 161 kb long CNV containing 67 protein encoding genes including *ERG11* (59).

A very recent analysis identified amplifications of *ERG11* in 21 azole resistant isolates of *C. parapsilosis* (60). The amplifications ranged from partial aneuploidy of chromosome 3, similar to *C. albicans* (47, 48), to smaller amplifications of 2.3 to 12.1 kb (60). These are similar to the amplifications that we observe in strain Cp27, where *ERG11* and its neighboring gene are amplified (5.6 kb). The CNV in strain Cp15 is distinctly different; only the *ERG11* promoter region is amplified. Importantly, both Cp27 and Cp15 isolates also have increased *ERG11* expression (logFC ~2.4, Table 1), strongly suggesting these duplications directly impact expression of this gene and fluconazole resistance. We have previously observed a similar localised gene amplification in *C. parapsilosis* in the gene *RTA3* (44). Several different CNVs, spanning either the whole gene or just the promoter region, led to overexpression of *RTA3* associated with increased resistance to the antimicrobial drug miltefosine. Promoter amplifications may therefore be a previously underexplored mechanism of drug resistance in *Candida* species.

The GWAS analysis performed in this study failed to find significant associations. Although disappointing, it is within expectation because the sample size (n=42) is very low for a study of this kind, and there was a large imbalance between the number of resistant (n=35) and susceptible (n=7) isolates. PowerBacGWAS, a tool used to find required sample sizes for GWAS in bacteria, estimates in a best case scenario where minor allele frequency is high and effect size is large, analysis of 500 isolates would be required to identify a single significant SNP in *Mycobacterium tuberculosis* (61). This issue is further compounded in our study by the presence of groups of highly related strains from outbreak infections, which reduces the effective sample size. Despite the lack of statistical power, the GWAS results guided us towards characterising the *CDR1B* amplification as a possible mechanism of fluconazole resistance. This discovery lends support to the use of GWAS in *C. parapsilosis,* particularly in studying easily definable phenotypes such as antifungal resistance.

Our analysis of the *CDR* family in the *C. parapsilosis* species complex has led to some interesting insights. We found that most of the *CDR* genes in the *C. parapsilosis* species complex have arisen by gene duplication events after the lineage diverged from *C. albicans*. The *CDR* gene content varies between *C. parapsilosis* and its two sister species. Many *CDR* copies are located at telomeres and there is evidence of exchange of duplicated genes between telomeres. The *CDR1B* locus is particularly variable between the three sister species; there is a single gene in *C. metapsilosis*, two distinct genes in *C. orthopsilosis*, and two highly similar genes in *C. parapsilosis*. It is possible that *C. orthopsilosis* represents the ancestral state, with one of the genes lost in the *C. metapsilosis* lineage and one gene overwritten in *C. parapsilosis* by gene conversion from its neighbour.

Previous analysis identified several potential azole resistance mechanisms in some of the strains described here (18, 40, 41). This includes the Y132F mutation in *ERG11* (Table 1) (18). We find that this variant is present in all the closely related isolates from Johannesburg, supporting our inference that these represent an infection outbreak. However, the range of MICs observed in these isolates cannot fully be explained by the presence of Y132F (Table 1). Two isolates (Cp35 and Cp38) collected in 2008, late in the inferred outbreak, have also acquired variants in

Tac1, with an associated rise in expression of *CDR1* (Table 1, logFC 0.4-0.7) (18). Our analysis suggests that in other isolates (e.g. Cp37, also collected in 2008) increased resistance is associated with, and may be caused by, amplification of the *CDR1B* locus (up to ~33 copies).

The presence of *ERG11* Y132F is increasingly associated with infection outbreaks (19-23, 25), suggesting that it may confer a transmission advantage beyond the effect on azole resistance. However, outbreaks can also be caused by strains without Y132F such as the Bloemfontein isolates in this study. For example, in a study investigating 60 *C. parapsilosis* strains involved in a large outbreak in a Brazilian ICU, only ~36% of isolates resistant to fluconazole had an *ERG11* mutation (62). Another outbreak among patients undergoing allo-hematopoietic cell transplant treatment was associated with isolates without the Y132F mutation (63).

All of strains from the Bloemfontein outbreak have increased expression of *MDR1B* (log_2FC 1.8-5.28, Table 1). Although the isolates are closely related, there is some variability in their *MRR1* alleles. Some isolates contain an A854V activating mutation in *MRR1*, which is known to result in overexpression of *MDR1B* (41). Seven are heterozygous for the A854V mutation in *MRR1*, one is homozygous for the mutation, and six do not have the mutation. The earliest cultured strains (Cp1 and Cp2 isolated in 2001), are both heterozygous for the mutation whereas some strains isolated later (e.g. Cp17 in 2003) are lacking the mutation entirely (Table 1). This indicates that there may be sub-populations of related strains existing concurrently in the clinic which are variable at *MRR1*. Multiple strains without Mrr1 A854V have increased expression of *MDR1B* (Table 1), showing that there may be additional unidentified factors contributing to *MDR1B* expression in this clade.

Eight of the resistant isolates in Table 1 contain none of the common variants associated with azole resistance. Our work identifies relevant CNVs in 4 of these. *ERG11* is amplified in two strains, Cp27 and Cp15, that are not associated with outbreaks. Strains Cp14 and Cp6, which were also not associated with outbreaks, both have >10 copies of *CDR1B* and no other identified resistance mechanism. The variation in copy number of *CDR1B* across the isolates suggests that *CDR1B* amplification may be a common mechanism of azole resistance in *C. parapsilosis*. It

is important to note that, of the multiple *CDR* genes in multiple *Candida* species, we observed CNVs at only the *C. parapsilosis CDR1B* locus. We propose that the existence of two nearly identical genes in tandem makes CNVs at this locus more likely to occur. In this way, *C. parapsilosis* may be primed to generate extra copies of *CDR1B*, and therefore be predisposed to develop fluconazole resistance. Amplification likely occurs during infection as described by Branco et al (46), where a strain of *C. parapsilosis* in a patient treated with fluconazole acquired a CNV amplifying *CDR1B* that was associated with reduced fluconazole susceptibility.

Methods

RNA Sequencing

C. parapsilosis isolates were maintained at -80°C in 40% glycerol stocks. Isolates were grown in YPD liquid media overnight and plated onto Sabouraud-Dextrose (BD companies) agarose plates in biological triplicate for 24h growth at 30°C. Sterile loops were used to inoculate 20mL RPMI with MOPS and 2% glucose to OD600=0.1. Cultures were incubated at 35°C with110rpm shaking for 8h, after which the cells were centrifuged at 4000rpm for 5 min. Supernatants were removed and pellets were stored at -80°C for a minimum of 24hrs. RNA isolation was performed via RiboPureTM Yeast (Invitrogen) kits per manufacturer's instructions. RNA Sequencing performed using Illumina NextSeq for stranded mRNA (Hartwell Center, St Jude Children's Research Hospital). Libraries were prepared with paired-end adapters using Illumina chemistries per manufacturer's instructions, with read lengths of approximately 150bp with at least 50 million raw reads per sample (Bioproject 14022043). RNA-sequencing was analyzed using CLC Genomics Workbench version 20.0 (QIAGEN), and reads were trimmed using default settings for failed reads and adaptor sequences and then subsequently mapped to the *C. parapsilosis* genome (GenBank accession: GCA_000182765.2) with paired reads counted as one and expression values set to RPKM. Principal-component analysis was utilized for initial assessment of biological replicate clusters. Whole transcriptome differential gene expression analysis was performed with the prescribed algorithm of CLC Genomics Workbench version 20.0. Mismatch, insertion, and deletion costs were set to default parameters and a Wald test was used to compare all isolates against the fluconazole susceptible isolate Cp13. Fold

changes for CPAR2_304370, CPAR2_405290, CPAR2_301760 were identified for all isolates and are reported in Table 1.

Whole Genome Sequencing

Genomic DNA was isolated from overnight yeast peptone dextrose (YPD) liquid media cultures utilizing a Triton SDS and phenol-chloroform method previously described by Amberg et al (64). DNA concentrations were quantified using both the Qubit Fluorometer and Nanodrop spectrophotometer using the manufacturers' protocols. Whole genome libraries were prepared and sequenced on the NovaSeq600 platform (150 bp, paired-end reads) by the University of Maryland School of Medicine Institute for Genomic Sciences.

Fluconazole Susceptibility Testing

Inoculums of YPD liquid media were prepared from original stocks, stored at −80°C in 40% glycerol. Inoculates were grown at 30°C with 220 rpm shaking overnight and subsequently plated onto Sabouraud dextrose agar for overnight incubation at 35°C. Minimum inhibitory concentrations (MIC) were determined in RPMI1640 (Roswell Park Memorial Institute) supplemented with MOPS (3-(N-morpholino) propanesulfonic acid) buffer and 2% glucose, pH 7.0 liquid media following CLSI M27-A4 methods for broth microdilution. Fluconazole (Sigma Aldrich) drug stocks were prepared in dimethyl sulfoxide (DMSO) at 100× the maximum plate concentration (256 mg/mL for resistant isolates and 16 mg/mL for susceptible isolates). MICs were determined visually as the concentration achieving 50% growth inhibition at 24 hours, and the modal value of biological triplicate measurements was considered the MIC reporting.

MinION sequencing

Samples were grown overnight in 50 mL of YPD broth at 30°C. Genomic DNA was extracted from 1.5 ml of liquid cultures saturated to 10 A600 units per millilitre using the Yeast Masterpure DNA purification kit (MPY80010) following manufacturer's instructions.

