Systematic Analysis of Copy Number Variations in the Pathogenic Yeast *Candida parapsilosis* Identifies a Gene Amplification in *RTA3* That is Associated with Drug Resistance

Sean A. Bergin,[a] Fang Zhao,[a] Adam P. Ryan,[a] Carolin A. Müller,[b] Conrad A. Nieduszynski,[c,d] Bing Zhai,[e,f,g] Thierry Rolling,[e,f] Tobias M. Hohl,[e,f,h] Florent Morio,[i] Jillian Scully,[a] Kenneth H. Wolfe,[j] Geraldine Butler[a]

aSchool of Biomolecular and Biomedical Science, Conway Institute, University College Dublin, Belfield, Dublin, Ireland

bSir William Dunn School of Pathology, University of Oxford, Oxford, United Kingdom

cEarlham Institute, Norwich, United Kingdom

dSchool of Biological Sciences, University of East Anglia, Norwich, United Kingdom

eInfectious Disease Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA

fImmunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, USA

gLaboratory of Quantitative Engineering Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

hDepartment of Medicine, Weill Cornell Medical College, New York, New York, USA

iNantes Université, CHU de Nantes, Cibles et Médicaments des Infections et de l'Immunité, IICiMed, Nantes, France

jSchool of Medicine, Conway Institute, University College Dublin, Belfield, Dublin, Ireland

Sean A. Bergin and Fang Zhao contributed equally to the manuscript. Author order was determined alphabetically.

ABSTRACT We analyzed the genomes of 170 *C. parapsilosis* isolates and identified multiple copy number variations (CNVs). We identified two genes, *RTA3* (CPAR2_104610) and *ARR3* (CPAR2_601050), each of which was the target of multiple independent amplification events. Phylogenetic analysis shows that most of these amplifications originated only once. For *ARR3*, which encodes a putative arsenate transporter, 8 distinct CNVs were identified, ranging in size from 2.3 kb to 10.5 kb with 3 to 23 copies. For *RTA3*, 16 distinct CNVs were identified, ranging in size from 0.3 kb to 4.5 kb with 2 to ~50 copies. One unusual amplification resulted in a DUP-TRP/INV-DUP structure similar to some human CNVs. *RTA3* encodes a putative phosphatidylcholine (PC) floppase which is known to regulate the inward translocation of PC in *Candida albicans*. We found that an increased copy number of *RTA3* correlated with resistance to miltefosine, an alkylphosphocholine drug that affects PC metabolism. Additionally, we conducted an adaptive laboratory evolution experiment in which two *C. parapsilosis* isolates were cultured in increasing concentrations of miltefosine. Two genes, CPAR2_303950 and CPAR2_102700, coding for putative PC flippases homologous to *S. cerevisiae DNF1* gained homozygous protein-disrupting mutations in the evolved strains. Overall, our results show that *C. parapsilosis* can gain resistance to miltefosine, a drug that has recently been granted orphan drug designation approval by the United States Food and Drug Administration for the treatment of invasive candidiasis, through both CNVs or loss-of-function alleles in one of the flippase genes.

IMPORTANCE Copy number variations (CNVs) are an important source of genomic diversity that have been associated with drug resistance. We identify two unusual CNVs in the human fungal pathogen *Candida parapsilosis*. Both target a single gene (*RTA3* or *ARR3*), and they have occurred multiple times in multiple isolates. The copy number of *RTA3*, a putative floppase that controls the inward translocation of lipids in the cell membrane, correlates with resistance to miltefosine, a derivative of phosphatidylcholine (PC) that was originally developed as an anticancer drug. In 2021, miltefosine was designated an orphan drug by the United States Food and Drug

Administration for the treatment of invasive candidiasis. Importantly, we find that resistance to miltefosine is also caused by mutations in flippases, which control the outward movement of lipids, and that many *C. parapsilosis* isolates are prone to easily acquiring an increased resistance to miltefosine.

KEYWORDS *Candida*, copy number variation, genomics, drug resistance evolution

Copy number variations (CNVs), changes in the number of copies at a genomic location, are common in biological systems (1, 2). Many CNVs in human cells are associated with disease, particularly cancers (3). In addition, CNVs in yeasts are frequently identified in industrial isolates and during evolution experiments that examine adaptations to exogenous compounds or to limiting environmental conditions. One of the best-studied examples is the amplification of the *CUP1* metallothionein locus in *Saccharomyces cerevisiae* in response to the presence of toxic copper (4). An increased resistance is associated with the tandem amplification of the *CUP1* open reading frame (ORF) (5, 6). Similarly, limiting sulfate in *S. cerevisiae* induces the amplification of the *SUL1* gene, encoding a sulfate transporter, whereas limiting glucose and amino acids leads to the amplification of the glucose transporter *HXT6* and the amino acid transporter *GAP1,* respectively (1, 2, 7–9). Limiting glutamine also induces the amplification of the urea permease *DUR3* (2). Several different amplifications of *SUL1* and *GAP1* were observed in laboratory evolution experiments, differing in copy number and in the boundaries of the amplification units (2, 7). In addition, the exposure of the pathogenic yeast *Candida albicans* to antifungal drugs, such as azoles, is associated with multiple changes in the genome, including CNVs (10–12).

Multiple CNVs have been identified in natural isolates of many yeasts, including *S. cerevisiae* (13), *C. albicans* (14, 15), and *Candida glabrata* (16). Many affect genes in subtelomeric regions, which are known hot spots for variation (17, 18). Some CNVs in *S. cerevisiae* are lineage-specific, occurring particularly in industrial isolates, and are associated with specific phenotypes (19, 20). However, CNVs in natural isolates that occur outside subtelomeric regions and that differ significantly in copy number and in the size and organization of the amplification unit are rare and are surprisingly poorly studied. *CUP1* is an exception; the copy number has been shown to vary from 0 to ~80 with different endpoints in natural and industrial isolates of *S. cerevisiae* (21–23), though most studies are of amplifications induced in the laboratory by growth in high copper concentrations (24, 25).

Changes in copy number can result from several different mechanisms (26). Where an array already exists (i.e., where there are already at least two copies of a gene at one allele), it can be expanded by unequal crossing-over (nonallelic homologous recombination, NAHR). A misalignment of the alleles results in different numbers of copies following mitosis or meiosis. This may underlie some of the natural variation observed at the *CUP1* locus in *S. cerevisiae* (22). NAHR between inverted repeats can also result in dicentric chromosomes, which are resolved through several breakage-fusion-bridge cycles and result in CNVs (27). These have been observed in azole-resistant isolates of *C. albicans* (10). Many CNVs in *C. albicans* that are induced by exposure to drugs are adjacent to long inverted repeat sequences (10). Some are complex structures, consisting, for example, of a symmetrical "stair-step" amplification, with copy number changes occurring in two steps.