Genomic DNA (1 µg) from each sample sequenced with MinION technology using the native barcoding kits (SQK-NBD-24 and SQK-NBD114-24) from Oxford Nanopore Technologies (ONT), following manufacturer's instructions. Library kit SQK-NBD112-24 was used for sequencing CDC317 on an R9.4.1 chemistry flowcell (FLO-MIN106D). Library kit SQK-NBD114-24 was used for sequencing Cp14 on an R10.4.1 chemistry flowcell (FLO-MIN114). Sequencing of both strains was performed on a MinION MK1C device with MinKNOW (ONT) versions 21.11.6 and 22.10.7 for CDC317 and Cp14 respectively. Both runs were set to the default fast configurations. Basecalling and demultiplexing was run within MinKNOW during sequencing. This generated 638,855 and 237,020 raw reads respectively for CDC317 and Cp14.

Sequence analysis

The Illumina reads for all 42 *C. parapsilosis* samples were downsampled to ~100x coverage using the Picard version 2.21.2 DownsampleSam on unmapped SAM files. These files were converted to FASTQ format using Picard SamToFastq and aligned to the *C. parapsilosis* CDC317 reference genome using bwa mem version 0.7.17 (65). GATK version 4.1.4.1 was used to mark duplicate reads and reorder the mapped BAM files with the tools MarkDuplicates and ReorderSam respectively (66). Variants were called on individual samples using GATK HaplotypeCaller with the "-ERC GVCF" tag. The GVCF outputs were combined into a multi-sample VCF with GATK CombineGVCFs, and then genotyped using GATK GenotypeGVCFs. SNP variants were hard filtered using GATK VariantFiltration with following parameters: QD < 2.0, QUAL < 30.0, SOR > 3.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0 and DP < 10. Indels were similarly filtered using parameters: QD < 2.0, QUAL < 30.0, FS > 200.0, and ReadPosRankSum < -20.0. All variants were filtered using GATK SelectVariants to remove multi-allelic sites, and sites that contained >10% 'no-call' genotypes. The two sets of variants were combined for GWAS analysis, and a file containing only SNPs was used for phylogenetic analysis. MinION reads for isolate CDC317 were filtered using NanoFilt to exclude reads with quality < 12 and length < 10000 (67). The filtered reads were assembled using Canu version 2.2 (68). Errors in this assembly were removed by incorporating Illumina read data using NextPolish version 1.4.0 (69). The assembly is

available under accession PRJNA1031570. Reads for *C. orthopsilosis* were obtained from Schröder et al. (70), Pryszcz et al. (71), Bergin et al. (44) and Zhai et al. (63), and reads for *C. metapsilosis* from Zhai et al. (63) and O'Brien et al. (72) (Table S2).

Phylogeny

A FASTA alignment of all sites containing a SNP in at least one isolate was created from the multi-sample VCF file using a custom script (https://github.com/CMOTsean/HetSiteRando). Heterozygous variants were randomly assigned to either allele on a per-site basis. A SNP tree was constructed with the alignment file using RAxML version v8.2.12 with the GTRGAMMA model of nucleotide substitution and 1000 bootstrap replicants (73).

GWAS

A binary phenotype matrix where all samples were scored as either resistant or susceptible was used as input to the GWAS. Variants entering the GWAS were filtered to only those likely to affect protein function, as annotated by SIFT4G (74). PLINK version 1.90b6.21 was used to reformat input data into BED, BIM, and FAM files for the GWAS (75). GEMMA version 0.98.5 was used to create a relatedness matrix between all strains, and then to conduct the GWAS itself using parameters '-hwe 0.0 -maf 0.0' (51).

Estimating repeat copy number and structure

To calculate the estimated copy number of the *CPAR2_304370* locus, the average coverage across the ORF in each isolate was found using BEDTools coverage version 2.29.2 and then divided by half the modal genome coverage (BEDTools genomecov) for that isolate (76).

To confirm copy number estimates and investigate the structure of the repeat, the nucleotide sequence of *CDR1B.1* was searched against MinION reads from strains CDC317 and Cp14 using BLASTN. Additionally, the Cp14 reads were aligned to the MinION CDC317 assembly using GraphMap version 0.3.0 (77).

C. parapsilosis species complex comparison

To create the synteny maps between *C. parapsilosis* and its sister species, the *C. parapsilosis* CDC317 reference genome was aligned against the *C. metapsilosis* BP57 reference genome (GCA_017655625.1) (52) and the *C. orthopsilosis* SY36 long-read assembly (PRJNA767198) using BLASTN. Hits from each query chromosome were assigned a colour and then plotted.

To construct the tree of CDR protein sequences, the sequences for *C. albicans*, *C. parapsilosis*, and *C. metapsilosis* were taken from CGOB with the exception of *CMET_1535* which came from the *C. metapsilosis* BP57 assembly.

Data availability

DNA sequence assembly and raw data is available under accession PRJNA1031570 and RNA sequencing is available at BioProject 14022043.

Acknowledgements

This work was supported by Science Foundation Ireland (grant numbers 19/FFP/6668 and 18/CRT/6214 to G.B. and 20/FFP-A/8795 to K.H.W.) and by NIAID grant U19AI110818 to the Broad Institute (C.A.C.). This work was also supported by NIH NIAID grant R01 A1058145 awarded to P.D.R. and in part by the National Cancer Institute of the National Institutes of Health under Award Number P30 CA021765 awarded to the Hartwell Center at St. Jude Children's Research Hospital. We thank Daniel J. Diekema at the University of Iowa for generously providing the clinical isolates used in this study. We are grateful to Adam Ryan, Eoin Ó Cinnéide, Athaliah Fubara and Conor Hession in UCD for help with MinION sequencing of Cp14 and CDC317, and Evelyn Zuniga for some MIC assays.

References

1. Hajjeh RA, Sofair AN, Harrison LH, Lyon GM, Arthington-Skaggs BA, Mirza SA, Phelan M, Morgan J, Lee-Yang W, Ciblak MA, Benjamin LE, Sanza LT, Huie S, Yeo SF, Brandt ME, Warnock DW. 2004. Incidence of bloodstream infections due to *Candida* species and i*n vitro* susceptibilities of isolates collected from 1998 to 2000 in a population-based active surveillance program. J Clin Microbiol 42:1519-1527.

2. Tóth R, Nosek J, Mora-Montes HM, Gabaldon T, Bliss JM, Nosanchuk JD, Turner SA, Butler G, Vágvölgyi C, Gácser A. 2019. *Candida parapsilosis*: from genes to the bedside. Clin Microbiol Rev 32.

3. Pfaller MA, Diekema DJ, Turnidge JD, Castanheira M, Jones RN. 2019. Twenty years of the SENTRY antifungal surveillance program: results for *Candida* species From 1997–2016. Open Forum Infect Dis 6:S79-S94.

4. Rolling T, Zhai B, Gjonbalaj M, Tosini N, Yasuma-Mitobe K, Fontana E, Amoretti LA, Wright RJ, Ponce DM, Perales MA. 2021. Haematopoietic cell transplantation outcomes are linked to intestinal mycobiota dynamics and an expansion of *Candida parapsilosis* complex species. Nature Microbiol 6:1505-1515.

5. Pammi M, Holland L, Butler G, Gacser A, Bliss JM. 2013. *Candida parapsilosis* is a significant neonatal pathogen: a systematic review and meta-analysis. Pediatr Infect Dis J 32:e206.

6. Arastehfar A, Unal N, Hosbul T, Alper Ozarslan M, Sultan Karakoyun A, Polat F, Fuentes D, Gumral R, Turunc T, Daneshnia F, Perlin DS, Lass-Florl C, Gabaldon T, Ilkit M, Nguyen MH. 2022. Candidemia among Coronavirus disease 2019 patients in Turkey admitted to Intensive Care Units: a retrospective multicenter study. Open Forum Infect Dis 9:ofac078.

7. Brescini L, Mazzanti S, Morroni G, Pallotta F, Masucci A, Orsetti E, Montalti R, Barchiesi F. 2022. Candidemia in internal medicine: Facing the new challenge. Mycopathologia 187:181-188.

8. Fitzpatrick DA, Logue ME, Stajich JE, Butler G. 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol 6:1-15.

9. Logue ME, Wong S, Wolfe KH, Butler G. 2005. A genome sequence survey shows that the pathogenic yeast *Candida parapsilosis* has a defective *MTL* **a***1* allele at its mating type locus. Eukaryot Cell 4:1009-1017.

10. Sai S, Holland LM, McGee CF, Lynch DB, Butler G. 2011. Evolution of mating within the *Candida parapsilosis* species group. Eukaryot Cell 10:578-587.

11. Berkow EL, Lockhart SR. 2017. Fluconazole resistance in *Candida species*: a current perspective. Infect Drug Resist 10:237-245.

12. Bard M, Lees ND, Turi T, Craft D, Cofrin L, Barbuch R, Koegel C, Loper JC. 1993. Sterol synthesis and viability of *erg11* (cytochrome P450 lanosterol demethylase) mutations in *Saccharomyces cerevisiae* and *Candida albicans*. Lipids 28:963-967.

13. Xiang M-J, Liu J-Y, Ni P-H, Wang S, Shi C, Wei B, Ni Y-X, Ge H-L. 2013. Erg11 mutations associated with azole resistance in clinical isolates of *Candida albicans*. FEMS Yeast Res 13:386-393.

14. Xu Y, Chen L, Li C. 2008. Susceptibility of clinical isolates of *Candida* species to fluconazole and detection of *Candida albicans ERG11* mutations. J Antimicrob Chemother 61:798-804.

15. Jordá T, Puig S. 2020. Regulation of ergosterol biosynthesis in *Saccharomyces cerevisiae*. Genes 11:795.

16. Rodrigues ML. 2018. The multifunctional fungal ergosterol. MBio 9:10.1128/mbio. 01755-18.

17. Escribano P, Guinea J. 2022. Fluconazole-resistant *Candida parapsilosis*: A new emerging threat in the fungi arena. Front Fungal Biol 3:1010782.

18. Doorley LA, Barker KS, Zhang Q, Rybak JM, D PR. 2023. Mutations in *TAC1* and *ERG11* are major drivers of triazole antifungal resistance in clinical isolates of *Candida parapsilosis*. Clin Microbiol Infect doi:10.1016/j.cmi.2023.08.030.

19. Arastehfar A, Daneshnia F, Hilmioğlu-Polat S, Fang W, Yaşar M, Polat F, Metin DY, Rigole P, Coenye T, Ilkit M. 2020. First report of candidemia clonal outbreak caused by emerging fluconazole-resistant *Candida*

parapsilosis isolates harboring Y132F and/or Y132F+ K143R in Turkey. Antimicrob Agents Chemother 64:10.1128/aac. 01001-20.

20. Corzo-Leon DE, Peacock M, Rodriguez-Zulueta P, Salazar-Tamayo GJ, MacCallum DM. 2021. General hospital outbreak of invasive candidiasis due to azole-resistant *Candida parapsilosis* associated with an Erg11 Y132F mutation. Med Mycol 59:664-671.

21. Díaz-García J, Gómez A, Alcalá L, Reigadas E, Sánchez-Carrillo C, Pérez-Ayala A, Gómez-García de la Pedrosa E, González-Romo F, Merino-Amador P, Cuétara MS. 2022. Evidence of fluconazole-resistant *Candida parapsilosis* genotypes spreading across hospitals located in Madrid, Spain and harboring the Y132F ERG11p substitution. Antimicrob Agents Chemother 66:e00710-22.

22. Presente S, Bonnal C, Normand A-C, Gaudonnet Y, Fekkar A, Timsit J-F, Kernéis S. 2023. Hospital clonal outbreak of fluconazole-resistant *Candida parapsilosis* harboring the Y132F ERG11p substitution in a French Intensive Care Unit. Antimicrob Agents Chemother 67:e01130-22.

23. Thomaz DY, de Almeida Jr JN, Sejas ON, Del Negro GM, Carvalho GO, Gimenes VM, de Souza MEB, Arastehfar A, Camargo CH, Motta AL. 2021. Environmental clonal spread of azole-resistant *Candida parapsilosis* with Erg11-Y132F mutation causing a large candidemia outbreak in a Brazilian cancer referral center. J Fungi 7:259.

24. Martini C, Torelli R, De Groot T, De Carolis E, Morandotti GA, De Angelis G, Posteraro B, Meis JF, Sanguinetti M. 2020. Prevalence and clonal distribution of azole-resistant *Candida parapsilosis* isolates causing bloodstream infections in a large Italian hospital. Front Cell Infect Microbiol 10:232.

25. Castanheira M, Deshpande LM, Messer SA, Rhomberg PR, Pfaller MA. 2020. Analysis of global antifungal surveillance results reveals predominance of Erg11 Y132F alteration among azole-resistant *Candida parapsilosis* and *Candida tropicalis* and country-specific isolate dissemination. Int J Antimicrob Agents 55:105799.

26. Flowers SA, Colón B, Whaley SG, Schuler MA, Rogers PD. 2015. Contribution of clinically derived mutations in *ERG11* to azole resistance in *Candida albicans*. Antimicrob Agents Chemother 59:450-460.

27. Healey KR, Kordalewska M, Jiménez Ortigosa C, Singh A, Berrío I, Chowdhary A, Perlin DS. 2018. Limited *ERG11* mutations identified in isolates of C*andida auris* directly contribute to reduced azole susceptibility. Antimicrob Agents Chemother 62:10.1128/aac. 01427-18.

28. Rybak JM, Sharma C, Doorley LA, Barker KS, Palmer GE, Rogers PD. 2021. Delineation of the direct contribution of *Candida auris ERG11* mutations to clinical triazole resistance. Microbiol Spect 9:e01585-21.

29. Xu Y, Sheng F, Zhao J, Chen L, Li C. 2015. *ERG11* mutations and expression of resistance genes in fluconazole-resistant *Candida albicans* isolates. Arch Microbiol 197:1087-1093.

30. Flowers SA, Barker KS, Berkow EL, Toner G, Chadwick SG, Gygax SE, Morschhäuser J, Rogers PD. 2012. Gain-of-function mutations in *UPC2* are a frequent cause of *ERG11* upregulation in azole-resistant clinical isolates of *Candida albicans*. Eukaryot Cell 11:1289-1299.

31. Heilmann CJ, Schneider S, Barker KS, Rogers PD, Morschhäuser J. 2010. An A643T mutation in the transcription factor Upc2p causes constitutive *ERG11* upregulation and increased fluconazole resistance in *Candida albicans*. Antimicrob Agents Chemother 54:353-359.

32. MacPherson S, Akache B, Weber S, De Deken X, Raymond M, Turcotte B. 2005. *Candida albicans* zinc cluster protein Upc2p confers resistance to antifungal drugs and is an activator of ergosterol biosynthetic genes. Antimicrob Agents Chemother 49:1745-1752.

33. Chen L, Xu Y, Zhou C, Zhao J, Li C, Wang R. 2010. Overexpression of *CDR1* and *CDR2* genes plays an important role in fluconazole resistance in *Candida albicans* with G487T and T916C mutations. J Int Med Res 38:536-545.

34. Marr KA, Lyons CN, Rustad T, Bowden RA, White TC. 1998. Rapid, transient fluconazole resistance in *Candida albican*s is associated with increased mRNA levels of *CDR*. Antimicrob Agents Chemother 42:2584-2589.

35. Prasad R, Banerjee A, Khandelwal NK, Dhamgaye S. 2015. The ABCs of *Candida albicans* Multidrug Transporter Cdr1. Eukaryot Cell 14:1154-1164.

36. Schillig R, Morschhäuser J. 2013. Analysis of a fungus-specific transcription factor family, the *Candida albicans* zinc cluster proteins, by artificial activation. Mol Microbiol 89:1003-1017.

37. Coste AT, Karababa M, Ischer F, Bille J, Sanglard D. 2004. *TAC1*, transcriptional activator of *CDR* genes, is a new transcription factor involved in the regulation of *Candida albicans* ABC transporters *CDR1* and *CDR2*. Eukaryotic Cell 3:1639-1652.

38. Lohberger A, Coste AT, Sanglard D. 2014. Distinct roles of *Candida albicans* drug resistance transcription factors *TAC1*, *MRR1*, and *UPC2* in virulence. Eukaryot Cell 13:127-142.

39. Morschhäuser J, Barker KS, Liu TT, Blaß-Warmuth J, Homayouni R, Rogers PD. 2007. The transcription factor Mrr1p controls expression of the MDR1 efflux pump and mediates multidrug resistance in *Candida albicans*. PLoS Pathog 3:e164.

40. Berkow EL, Manigaba K, Parker JE, Barker KS, Kelly SL, Rogers PD. 2015. Multidrug transporters and alterations in sterol biosynthesis contribute to azole antifungal resistance in *Candida parapsilosis*. Antimicrob Agents Chemother 59:5942-50.

41. Doorley LA, Rybak JM, Berkow EL, Zhang Q, Morschhäuser J, Rogers PD. 2022. *Candida parapsilosis* Mdr1B and Cdr1B are drivers of Mrr1-mediated clinical fluconazole resistance. Antimicrob Agents Chemother 66:e00289-22.

42. Rybak JM, Muñoz JF, Barker KS, Parker JE, Esquivel BD, Berkow EL, Lockhart SR, Gade L, Palmer GE, White TC. 2020. Mutations in *TAC1B*: a novel genetic determinant of clinical fluconazole resistance in *Candida auris*. MBio 11:10.1128/mbio. 00365-20.

43. Souza ACR, Fuchs BB, Pinhati HM, Siqueira RA, Hagen F, Meis JF, Mylonakis E, Colombo AL. 2015. *Candida parapsilosi*s resistance to fluconazole: molecular mechanisms and *in vivo* impact in infected *Galleria mellonella* larvae. Antimicrob Agents Chemother 59:6581-6587.

44. Bergin SA, Zhao F, Ryan AP, Müller CA, Nieduszynski CA, Zhai B, Rolling T, Hohl TM, Morio F, Scully J. 2022. Systematic analysis of copy number variations in the pathogenic yeast *Candida parapsilosis* identifies a gene amplification in *RTA3* that is associated with drug resistance. Mbio 13:e01777-22.

45. Grossman NT, Pham CD, Cleveland AA, Lockhart SR. 2015. Molecular mechanisms of fluconazole resistance in *Candida parapsilosis* isolates from a U.S. surveillance system. Antimicrob Agents Chemother 59:1030-1037.

46. Branco J, Ryan AP, Pinto ESA, Butler G, Miranda IM, Rodrigues AG. 2022. Clinical azole cross-resistance in *Candida parapsilosis* is related to a novel *MRR1* gain-of-function mutation. Clin Microbiol Infect 28:1655 e5-1655 e8.