Many CNVs have been hypothesized to be caused by replication-mediated mechanisms. For example, the *CUP1* locus is adjacent to an origin of replication, and the induction of the expression of *CUP1* causes the stalling of the replication fork at this origin (25). The stalled replication fork is repaired by strand invasion in a mechanism similar to break induced replication (BIR) (28). Errors in BIR at repeat regions can result in copy number variation. Microhomology-mediated BIR (MMBIR) occurs where there are short (micro) regions of homology between the collapsed fork and other single-stranded DNA (29). MMBIR may result in expansion at the *DUR3* locus and in many of the *GAP1* expansions described in *S. cerevisiae* (2).

Some unusual amplifications may require a combination of mechanisms. These include a structure with a triplicated inverted central copy surrounded by duplicated

regions (DUP-TRP/INV-DUP) seen in some human CNVs (30). A similar structure was observed in some amplifications of *SUL1* in *S. cerevisiae* (31). The ODIRA (Origin-dependent inverted-repeat amplification) model proposes that these amplifications occur at regions containing an origin of replication flanked by short, inverted repeats. Slippage at one fork could generate closed, circular, self-complementary extrachromosomal intermediates, which are subsequently integrated at the original site (31).

Pryszcz et al. (32) described several CNVs in four isolates of the human fungal pathogen *Candida parapsilosis*, including the amplification of *ARR3*, a putative arsenate transporter. In addition, in 2020, we described an amplification in 23 related *C. parapsilosis* isolates that resulted in a dramatically increased copy number (24 to 33×) of the *RTA3* gene (33). Here, we used genome sequencing to explore CNVs in 170 isolates of *C. parapsilosis* from different sources. We identified 8 different amplifications of *ARR3*, and 3 examples of large stair-step amplifications that are similar to those described in *C. albicans* (10). Notably, we found 16 distinct amplifications of the region surrounding *RTA3* that have unique endpoints, indicative of independent and parallel amplification events. We also found an in-frame fusion to a related neighboring gene, *RTA2*. We identify one CNV with a DUP-TRP/INV-DUP structure at *RTA3* in four isolates. Some allelic expansions of *RTA3* may occur through NAHR, but we found that many isolates have only one copy of *RTA3* at each allele, suggesting that amplification must also occur by other means.

The amplification of *RTA3* increases its transcription and probably its translation. Rta3 is a member of the Rta1/Rsb1-like family in *S. cerevisiae*, which encodes putative transporters with seven transmembrane (TM) domains. ScRsb1 controls the localization of sphingoid bases in *S. cerevisiae*, including phytosphingosine and dihydrosphingosine (34, 35). Rsb1 is localized to the membrane and is likely to act as a membrane transporter (a floppase) (36) or possibly as a regulator of transporters (37). In *C. albicans*, CaRta3 localizes to the plasma membrane; a fluorophore-labeled PC accumulates in the inner leaflet of the membrane in a *CaRTA3* deletion (38). Deleting *CaRTA3* also decreases resistance to the drug miltefosine (hexadecylphosphocholine), an alkylphosphocholine derivative that inhibits PC biosynthesis or localization (39). *CaRTA3* is therefore likely to encode a PC floppase (38).

We find that the copy number of *RTA3* correlates with resistance to miltefosine in *C. parapsilosis*. However, experimentally-induced adaptation in the presence of miltefosine did not induce the amplification of *RTA3* but instead resulted in the inactivation of two flippase genes. Therefore, we show that natural copy number variation in *C. parapsilosis* results in drug resistance but that miltefosine is unlikely to be the cause of selection for amplification. In addition, we show that miltefosine, which was granted orphan drug designation approval by the United States Food and Drug Administration (FDA) for the treatment of invasive candidiasis (https://www.accessdata.fda.gov/scripts/opdlisting/oopd/detailedIndex.cfm?cfgridkey=843921), is a poor treatment choice for *C. parapsilosis* because resistant isolates arise easily in the population.

RESULTS

Phylogeny and CNVs in 170 *C. parapsilosis* isolates. We explored the relationships of 170 *Candida parapsilosis* isolates using a genome-wide SNP alignment approach, generating the largest phylogeny of *C. parapsilosis* isolates to date (Fig. 1). Isolates were obtained from several locations, including some that were previously published and many that are sequenced and described here for the first time (Table S1). Most isolates originated from the Memorial Sloan Kettering Cancer Center in New York (MSK) and the Centre Hospitalier Universitaire de Nantes, France (FM) (Fig. 1).