47. Coste A, Selmecki A, Forche A, Diogo D, Bougnoux M-E, d'Enfert C, Berman J, Sanglard D. 2007. Genotypic evolution of azole resistance mechanisms in sequential *Candida albicans* isolates. Eukaryot Cell 6:1889-1904.

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

49. Selmecki A, Gerami-Nejad M, Paulson C, Forche A, Berman J. 2008. An isochromosome confers drug resistance *in vivo* by amplification of two genes, *ERG11* and *TAC1*. Mol Microbiol 68:624-641.

50. Feng W, Yang J, Xi Z, Qiao Z, Lv Y, Wang Y, Ma Y, Wang Y, Cen W. 2017. Mutations and/or overexpressions of *ERG4* and *ERG11* genes in clinical azoles-resistant isolates of *Candida albicans*. Microb Drug Resist 23:563-570.

51. Zhou X, Stephens M. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 11:407-409.

52. Mixão V, Nunez-Rodriguez JC, del Olmo V, Ksiezopolska E, Saus E, Boekhout T, Gacser A, Gabaldón T. 2023. Evolution of loss of heterozygosity patterns in hybrid genomes of *Candida* yeast pathogens. BMC Biol 21:105.

53. Jeffares DC, Jolly C, Hoti M, Speed D, Shaw L, Rallis C, Balloux F, Dessimoz C, Bähler J, Sedlazeck FJ. 2017. Transient structural variations

have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun 8:14061.

54. Mishra S, Whetstine JR. 2016. Different facets of copy number changes: permanent, transient, and adaptive. Mol Cell Biol 36:1050-1063.

55. Todd RT, Selmecki A. 2020. Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. eLife 9:NA.

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

57. Hu T, Wang S, Bing J, Zheng Q, Du H, Li C, Guan Z, Bai FY, Nobile CJ, Chu H, Huang G. 2023. Hotspot mutations and genomic expansion of *ERG11* are major mechanisms of azole resistance in environmental and human commensal isolates of *Candida tropicalis*. Int J Antimicrob Agents 62:107010.

58. Chow NA, Muñoz JF, Gade L, Berkow EL, Li X, Welsh RM, Forsberg K, Lockhart SR, Adam R, Alanio A. 2020. Tracing the evolutionary history and global expansion of *Candida auris* using population genomic analyses. MBio 11:10.1128/mbio. 03364-19.

59. Carolus H, Pierson S, Muñoz JF, Subotić A, Cruz RB, Cuomo CA, Van Dijck P. 2021. Genome-wide analysis of experimentally evolved *Candida auris* reveals multiple novel mechanisms of multidrug resistance. MBio 12:10.1128/mbio. 03333-20.

60. Ning Y, Dai R, Zhang L, Xu Y, Xiao M. 2023. Copy number variants of *ERG11*: mechanism of azole resistance in *Candida parapsilosis*. Lancet Microbe doi:10.1016/S2666-5247(23)00294-X.

61. Coll F, Gouliouris T, Bruchmann S, Phelan J, Raven KE, Clark TG, Parkhill J, Peacock SJ. 2022. PowerBacGWAS: a computational pipeline to perform power calculations for bacterial genome-wide association studies. Commun Biol 5:266.

62. Daneshnia F, de Almeida Júnior JN, Arastehfar A, Lombardi L, Shor E, Moreno L, Verena Mendes A, Goreth Barberino M, Thomaz Yamamoto D, Butler G. 2022. Determinants of fluconazole resistance and echinocandin tolerance in *C. parapsilosis* isolates causing a large clonal candidemia

outbreak among COVID-19 patients in a Brazilian ICU. Emerg Microbes & Infect 11:2264-2274.

63. Zhai B, Ola M, Rolling T, Tosini NL, Joshowitz S, Littmann ER, Amoretti LA, Fontana E, Wright RJ, Miranda E. 2020. High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. Nat Med 26:59-64.

64. Amberg DC, Burke DJ, Strathern JN. 2006. Preparation of genomic DNA from yeast using glass beads. Cold Spring Harb Protoc 2006:pdb. prot4151-pdb. prot4151.

65. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997.

66. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297-1303.

67. De Coster W, D'hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34:2666-2669.

68. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722-736.

69. Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics 36:2253-2255.

70. Schroder MS, Martinez de San Vicente K, Prandini TH, Hammel S, Higgins DG, Bagagli E, Wolfe KH, Butler G. 2016. Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. PLoS Genet 12:e1006404.

71. Pryszcz LP, Nemeth T, Gacser A, Gabaldon T. 2014. Genome comparison of *Candida orthopsilosis* clinical strains reveals the existence of hybrids between two distinct subspecies. Genome Biol Evol 6:1069-78.

72. O'Brien CE, Zhai B, Ola M, Bergin SA, E OC, O'Connor I, Rolling T, Miranda E, Babady NE, Hohl TM, Butler G. 2022. Identification of a novel *Candida metapsilosis* isolate reveals multiple hybridization events. G3 12.

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

74. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. 2016. SIFT missense predictions for genomes. Nat Protoc 11:1-9.

75. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559-575.

76. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841-842.

77. Sović I, Šikić M, Wilm A, Fenlon SN, Chen S, Nagarajan N. 2016. Fast and sensitive mapping of nanopore sequencing reads with GraphMap. Nat Commun 7:11307.

78. Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178-192.

79. Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL. 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459:657-662.

80. Ola M, O'Brien CE, Coughlan AY, Ma Q, Donovan PD, Wolfe KH, Butler G. 2020. Polymorphic centromere locations in the pathogenic yeast *Candida parapsilosis*. Genome Res 30:684-696.

81. Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221-224.

Legends

Figure 1.

Left: Maximum-likelihood tree of 42 *Candida parapsilosis* isolates. RAxML (73) was used to construct the tree from an alignment of 15,582 SNPs genome-wide, using the GTRGAMMA model of nucleotide substitution. Branches with bootstrap values < 100 after 1000 iterations of bootstrap sampling are coloured according to value,

ranging from red (0) to green (99). Clades are marked with asterisks to denote potential outbreaks in Bloemfontein (*) and Johannesburg (**). *Right*: Bar chart showing estimated copy number of the *CDR1B* locus (*CPAR2_304370* in the reference genome assembly). Fluconazole-susceptible isolates are denoted with a blue bar, resistant isolates with a green bar. Copy number was estimated by taking mean coverage across the *CPAR2_304370* ORF and dividing by half the modal coverage for the isolate.

Figure 2.

Amplification of *ERG11* in fluconazole resistant isolates *C. parapsilosis* Cp15 and Cp27. Screenshot from IGV showing coverage (read depth) tracks (78). Coverage value is marked on the right for the peak of the amplification, and for the region average. Estimated copy numbers are in parentheses. Thick bars at bottom show CDS of genes, and thinner bars represent transcribed regions.

Figure 3.

A. Diagram showing the likely ancestral state of the *CDR1B* (*CPAR2_304370*) locus, with two highly similar genes *CDR1B.1* and *CDR1B.2* in tandem as occurs in the corrected genome sequence of *C. parapsilosis* CDC317 (based on MinION sequencing). The dashed grey arrow indicates how mis-assembly of this locus in the original reference genome sequence for CDC317 (79) led to an erroneous fusion between *CDR1B.1* and *CDR1B.2*, resulting in annotation of the incorrect gene structure *CPAR2_304370*. The solid grey arrows show CNVs formed by array expansion and or contraction in other isolates, such as Cp14 and FM16.
B. Coverage tracks in IGV of short-read data from three isolates aligned to the original *C. parapsilosis* CDC317 assembly. Both CDC317 and Cp14 have increased coverage at the *CPAR2_304370* locus, whereas FM16 does not.
C. MinION reads of isolate Cp14 aligned to a long-read CDC317 assembly show evidence of a tandem array of chimeric genes bounded by parental genes. The bottom track shows the annotated long-read assembly of CDC317, containing both *CDR1B.1* and *CDR1B.2*. Each grey bar is a single read aligned to the *CDR1B* locus where grey denotes read sequence matching the

reference, and coloured dashes are mismatched positions, from modified screenshots of IGV (70). Reads aligned to this locus belong to one of three groups: (i) reads that align fully to *CDR1B.1* and left flanking DNA while partially aligning to *CDR1B.2*, (ii) reads that align partially to both *CDR1B.1* and *CDR1B.2*, and (iii) reads that align fully to *CDR1B.2* and right flanking DNA while partially aligning to *CDR1B.1*. A schematic showing the genic content for each read group is shown above for clarity. Red boxes highlight regions where the read sequence does not match the reference sequence. Reads from groups (i) and (iii) contain one parent gene and one or more copies of the tandem chimera. Reads from group (ii) contain multiple copies of the tandem chimera but neither of the parent genes.

Figure 4.

Left: Schematic of synteny shared between members of the *C. parapsilosis* species complex. The *C. parapsilosis* genome was aligned to the genomes of *C. metapsilosis* (GCA_017655625.1) (52) and *C. orthopsilosis* (PRJNA767198) using BLASTN. These genomes were then coloured according to the *C. parapsilosis* chromosome aligned to each region. Previously identified centromeres are marked with a white circle (80). The locations of *CDR* genes are shown for each species. Right: Phylogenetic tree of CDR protein sequences from *C. parapsilosis, C. orthopsilosis*, *C. metapsilosis*, and *C. albicans*. Pink backgrounds denote genes most closely related to *CDR1B.1* and *CDR1B.2*. Three clades of genes located near telomeres are marked with boxes. Protein sequences were aligned using the Clustal Omega method in Seaview (81). The tree was constructed with the LG model within Seaview, using 100 bootstrap replicates. Bootstrap values >85 have been omitted for clarity.

Figure S1.

Maximum-likelihood tree of 207 *Candida parapsilosis* isolates constructed as in Fig. 1. Additional isolates, and clade designations were taken from (44). Susceptible isolates in this paper are marked with a blue circle, and resistant isolates are marked with a green circle. Coloured branches indicate bootstrap values after 1000 iterations of bootstrap sampling for that branch ranging from 0 (red) to 100 (green).

Figure S2.

Genome-wide coverage tracks for aneuploid strains. Coverage was calculated as the log2 mean coverage of 1kb windows and plotted. Dashed red line shows log2 value = 1, i.e. typical 2x diploid coverage

Table S1. GEMMA (GWAS) analysis

Table S2. Source of sequencing data for isolates.

Table 1 Expression and variant analysis of *C. parapsilosis* isolates

Isolate	Origin	Date	[a]MIC	Expression relative to Cp13 (log_2FC)[c]					*CDR1B* Copy Number	[a]Variants		
				ERG11	*MDR1*	*MDR1B*	*CDR1*	*CDR1B*		*ERG11*	*MRR1*	*TAC1*
Cp37	Johannesburg	07/2008	256	0.53	0.89	0.25	-0.45	1.84	33	**Y132F, R398I**		L978W
Cp38	Johannesburg	08/2008	32	0.11	0.6	-0.31	0.73	2.03	6	**Y132F, R398I**		**G650E**
Cp35	Johannesburg	01/2008	32	0.38	0.78	0	0.37	1.48	6	**Y132F, R398I**		**G650E**
Cp34	Johannesburg	03/2008	32	-0.07	0.48	0.91	-0.29	1.65	27	**Y132F, R398I**		
Cp26	Johannesburg	05/2004	32	0.12	0.46	-0.25	-0.23	1.15	11	**Y132F, R398I**		
Cp22	Johannesburg	07/2003	32	0.25	0.6	0.08	-0.36	1.26	19	**Y132F, R398I**		
Cp31	Johannesburg	07/2006	32	0.27	0.35	0.16	-0.17	1.23	11	**Y132F, R398I**		
Cp24	Johannesburg	09/2003	16	-0.1	0.2	0.05	-0.07	0.83	13	**Y132F, R398I**		
Cp25	Johannesburg	09/2003	16	0.12	0.51	0.03	-0.41	1.24	12	**Y132F, R398I**		
Cp4	Johannesburg	05/2001	16	0.82	1.01	1.47	-0.15	1.65	18	Y132F, **R398I**		
Cp7	Johannesburg	05/2002	16	0.63	1.44	2.58	0.09	1.75	13	**Y132F, R398I**		
Cp21	Ann Arbor	05/2003	8	0.35	0.69	0.3	-0.43	1.77	22	**Y132F, R398I**		L978W
Cp28	Bloemfontein	02/2005	64	0.34	-0.21	3.63	0.2	0.48	5		A854V	
Cp29	Bloemfontein	02/2005	64	0.11	2.53	5.28	-0.22	2.18	5		**A854V**	
Cp20	Bloemfontein	04/2003	16	0.22	-0.04	3.94	0.59	0.94	6		A854V	
Cp11	Bloemfontein	03/2002	16	-0.19	-0.02	4.56	-0.08	-0.17	5			
Cp2	Bloemfontein	05/2001	16	0.48	1.41	4.41	0.36	1.25	5		A854V	
Cp1	Bloemfontein	03/2001	16	0.57	0.21	3.65	0.23	0.98	5		A854V	
Cp18	Bloemfontein	04/2003	16	0.5	-0.18	2.31	0.4	0.07	6			
Cp17	Bloemfontein	01/2003	16	0.14	-0.07	1.76	-0.6	-0.01	5			
Cp8	Bloemfontein	01/2002	16	0.72	1.46	4.13	0.25	0.39	5			
Cp10	Bloemfontein	03/2002	16	0.45	0.45	4.55	0.26	1.03	5		A854V	
Cp12	Bloemfontein	07/2002	16	-0.68	0.9	2.28	-0.4	0.15	4		P255L, A854V	
Cp9	Bloemfontein	03/2002	16	-0.29	0.45	2.18	-0.89	-0.6	5		A854V	
Cp16	Bloemfontein	01/2003	8	0.37	-0.29	3.44	0.39	-0.12	5			

Cp19	Bloemfontein	04/2003	8	-0.72	-0.93	1.8	-0.22	-0.69	5			
Cp30	Bratislava	09/2005	128	-0.18	5.14	7.38	-0.67	3.39	4	**R398I**	**R479K**	
Cp32	Caracas	06/2007	128	0.92	0.48	-0.69	0.09	-0.08	4	**R398I**		I221T
Cp36	Detroit	04/2008	64	-0.69	4.09	6.71	-0.53	3.11	2		**I283R**	
Cp39	Bratislava	07/2009	32	0.75	0.33	3.26	-0.43	1	4	**Y132F**	G294E	**R208G**
Cp27	Hershey	11/2004	32	2.45	0.25	2.7	-0.06	2.01	4		K129fs, G982R	**R208G**
Cp14	Helsinki	06/2002	16	0.08	-0.47	0.28	0.42	2.09	16			
Cp15	Quito	01/2003	16	2.38	-0.26	0.58	-0.01	0.17	9	**F145L**		**N900D**
Cp6	Turino	05/2001	16	0.93	0.45	0.8	-0.92	0.94	10			
Cp5	New York	N/A	0.5	0.42	0.54	0.7	-1.26	-0.57	2		**K177N**, Q1053*	
Cp3	Kuala Lumpur	04/2001	0.25	0.8	0.25	0.26	-0.14	0.82	5	**R398I**		
Cp13	New York	N/A	0.25	0	0	0	0	0	4	**R398I**, S216L		
CDC317	Mississippi	N/A	64[b]	-	-	-	-	-	5	Y132F		
FM16	Nantes	N/A	2[b]	-	-	-	-	-	2			
MSK809	New York	N/A	2[b]	-	-	-	-	-	4	**R398I**		
73-037	Leeds	N/A	2[b]	-	-	-	-	-	4			**R208G**
73-114	Leeds	N/A	2[b]	-	-	-	-	-	4	**R398I**	D615G	**L877P**

[a]Fluconazole MIC values (24 h) and variants in *ERG11, MMR1* and *TAC1* were described previously (18, 41). Underlined mutations have been experimentally determined to increase fluconazole resistance. Homozygous variants are shown in bold. [b]MIC values were measured in a separate assay. [c] Expression values show $\log_2$ fold-changes relative to Cp13 from RNA-Seq. - expression data not available.