Isolates fall into five major clades, with almost half (84 out of 170), belonging to Clade 1. Most Clade 1 isolates were isolated at MSK. Half (8 out of 16) of the FM isolates are found in Clade 4. There are 49 highly similar isolates in Clade 1 and 19 highly similar isolates in Clade 4, all from patients at MSK. These may have originated recently from two single isolates, and they are indicated with blue triangles in Fig. 1. Some other
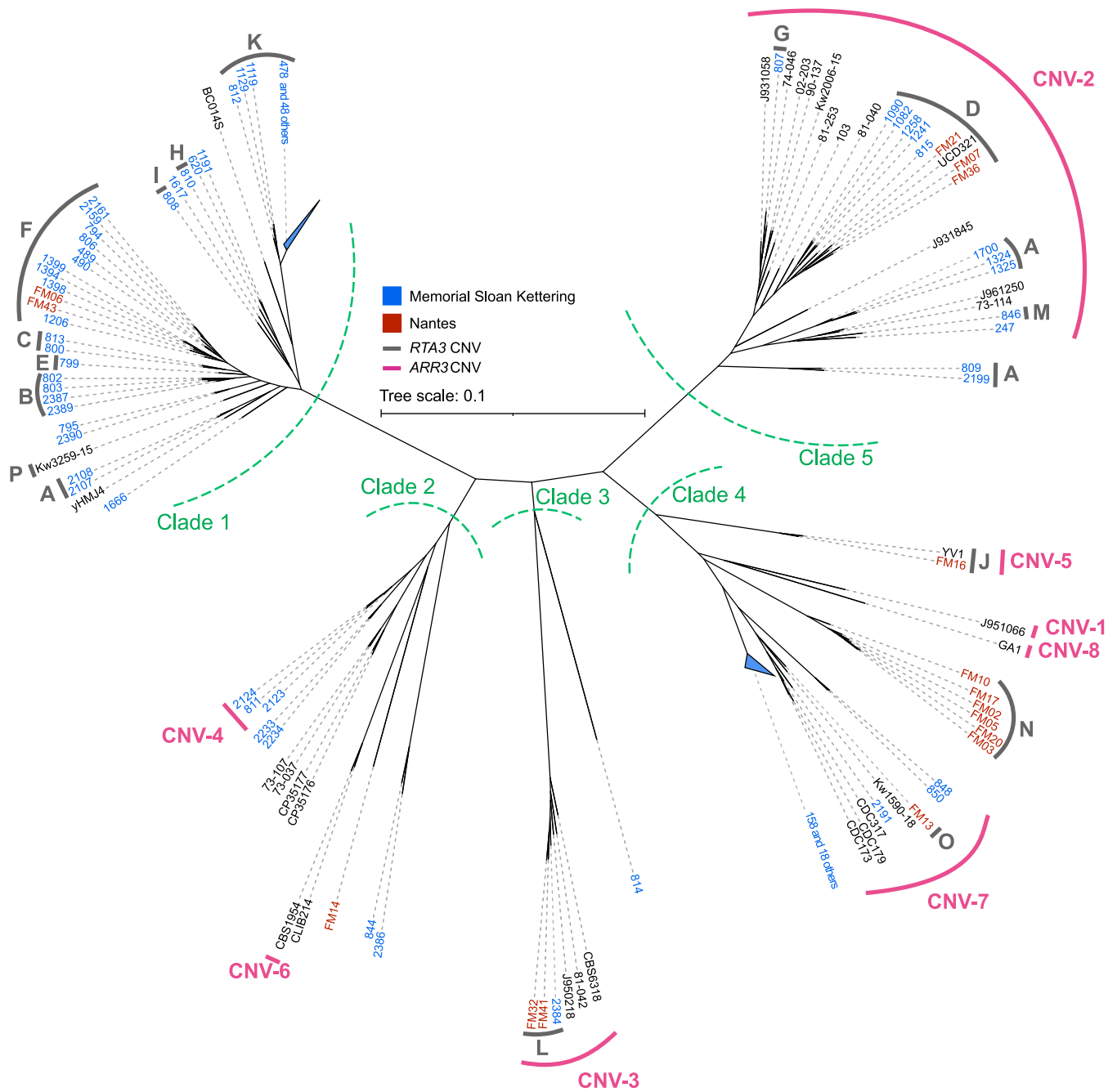
FIG 1 SNP-based unrooted phylogeny of 170 *C. parapsilosis* isolates. Isolates from Memorial Sloan Kettering Cancer Center (MSK) are named in blue (the prefix MSK is omitted from these strain names), and isolates from CHU de Nantes (FM) are named in red. Green dashed lines label each of five apparent clades. 49 similar isolates in Clade 1 and 19 similar isolates in Clade 4 were grouped and are shown as blue triangles. Isolates that harbor a CNV at *RTA3* are marked with a thick gray bar and a letter (from A to P) that corresponds to each of the 16 CNVs. The CNVs at *ARR3* are marked with a pink bar and a number (from CNV-1 to CNV-8). The phylogeny was constructed by calling SNPs for each sample using the GATK HaplotypeCaller tool (77). Filtered heterozygous sites were resolved using 1,000 iterations of random repeated haplotype sampling to provide haploid inputs for tree construction (71). The tree was then constructed using RAxML with the GTRGAMMA model of nucleotide substitution.

isolates (e.g., MSK2233 and MSK2234 in Clade 2, both of which were isolated from the same patient [Table S1]) may also share a recent origin. However, overall, there is little evidence for geographical clustering. Each clade includes at least one isolate from both MSK and FM (Fig. 1). Three isolates from a clinical setting in Kuwait (40) are each located in a different clade (designated by Kw in Fig. 1). Interestingly, an environmental sample isolated from Irish soil, UCD321, groups with clinical isolates in Clade 5 from both MSK and FM (i.e., both the United States of America and Europe). The diversity of isolates
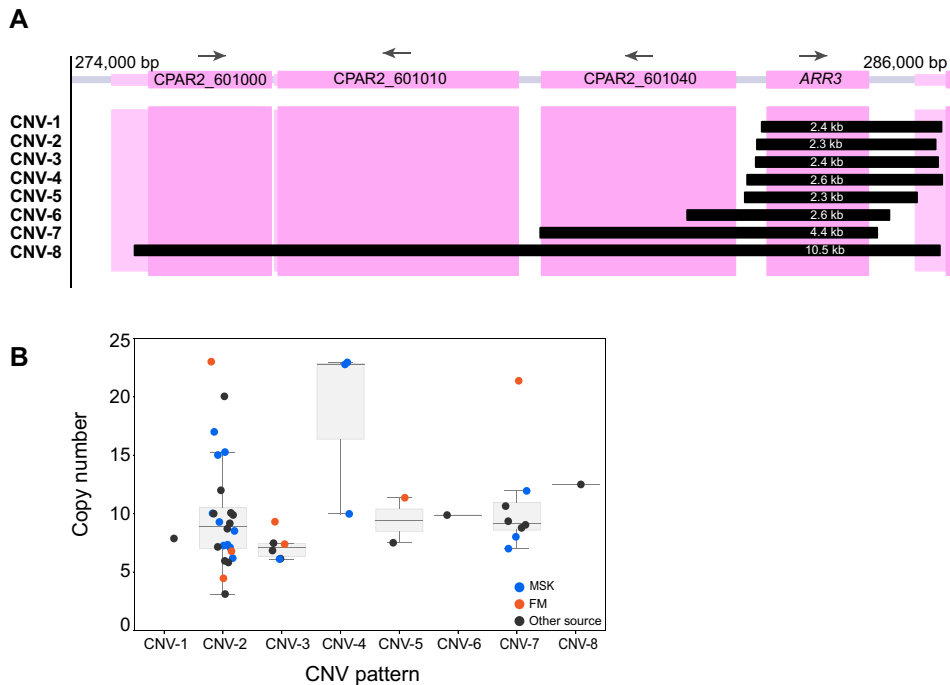
FIG 2 CNVs at the *ARR3* locus. (A) Span of the 8 different CNVs at the *ARR33* locus in *C. parapsilosis*. The coding sequences are shown by a dark pink box, and the flanking UTRs are shown in a lighter color. The extents of the tandemly repeating units in each of the 8 CNVs are shown by black boxes, labeled CNV-1 to CNV-8 on the left, and their lengths are indicated in white. Exact breakpoints were identified by interrogating split reads from the Illumina data, and the array structure of CNV-2 was verified by MinION sequencing of isolate UCD321. (B) Copy number of *ARR3* in each *C. parapsilosis* isolate. Each dot represents a single isolate: blue isolates from MSK, red from CHU de Nantes (FM), and dark gray from other sources. Medians and interquartile ranges are shown for CNVs present in more than one strain. Dots are jittered for clarity.

obtained from the same clinical setting and the close relationship between isolates from different geographical settings highlight the global nature of *C. parapsilosis* as a human pathogen.