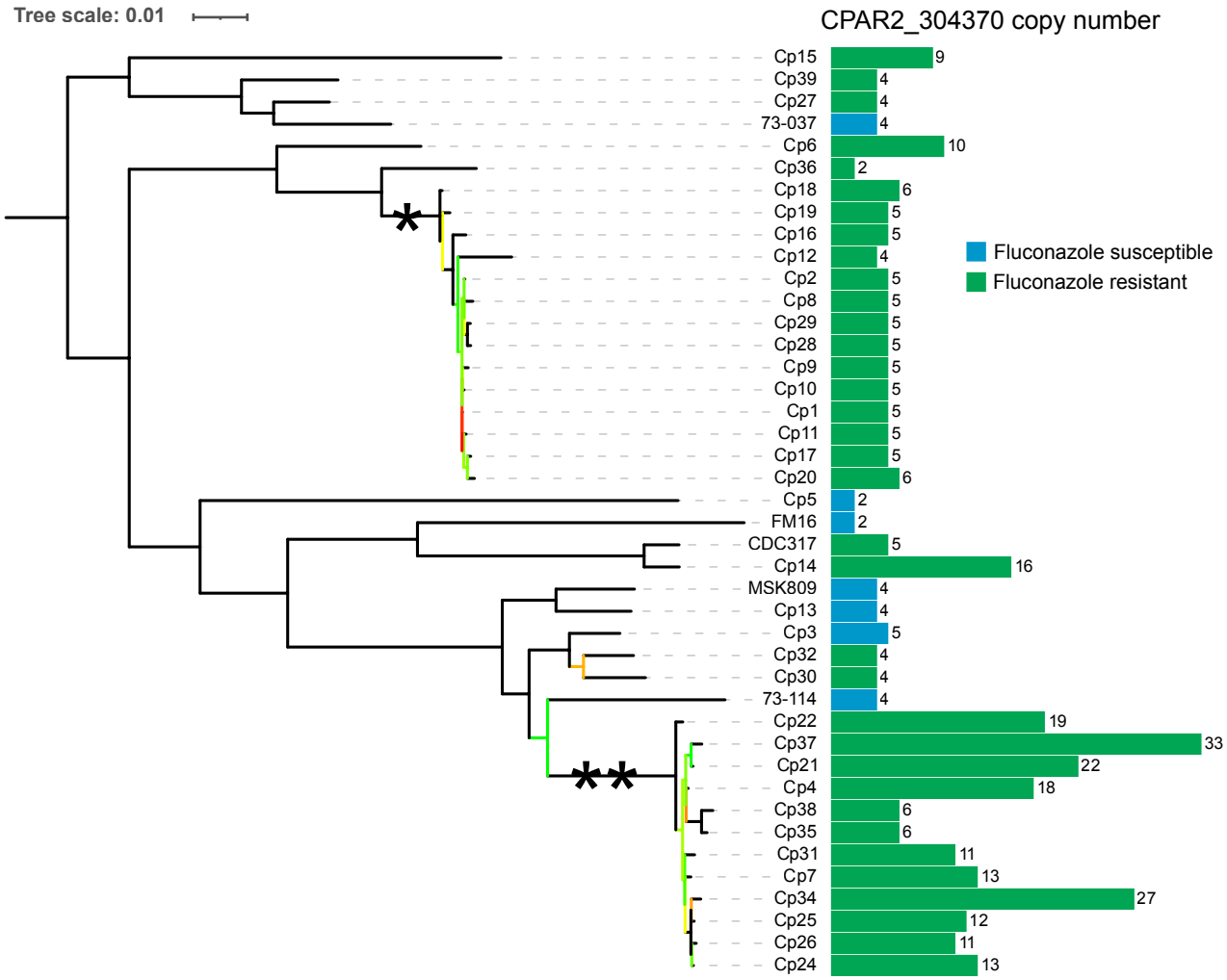

Figure 1.
Left: Maximum-likelihood tree of 42 *Candida parapsilosis* isolates. RAxML (71) was used to construct the tree from an alignment of 15,582 SNPs genome-wide, using the GTRGAMMA model of nucleotide substitution. Branches with bootstrap values < 100 after 1000 iterations of bootstrap sampling are coloured according to value, ranging from red (0) to green (99). Clades are marked with asterisks to denote potential outbreaks in Bloemfontein (*) and Johannesburg (**). Right: Bar chart showing estimated copy number of the *CDR1B* locus (*CPAR2_304370* in the reference genome assembly). Fluconazole-susceptible isolates are denoted with a blue bar, resistant isolates with a green bar. Copy number was estimated by taking mean coverage across the *CPAR2_304370* ORF and dividing by half the modal coverage for the isolate.

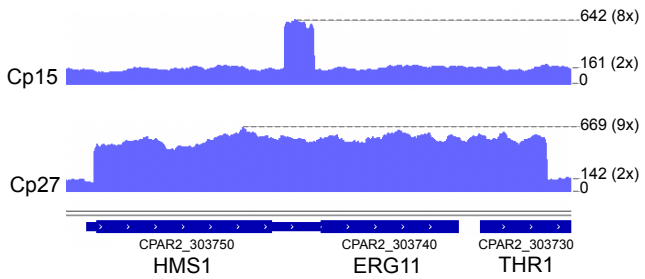

Figure 2.
Amplification of *ERG11* in fluconazole resistant isolates *C. parapsilosis* Cp15 and Cp27. Screenshot from IGV showing coverage (read depth) tracks (75). Coverage value is marked on the right for the peak of the amplification, and for the region average. Estimated copy numbers are in parentheses. Thick bars at bottom show CDS of genes, and thinner bars represent transcribed regions.

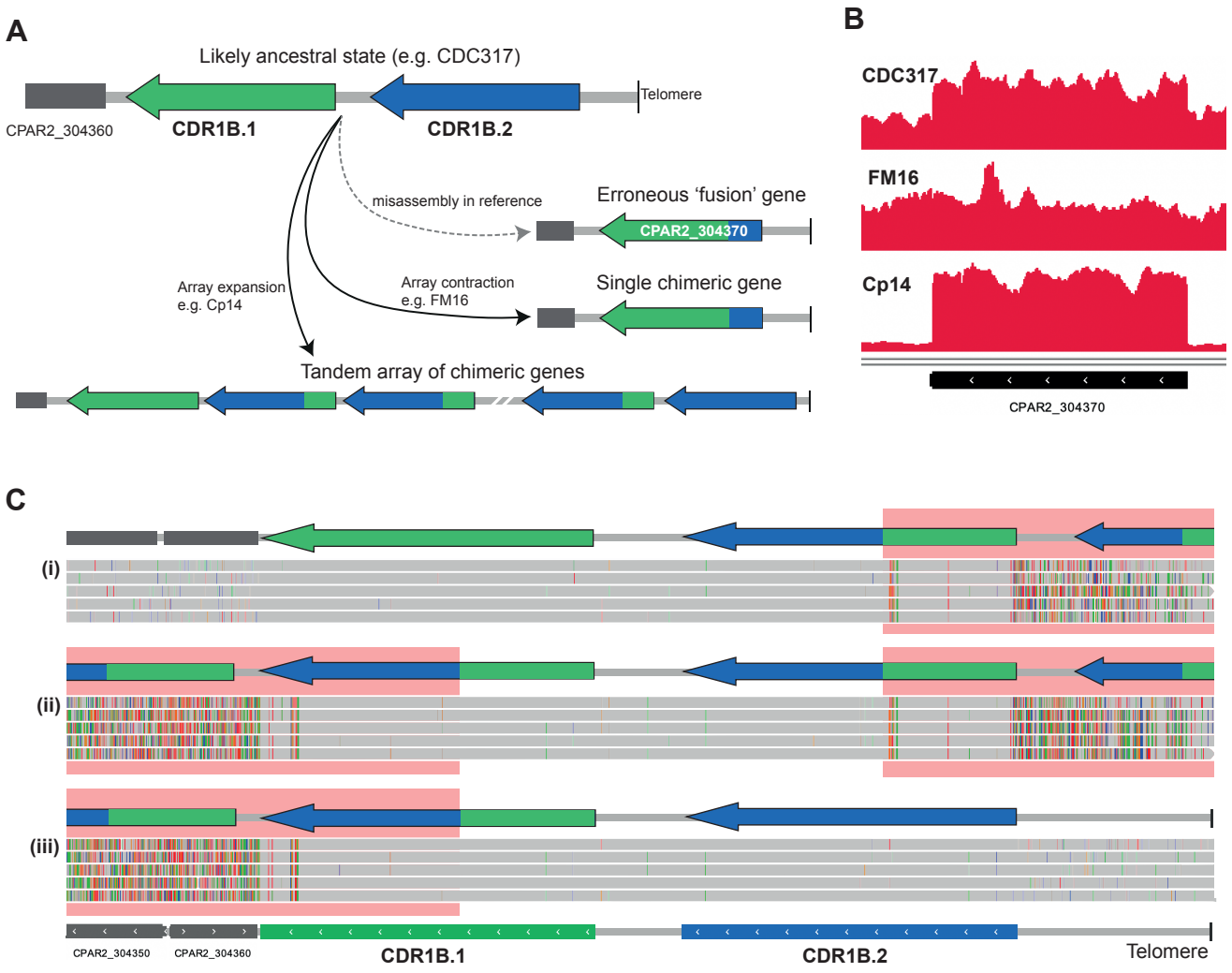

Figure 3.

A. Diagram showing the likely ancestral state of the *CDR1B* (*CPAR2_304370*) locus, with two highly similar genes *CDR1B.1* and *CDR1B.2* in tandem as occurs in the corrected genome sequence of C. parapsilosis CDC317 (based on MinION sequencing). The dashed grey arrow indicates how mis-assembly of this locus in the original reference genome sequence for CDC317 (76) led to an erroneous fusion between *CDR1B.1* and *CDR1B.2*, resulting in annotation of the incorrect gene structure *CPAR2_304370*. The solid grey arrows show CNVs formed by array expansion and or contraction in other isolates, such as Cp14 and FM16.

B. Coverage tracks in IGV of short-read data from three isolates aligned to the original *C. parapsilosis* CDC317 assembly. Both CDC317 and Cp14 have increased coverage at the *CPAR2_304370* locus, whereas FM16 does not.

C. MinION reads of isolate Cp14 aligned to a long-read CDC317 assembly show evidence of a tandem array of chimeric genes bounded by parental genes. The bottom track shows the annotated long-read assembly of CDC317, containing both *CDR1B.1* and *CDR1B.2*. Each grey bar is a single read aligned to the *CDR1B* locus where grey denotes read sequence matching the reference, and coloured dashes are mismatched positions, from modified screenshots of IGV (70). Reads aligned to this locus belong to one of three groups: (i) reads that align fully to *CDR1B.1* and left flanking DNA while partially aligning to CDR1B.2, (ii) reads that align partially to both *CDR1B.1* and *CDR1B.2*, and (iii) reads that align fully to CDR1B.2 and right flanking DNA while partially aligning to *CDR1B.1*. A schematic showing the genic content for each read group is shown above for clarity. Red boxes highlight regions where the read sequence does not match the reference sequence. Reads from groups (i) and (iii) contain one parent gene and one or more copies of the tandem chimera. Reads from group (ii) contain multiple copies of the tandem chimera but neither of the parent genes.