Only 9 of the 163 isolates that could be analyzed were aneuploid (one extra copy of chromosome 3, 4, 5, or 6) (Table S2A). Large segmental amplifications were also relatively rare, excluding telomeric and subtelomeric regions, which contained multiple variations in copy number (41, 42). We identified 11 amplifications and 5 deletions of >10 kb in size, 13 of which were found in only one isolate (Table S2B). Three large amplifications (from 125 to 250 kb) in three different isolates have a complex stair-step structure, similar to those described by Todd and Selmecki (10) in the azole-resistant isolates of *C. albicans* (Fig. S1). We also identified ~167 CNVs that are <10 kb in size, 85 of which are found in only one isolate (Table S2B). We further characterized two amplified regions with particularly interesting patterns. These are the only two amplifications that have occurred multiple times in multiple isolates and include an open reading frame.

The first encompasses the gene *CPAR2_601050*, an ortholog of *S. cerevisiae ARR3*, an arsenite transporter (43). Amplifications of *ARR3* with different endpoints were previously described in four *C. parapsilosis* isolates (32). Pryszcz et al. (32) suggested that *ARR3* amplification may be induced in environmental conditions. *ARR3* is amplified in two of the three truly environmental (i.e., non-human-associated) isolates in our analysis (CBS1954, which was isolated from an olive tree in Italy, and UCD321, which was isolated from soil in Ireland [Table S1]) but not in the third (yHMJ4, which was isolated from berries in the United States [44] [Table S1]). In addition, we found *ARR3* amplifications in 44 other isolates associated with humans (Fig. 1; Table S1). In total, 8 different CNV patterns were identified with unique endpoints that ranged in size from 2.3 to 10.5 kb and had copy numbers ranging from 3 to 23 (Fig. 2; Table S1). Three CNVs

extend into the adjacent gene *CPAR2_601040*, and one (previously described by Pryszcz et al. [32]) covers four genes (Fig. 2). An analysis of the strain phylogeny suggests that each amplification type arose only once (Fig. 1). The amplification of a telomeric gene cluster containing *ARR3* together with the transcription factor *ARR1* and the arsenate reductase *ARR2* has been previously reported in natural isolates of *S. cerevisiae* and *Saccharomyces paradoxus* (45), and a large subtelomeric amplicon that includes *ARR3* has been described in isolates from one clade of *Cryptococcus neoformans* var. *grubii* (46). Copy number correlates with resistance to arsenate. *ARR* genes are not clustered in *Candida* species, and *ARR3* is not located at the telomere.

The second amplification is linked to *RTA3* (*CPAR2_104610*). We previously showed that *RTA3*, encoding a putative PC floppase, had undergone extensive copy number amplification in 23 closely related *C. parapsilosis* isolates (33). Amplification was also observed in a small number of isolates by West et al. (47). We now show that the *RTA3* copy number is highly variable. It is increased in 104 of 170 isolates with multiple amplification patterns (Fig. 1 and 3).

16 unique CNVs amplify *RTA3*. We found that *RTA3* has been amplified in 16 different types of CNV patterns, each with unique endpoints (each assigned a letter from A to P) (Fig. 1 and 3). Nine different CNV patterns were observed in isolates from Clade 1, whereas CNV-L is found only in isolates in Clade 3, and there are no CNVs in the Clade 2 isolates (Fig. 1). Most (15 out of 16) of the CNV patterns have a single evolutionary origin, and some are present in only a single isolate (Fig. 1). However, CNV-A may have originated three times (once in Clade 1 and twice in Clade 5).

To determine if the *RTA3* amplifications occur in tandem (and do not form extrachromosomal circles, such as those seen with *CUP1* [48]), we used MinION technology to sequence the genomes of 5 isolates (MSK478, MSK802, MSK803, MSK812, and UCD321), representing three *RTA3* CNV patterns (B, D, and K). In each case, some reads extended across part of the repeat unit and into the flanking DNA on each side of the repeat. This shows that at least for these isolates (and likely for all isolates), the repeats are in tandem on Chromosome 1 and are not extrachromosomal copies.

Thirteen CNV patterns result from tandem duplications that amplify the entire *RTA3* coding sequence (Fig. 3A), and one amplifies the promoter region only (CNV-B). The repeat units that include the *RTA3* ORF range in size from 2.3 to 4.5 kb. Four of these CNVs (CNV-I, J, K, and L) extend into the coding sequence of the upstream neighboring gene *MAK16*, and two (CNV-N and CNV-P) extend into the downstream gene *RTA2*. The copy number of *RTA3* varies both in isolates that share the same CNV pattern and in isolates with different CNV patterns (Fig. 3B). Isolate MSK807 (CNV-G) has the lowest estimated copy number of *RTA3* among the isolates with CNVs at four copies, whereas isolate FM16 (CNV-J) has the highest copy number at 50 copies.

For CNV-K, three isolates have roughly half the *RTA3* copy number of the other isolates with the same CNV pattern ($\sim$13$\times$ instead of $\sim$26$\times$) (Fig. 3B; Table S1). We considered the possibilities that only one allele of *RTA3* was amplified in these isolates (*C. parapsilosis* is diploid) and that both alleles were amplified in other isolates. We explored this issue by using long read sequencing (Oxford Nanopore) of *C. parapsilosis* MSK812, which is one of the CNV-K isolates with fewer copies of *RTA3* (estimated 14 copies) (Table S1). We found that it has 8 copies of *RTA3* at one allele and 6 copies at the other (Fig. S2). Similarly, the sequencing of isolate MSK478 (which also has CNV-K with $\sim$25 copies of *RTA3*) showed that it has at least 11 copies at both alleles (Fig. S2B). Therefore, the variation in copy number among isolates with CNV-K is due to the expansion or contraction of the array in both alleles and is not due to hemizygosity for the amplification.

Two of the *RTA3* CNVs alter the structure of the encoded protein. The amplification in CNV-P starts within the *RTA3* open reading frame and ends within the related adjacent gene *RTA2* (Fig. 3A and C). This repeat structure generates an array of in-frame *RTA2/RTA3* gene fusions, with the N terminus derived from *RTA2* and the C terminus derived from *RTA3* (Fig. 3C). *RTA3* and *RTA2* probably arose from an ancient duplication event, and both are present in many *Candida* clade species, including *C. albicans*
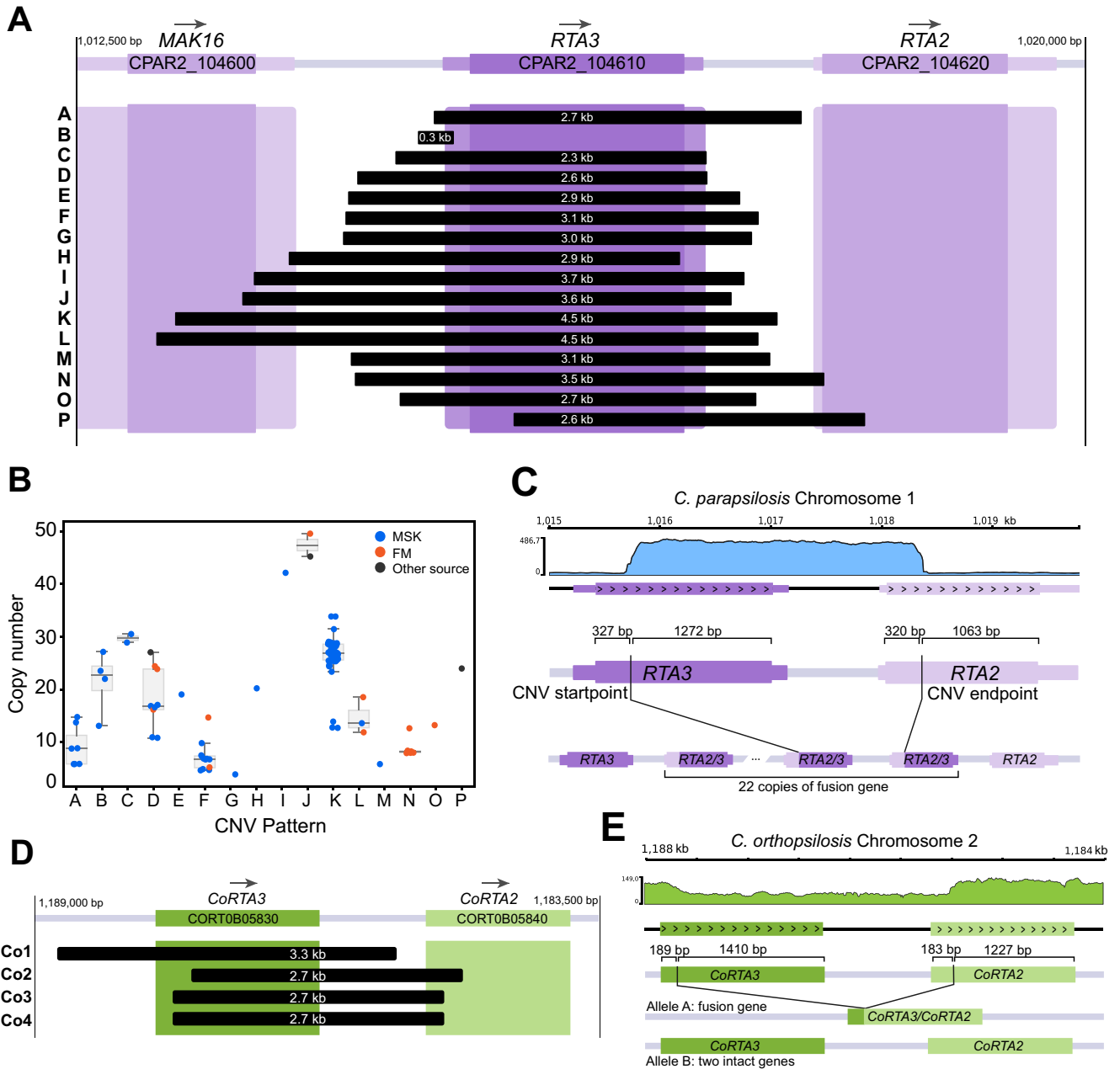
FIG 3 CNVs amplify different sequences at the *RTA3* locus. (A) Span of 16 different CNVs (from A to P) at the *RTA3* locus in *C. parapsilosis*. The coding sequence of *RTA3* is shown by a dark purple box, and the flanking UTRs are shown in a lighter color. The neighboring genes *MAK16* and *RTA2* are also shown. The extents of the tandemly repeating units in each of 16 CNVs are shown by black boxes, labeled (from A to P) on the left, and their lengths are indicated in white. Exact breakpoints were identified by interrogating split reads from the Illumina data, and the array structures of CNVs D and K were verified by MinION sequencing of isolates UCD321, MSK478, and MSK812. (B) Copy number of *RTA3* in each *C. parapsilosis* isolate. Each dot represents a single isolate: blue isolates from MSK, red from CHU de Nantes (FM), and dark gray from other sources. Medians and interquartile ranges are shown for CNVs present in more than one strain. Dots are jittered for clarity. (C) *RTA2/3* fusion gene in strain Kw3259-15 containing CNV-P. A coverage plot created by PyGenomeTracks (78) is shown at the top. The schematic shows the position of CNV breakpoints in relation to the CDS of both genes and the structure of the inferred array of fusion genes. (D) Amplifications of *RTA3* in *C. orthopsilosis* in strain 434. The orientation of the chromosome has been flipped to highlight the similarities to *RTA3* CNVs in *C. parapsilosis*. (E) A deletion in *C. orthopsilosis* generates a *CoRTA3/CoRTA2* fusion. A coverage plot of the locus is shown at the top. The schematic shows CNV breakpoints in both *CoRTA3* and *CoRTA2* and how the resulting fusion gene is formed at one allele while leaving the other allele intact.

(Fig. S3A). However, both *RTA2* and *RTA3* from the *C. parapsilosis* species complex are more closely related to *C. albicans RTA3* than either is to *C. albicans RTA2* (Fig. S3A). This likely resulted from a gene conversion event in the *C. parapsilosis* lineage. The sequences subsequently diverged, including an extension of the C terminus in Rta3

(532 aa) in *C. parapsilosis* that is not present in Rta2 (460 aa) (Fig. S3B). In CNV-H, the Rta3 protein is slightly truncated because this CNV consists of a tandem duplication with an endpoint 31 bp upstream of the stop codon of *RTA3*, resulting in a protein that is 10 amino acids shorter than its wild type counterpart.