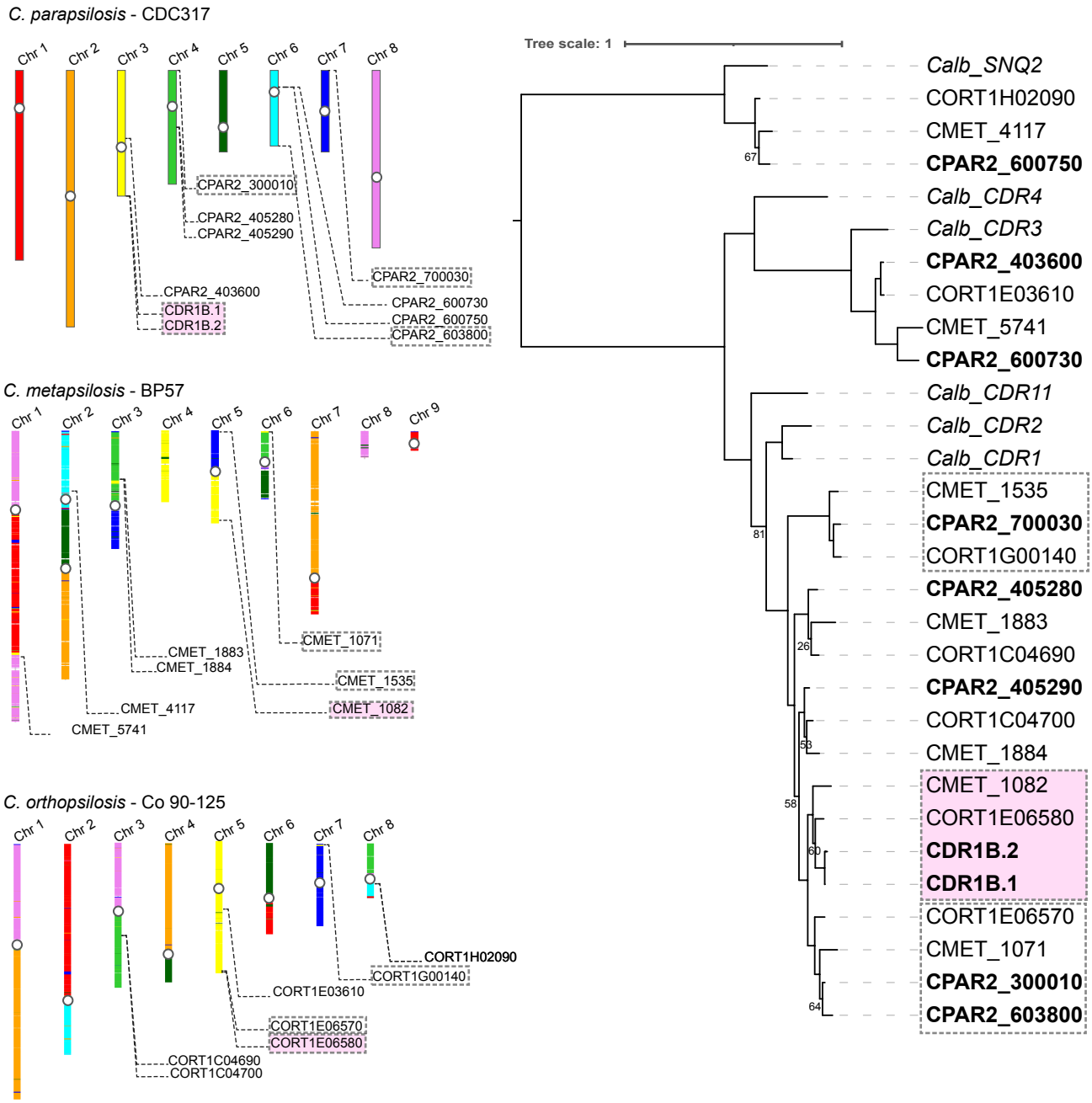

Figure 4.

Left: Schematic of synteny shared between members of the *C. parapsilosis* species complex. The *C. parapsilosis* genome was aligned to the genomes of *C. metapsilosis* (GCA_017655625.1) (55) and *C. orthopsilosis* (PRJNA767198) using BLASTN. These genomes were then coloured according to the *C. parapsilosis* chromosome aligned to each region. Previously identified centromeres are marked with a white circle (77). The locations of CDR genes are shown for each species.

Right: Phylogenetic tree of CDR protein sequences from *C. parapsilosis*, *C. orthopsilosis*, *C. metapsilosis*, and *C. albicans*. Pink backgrounds denote genes most closely related to *CDR1B.1* and *CDR1B.2*. Three clades of genes located near telomeres are marked with boxes. Protein sequences were aligned using the Clustal Omega method in Seaview (78). The tree was constructed with the LG model within Seaview, using 100 bootstrap replicates. Bootstrap values >85 have been omitted for clarity.

Supplementary Material

Table S1: Excel file

Table S2 Sequencing data for isolates

Isolate name	City	Country	Sequence data source	Reference
Candida parapsilosis				
Cp1	Bloemfontein	South Africa	This study	(40)
Cp10	Bloemfontein	South Africa	This study	(40)
Cp11	Bloemfontein	South Africa	This study	(40)
Cp12	Bloemfontein	South Africa	This study	(40)
Cp13	New York	United States	This study	(40)
Cp14	Helsinki	Finland	This study	(40)
Cp15	Quito	Ecuador	This study	(40)
Cp16	Bloemfontein	South Africa	This study	(40)
Cp17	Bloemfontein	South Africa	This study	(40)
Cp18	Bloemfontein	South Africa	This study	(40)
Cp19	Bloemfontein	South Africa	This study	(40)
Cp2	Bloemfontein	South Africa	This study	(40)
Cp20	Bloemfontein	South Africa	This study	(40)
Cp21	Ann Arbor	United States	This study	(40)
Cp22	Johannesburg	South Africa	This study	(40)
Cp24	Johannesburg	South Africa	This study	(40)
Cp25	Johannesburg	South Africa	This study	(40)
Cp26	Johannesburg	South Africa	This study	(40)
Cp27	Hershey	United States	This study	(40)
Cp28	Bloemfontein	South Africa	This study	(40)
Cp29	Bloemfontein	South Africa	This study	(40)
Cp3	Kuala Lumpur	Malaysia	This study	(40)
Cp30	Bratislava	Slovakia	This study	(40)
Cp31	Johannesburg	South Africa	This study	(40)
Cp32	Caracas	Venezuela	This study	(40)
Cp34	Johannesburg	South Africa	This study	(40)
Cp35	Johannesburg	South Africa	This study	(40)
Cp36	Detroit	United States	This study	(40)
Cp37	Johannesburg	South Africa	This study	(40)
Cp38	Johannesburg	South Africa	This study	(40)
Cp39	Bratislava	Slovakia	This study	(40)
Cp4	Johannesburg	South Africa	This study	(40)
Cp5	New York	United States	This study	(40)
Cp6	Turino	Italy	This study	(40)
Cp7	Johannesburg	South Africa	This study	(40)
Cp8	Bloemfontein	South Africa	This study	(40)
Cp9	Bloemfontein	South Africa	This study	(40)
CDC317	Mississippi	USA	PRJNA795920	(79)
FM16	Nantes	France	PRJNA795920	(44)
MSK809	New York	USA	PRJNA748054	(44)
73-037	Leeds	UK	PRJNA795920	(44)
73-114	Leeds	UK	PRJNA795920	(44)
Candida orthopsilosis				
151	Atlanta	USA	PRJNA322245	(70)
185	Atlanta	USA	PRJNA322245	(70)
282	Baltimore	USA	PRJNA322245	(70)
320	Baltimore	USA	PRJNA322245	(70)
421	Pisa	Italy	PRJNA322245	(70)
422	Pisa	Italy	PRJNA322245	(70)
423	L'Aquila	Italy	PRJNA322245	(70)

424	Pisa	Italy	PRJNA322245	(70)
425	Pisa	Italy	PRJNA322245	(70)
426	Varese	Italy	PRJNA322245	(70)
427	Pisa	Italy	PRJNA322245	(70)
428	Hong Kong	Hong Kong	PRJNA322245	(70)
433	Sint-Niklaas	Belgium	PRJNA322245	(70)
434	Bristol	UK	PRJNA322245	(70)
435	Pisa	Italy	PRJNA322245	(70)
436	Pisa	Italy	PRJNA322245	(70)
437	Pisa	Italy	PRJNA322245	(70)
498	Baltimore	USA	PRJNA322245	(70)
504	Baltimore	USA	PRJNA322245	(70)
599	Baltimore	USA	PRJNA322245	(70)
748	Atlanta	USA	PRJNA322245	(70)
831	Atlanta	USA	PRJNA322245	(70)
1540	Baltimore	USA	PRJNA322245	(70)
1799	Atlanta	USA	PRJNA322245	(70)
1825	Baltimore	USA	PRJNA322245	(70)
90-125	N/A	N/A	PRJNA431439	(70)
B-8274	N/A	Pakistan	PRJNA322245	(70)
B-8323	N/A	Pakistan	PRJNA322245	(70)
MCO456	N/A	N/A	PRJEB4430	(71)
MSK477	New York	USA	PRJNA579121	(63)
MSK479	New York	USA	PRJNA579121	(63)
MSK616	New York	USA	PRJNA748054	(44)
MSK636	New York	USA	PRJNA579121	(63)
MSK638	New York	USA	PRJNA579121	(63)
MSK639	New York	USA	PRJNA579121	(63)
MSK805	New York	USA	PRJNA748054	(44)
Candida metapsilosis				
MSK403	New York	USA	PRJNA579121	(63)
MSK404	New York	USA	PRJNA579121	(63)
MSK413	New York	USA	PRJNA579121	(63)
MSK414	New York	USA	PRJNA748054	(72)
MSK415	New York	USA	PRJNA579121	(63)
MSK416	New York	USA	PRJNA579121	(63)
MSK417	New York	USA	PRJNA579121	(63)
MSK418	New York	USA	PRJNA579121	(63)
MSK429	New York	USA	PRJNA579121	(63)
MSK430	New York	USA	PRJNA579121	(63)
MSK431	New York	USA	PRJNA579121	(63)
MSK432	New York	USA	PRJNA579121	(63)
MSK433	New York	USA	PRJNA579121	(63)
MSK434	New York	USA	PRJNA579121	(63)
MSK445	New York	USA	PRJNA579121	(63)
MSK446	New York	USA	PRJNA579121	(63)
MSK447	New York	USA	PRJNA579121	(63)
MSK448	New York	USA	PRJNA579121	(63)
MSK449	New York	USA	PRJNA579121	(63)
MSK450	New York	USA	PRJNA579121	(63)
MSK461	New York	USA	PRJNA579121	(63)
MSK462	New York	USA	PRJNA579121	(63)
MSK463	New York	USA	PRJNA579121	(63)
MSK464	New York	USA	PRJNA579121	(63)
MSK465	New York	USA	PRJNA579121	(63)
MSK466	New York	USA	PRJNA579121	(63)
MSK606	New York	USA	PRJNA748054	(72)