CNV-B is substantially different from the others because it is an amplification of a small region (269 bp) that resides upstream and completely outside the *RTA3* coding sequence (Fig. 3A). Sequencing read coverage analysis suggested that the repeat has a complex organization in which direct repeats are interspersed with inverted copies of a central segment flanked by inverted-repeats, resulting in an N:($2N$-1):N copy number pattern (Fig. S4A). This repeat structure was confirmed in MinION reads from isolates MSK802 and MSK803. The structure of the CNV-B repeat is similar to the DUP-TRP/INV-DUP structure seen in some human CNVs (30), although CNV-B repeats are much smaller and are reminiscent of amplifications formed via origin-dependent inverted repeat amplification (ODIRA) (9, 31). ODIRA results in complex CNVs with repeat units in head-to-head and tail-to-tail arrangements that are similar to the CNV-B structure. If CNV-B were caused by ODIRA, we would anticipate that the central region of the repeat would contain an origin of replication (Fig. S4B). However, we determined the temporal order of replication in *C. parapsilosis* using SORT-seq (49) (Fig. S4C), and we found no evidence that there is an origin near *RTA3*.

We also looked for evidence of *RTA3* and *ARR3* amplification in other *Candida* species. *RTA3* is not amplified in 200 sequenced *C. albicans* genomes (50–53). However, we identified 4 different *RTA3* CNVs among 36 strains of *Candida orthopsilosis* (33, 52, 53) (Table S1), a close relative of *C. parapsilosis* (Fig. 3D). Only one of these (CNV-Co1) consists of a simple tandem amplification of the entire *CoRTA3* ORF. Two others, CNV-Co2 and CNV-Co3, amplify a fusion of the *CoRTA2* and *CoRTA3* genes, similar to CNV-P in *C. parapsilosis* (Fig. 3D). Interestingly, the last *C. orthopsilosis* CNV (CNV-Co4) involves a deletion. In the two strains containing CNV-Co4, the 3′ end of *CoRTA3*, the 5′ end of *CoRTA2*, and the intergenic space between them are deleted at one allele (Fig. 3E). This results in a new fusion gene with the N terminus derived from *CoRTA3* and the C terminus derived from *CoRTA2*. Single copies of *CoRTA2* and *CoRTA3* are intact on the other allele (Fig. 3E). This is the only example we have seen of an *RTA3* deletion in *C. parapsilosis*, *C. orthopsilosis*, or *C. albicans*. We did not observe the amplification of *ARR3* in *C. albicans* or *C. orthopsilosis*.

The copy number of *RTA3* is correlated with miltefosine, but not fluconazole, resistance. The deletion of *RTA3* in *C. albicans* has been shown to increase susceptibility to miltefosine (38) and possibly to fluconazole (54). Therefore, we investigated the effect of copy number on the resistance of *C. parapsilosis*. Fig. 4A shows that the amplification of *RTA3* is associated with miltefosine resistance. For example, *C. parapsilosis* strains with only two copies of *RTA3* (one at each allele, e.g., the reference strain CLIB214) fail to grow at miltefosine concentrations of 10 μg/mL, whereas all of the strains with *RTA3* amplifications can grow at this concentration. Strains with CNVs A, H, I, J, and L can tolerate miltefosine concentrations up to at least 30 μg/mL, as can isolates with CNV-B (which amplifies only the region upstream of *RTA3*). The CNV with the weakest effect on MF resistance is CNV-G, which can tolerate 10 to 15 μg/mL, which is still higher than the tolerance of the reference strain CLIB214.

When different isolates carrying the same CNV pattern are compared, miltefosine resistance correlated with the copy number of *RTA3* (Fig. 4B). For example, *C. parapsilosis* isolates MSK812 and MSK1129, which have 13 to 14 copies of CNV-K, can tolerate a miltefosine concentration of up to ~20 μg/mL, whereas *C. parapsilosis* MSK2086, which has 27 copies, survives up to 30 μg/mL. Similarly, *C. parapsilosis* MSK815 with 17 copies of CNV-D and MSK794 with 5 copies of CNV-F tolerate miltefosine concentrations up to ~20 to 25 μg/mL, whereas UCD321 and FM43 with 27 and 15 copies of CNV-D and CNV-F, respectively, survive up to ~30 μg/mL (Fig. 4B).

Although the isolates have variable levels of sensitivity to fluconazole (Fig. 4), the copy number of *RTA3* does not correlate with susceptibility to this drug. For example, isolates MSK815 and UCD321, which have 17 and 27 copies of CNV-D, differ in their susceptibility to miltefosine, but they both tolerate fluconazole levels of 6 μg/mL.
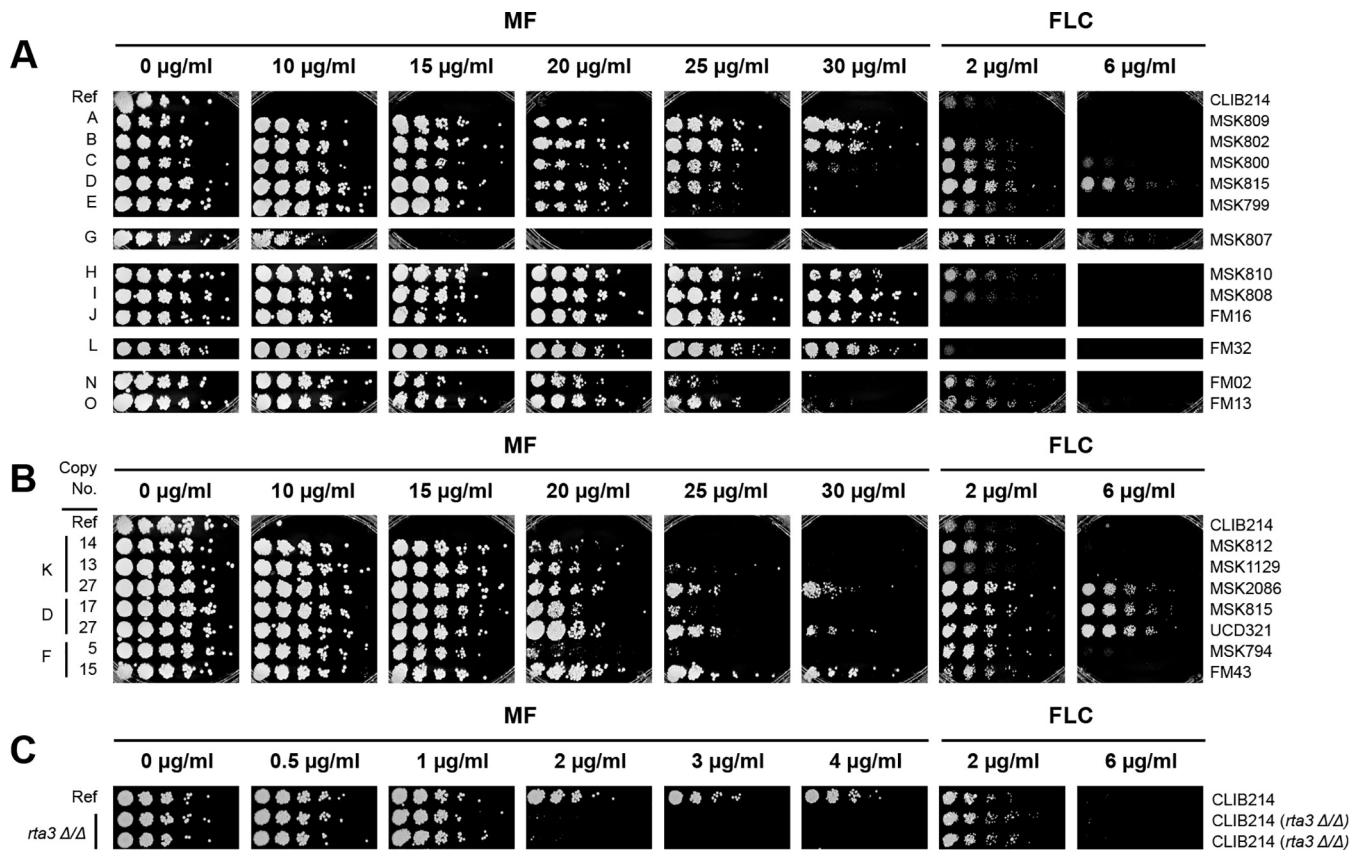
FIG 4 CNVs of *RTA3* correlate with resistance to miltefosine. (A) Multiple *RTA3* CNVs are associated with miltefosine resistance. A representative isolate of most CNV patterns (Fig. 2) was grown on YPD with increasing concentrations of miltefosine (MF) or fluconazole (FLC) as shown. Cells were serially diluted (1/5). The CNV pattern is indicated on the left, and the strain name is shown on the right. MF and FLC plates were incubated for 48 h. "Ref" indicates the reference strain, *C. parapsilosis* CLIB214. (B) Copy number is directly correlated with miltefosine resistance. Isolates with different copy numbers of CNV-K, CNV-D and CNV-F were grown as in (A). The *RTA3* copy number of each isolate is shown on the left. (C) Deleting *RTA3* reduces resistance to miltefosine. *RTA3* was deleted in *C. parapsilosis* CLIB214 using CRISPR-Cas9. The growth of two biological replicates is shown as in (A), except that the incubations on FLC were for 72 h.

We found that deleting *RTA3* in the reference isolate of *C. parapsilosis* (CLIB214) by CRISPR-Cas9 editing (55) results in an increased sensitivity to miltefosine (Fig. 4C). Low levels of miltefosine (up to 4 μg/mL) were used because the parental strain without any *RTA3* amplifications is highly sensitive. However, susceptibility to fluconazole was not affected by the deletion of *RTA3* (Fig. 4C).

Increasing the *RTA3* copy number correlates with increased expression, as shown by West et al. (47). However, we wondered whether amplifying the upstream region (CNV-B) has the same effect as amplifying the entire open reading frame. To explore this further, we measured RTA3 expression by reverse transcription polymerase chain reaction (RT-PCR) in one isolate of *C. parapsilosis*, MSK808, with approximately 42 copies of *RTA3* (CNV-I), and two other isolates (MSK802 and MSK803) in which only the promoter region was amplified (CNV-B). We found that *RTA3* expression is approximately 22-fold higher in *C. parapsilosis* MSK808 and is 2.8 to 6.6-fold higher in *C. parapsilosis* MSK802 and MSK803 than in the reference strain CLIB214 (Table 1). There are 28 copies of the direct repeat in CNV-B in *C. parapsilosis* MSK802 and 24 in *C. parapsilosis* MSK803. Thus, *RTA3* promoter region amplification (CNV-B) can also lead to increased expression.

Generation of miltefosine-resistant strains by experimental evolution. The prevalence of *RTA3* amplification in *C. parapsilosis* isolates suggests that there may be some strong selective pressure inducing amplification. To determine whether miltefosine was the driving force, we characterized the effect of exposing isolates to increasing concentrations of miltefosine in an adaptive laboratory evolution approach. We started with two isolates with only one copy of *RTA3* at each allele that were in the same clade as other

TABLE 1 Expression of *RTA3* in strains with different amplifications

Isolate	Relative expression	Range	*P* value
MSK802 (CNV-B)	6.57	2.12 to 20.35	7.37E−03
MSK803 (CNV-B)	2.79	1.24 to 6.25	3.39E−02
MSK808 (CNV-I)	22.50	6.52 to 77.62	6.92E−05
CLIB214 (Reference)	1.00	0.19 to 5.27	NA

isolates that had undergone amplification: *C. parapsilosis* MSK795 from Clade 1, which is related to isolates that have undergone 4 different CNVs (B, C, E, and F), and *C. parapsilosis* MSK247 from Clade 5, which is related to isolates with CNVs A, D, J, and M (Fig. 1). Five lineages were evolved from MSK247 (247A to 247E), and one was evolved from MSK795 (795B). Isolates were cultured in YPD with increasing concentrations of miltefosine, up to a maximum of 32 μg/mL, over a 26-day period (Fig. 5A). 16 randomly picked evolved colonies from each lineage tolerated miltefosine levels of 40 μg/mL (Fig. 4B). The genomes of 10 isolates derived from MSK247 and two from MSK795 were sequenced together with the parental strains. The sequenced strains are listed in the Materials and Methods section.

Miltefosine-resistant isolates acquired homozygous disruptions in two flippase genes. We did not find any evidence of the amplification of the *RTA3* locus in any of the evolved miltefosine-resistant strains. However, by comparing the sequences of the evolved strains to those of the parental strains, we identified two genes with homozygous loss-of-function variants in all 10 resistant isolates: *CPAR2_102700* and *CPAR2_303950*. These variants include frameshifts, nonsense mutations, and missense mutations that are predicted to be deleterious (Fig. 6A). *CPAR2_102700* and *CPAR2_303950* encode putative Class 3 P4-ATPases and are homologs of the PC flippase genes *DNF1* and *DNF2* in *S. cerevisiae* (Fig. 6B).