MSK607	New York	USA	PRJNA748054	(72)
MSK798	New York	USA	PRJNA748054	(72)
MSK801	New York	USA	PRJNA748054	(72)

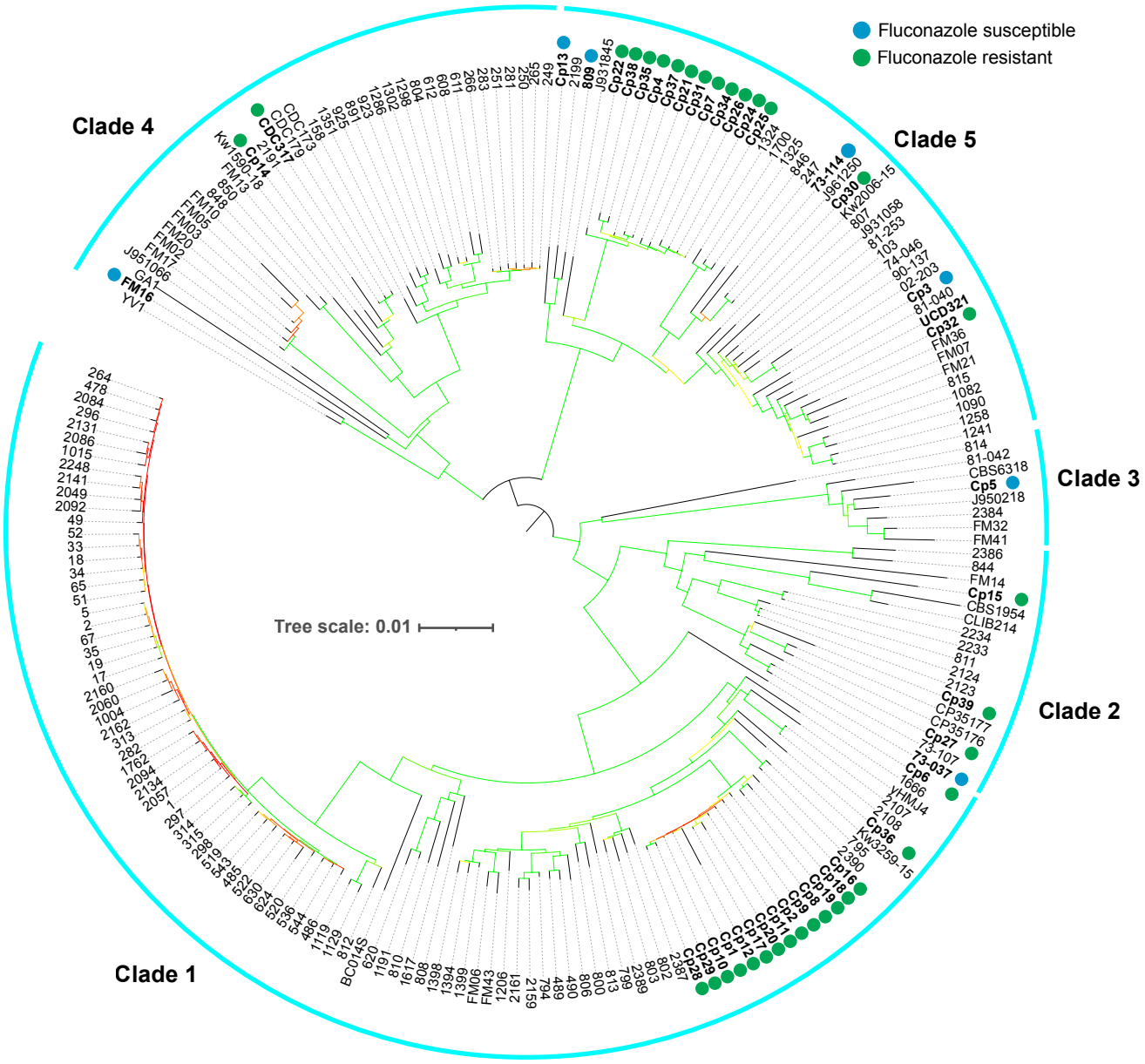

Figure S1.

Maximum-likelihood tree of 207 *Candida parapsilosis* isolates constructed as in Figure 1. Additional isolates, and clade designations were taken from (44). Susceptible isolates in this paper are marked with a blue circle, and resistant isolates are marked with a green circle. Coloured branches indicate bootstrap values after 1000 iterations of bootstrap sampling for that branch ranging from 0 (red) to 100 (green).

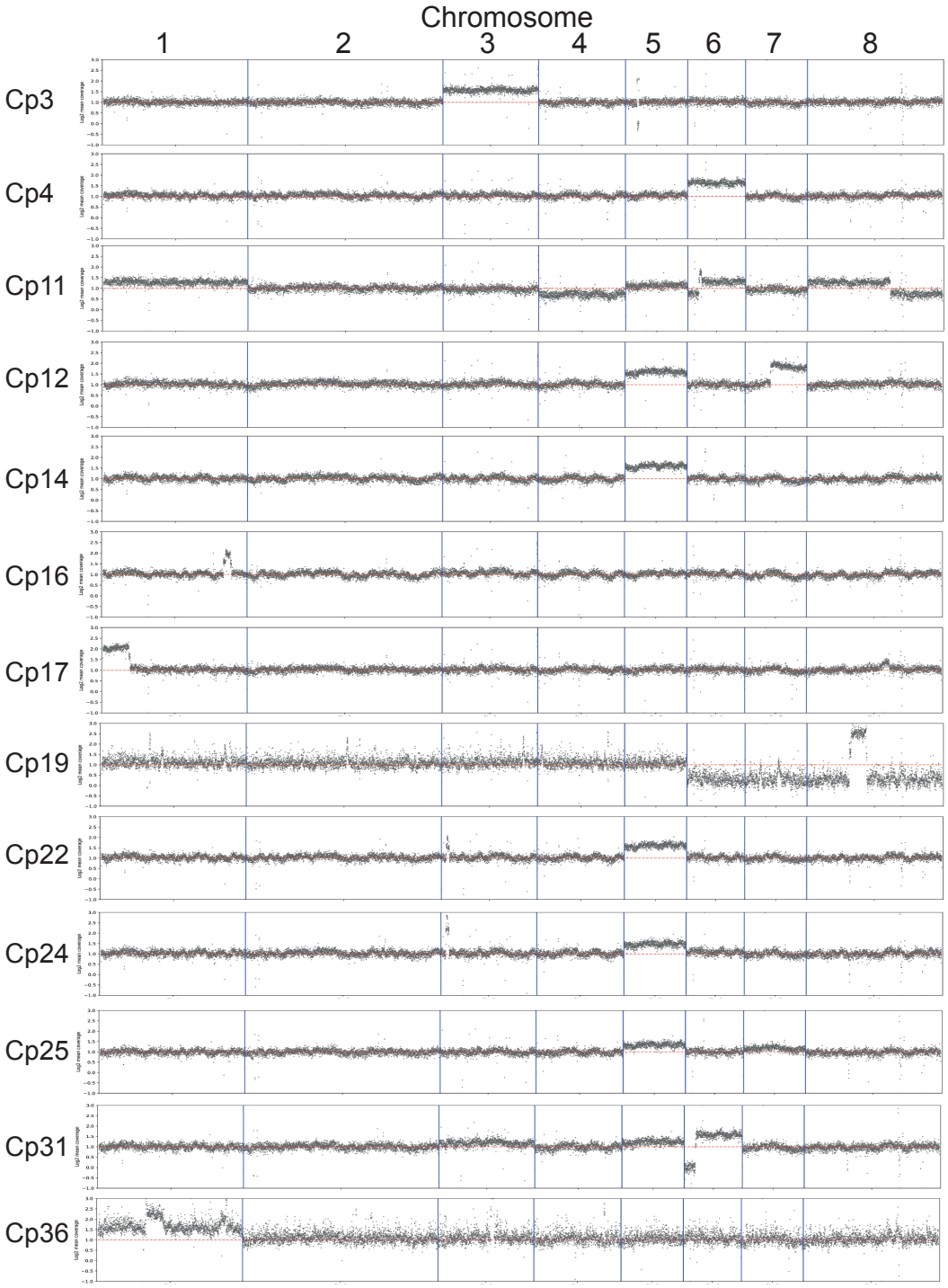

Figure S2.
Genome-wide coverage tracks for aneuploid strains. Coverage was calculated as the log2 mean coverage of 1kb windows and plotted. Dashed red line shows log2 value = 1, i.e. typical 2x diploid coverage