The two sequenced isolates derived from *C. parapsilosis* MSK795 (795B1 and 795B16) acquired mutations in both *CPAR2_102700* and *CPAR2_303950*, whereas the lineages derived from *C. parapsilosis* MSK247 acquired mutations only in *CPAR2_303950* (Fig. 6A). However, subsequent analysis revealed that *C. parapsilosis* MSK247 contains a homozygous natural variant in *CPAR2_102700*, resulting in a Trp-to-Stop nonsense mutation (W1280X) (Fig. 6A). *C. parapsilosis* MSK247 tolerates miltefosine concentrations of 5 μg/mL, whereas *C. parapsilosis* MSK795 fails to grow (Fig. 6C). Derivatives of both parents that carry homozygous inactivating mutations in both *CPAR2_303950* and *CPAR2_102700* can grow up to 30 μg/mL.

This result suggests that the reaching maximum level of miltefosine resistance requires the inactivation of both genes, *CPAR2_303950* and *CPAR2_102700*. To test this hypothesis, we deleted them, both separately and together, in the *C. parapsilosis* CLIB214 genetic background using CRISPR-Cas9 editing (Fig. 6C). Deleting *CPAR2_102700* alone in *C. parapsilosis* CLIB214 allows growth up to 5 μg/mL miltefosine, whereas deleting *CPAR2_303950* alone has no effect (Fig. 6C). However, strains in which both *CPAR2_303950* and *CPAR2_102700* were deleted tolerated at least 30 μg/mL miltefosine, similar to strains derived from the experimentally evolved isolates (MSK795B1, MSK247C1) (Fig. 6C). Deleting *CPAR2_102700* alone or in combination with *CPAR2_303950* slightly increases sensitivity to fluconazole (Fig. 6C).

DISCUSSION

We identified two examples of CNVs in *C. parapsilosis* with unusual patterns that are likely to provide interesting models for studying amplification mechanisms. The amplifications occur in tandem, and, at least for *RTA3*, the copy number at each allele can vary. This suggests that the expansion and contraction of the CNV may occur by NAHR following misalignments at alleles with several gene copies. It is likely that the *RTA3/RTA2* gene fusion in CNV-P also originated by NAHR, due to the high sequence similarity between *RTA3* and *RTA2*. However, many *C. parapsilosis* isolates have only one copy of *RTA3* and *ARR3* at each allele (Fig. 1), meaning that some other mechanism must therefore underlie the initial amplification step. Possibilities include BIR and MMBIR, which are used to restart replication at collapsed replication forks. For example, a
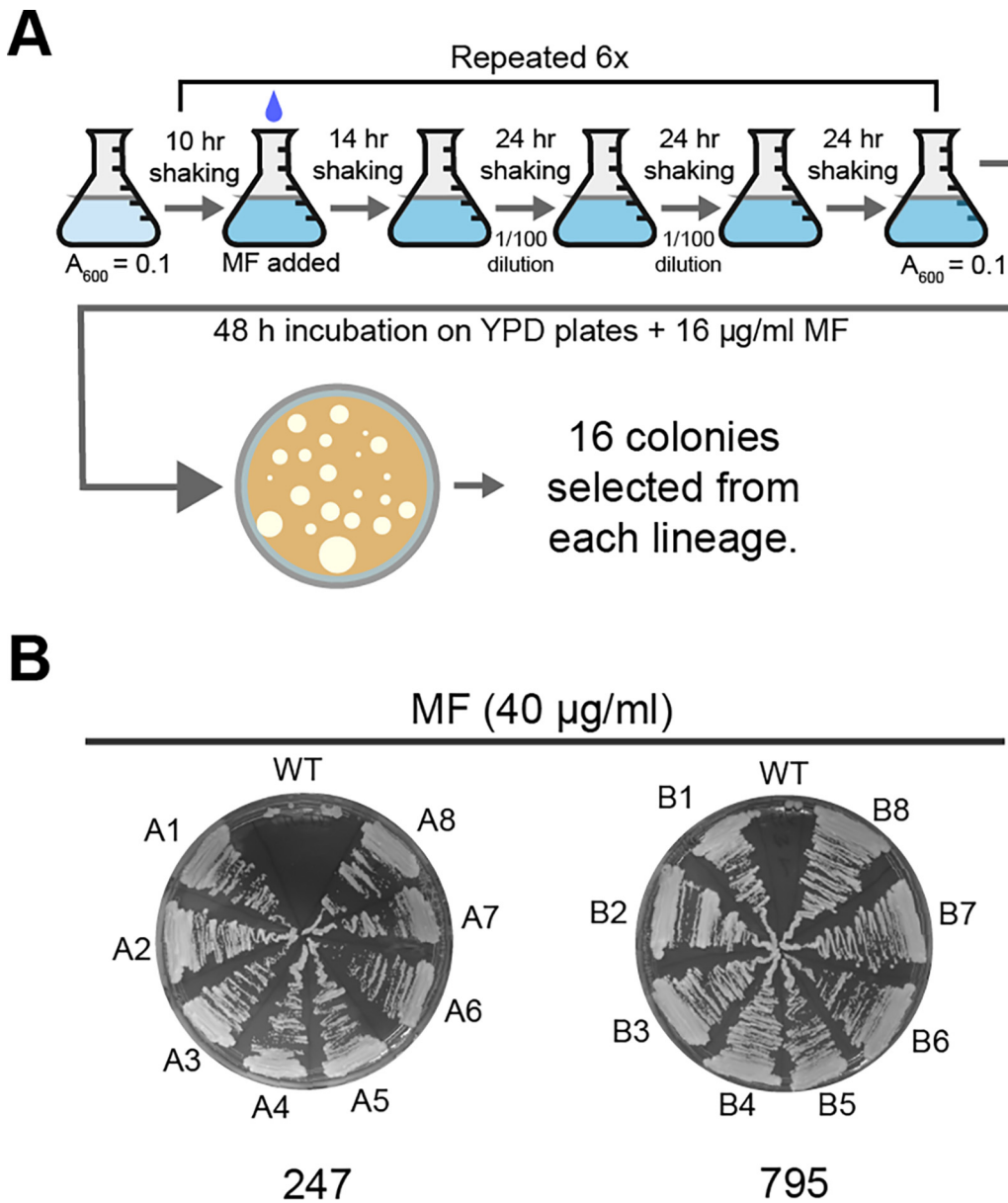
FIG 5 Generation of miltefosine-resistant *C. parapsilosis* isolates by adaptive laboratory evolution. (A) Six colonies (5 derived from *C. parapsilosis* MSK247 and one from *C. parapsilosis* MSK795) were inoculated in YPD at A_{600} = 0.1 and incubated at 30℃ for 10 h. Miltefosine (indicated with a blue drop) was added to a final concentration of 1 μg/mL. The cultures were incubated for a further 14 h, then reinoculated into the same media using a 1/100 dilution, and incubated for 24 h. The dilution was repeated every 24 h for 3 days. The cells were then inoculated into fresh media and grown for 10 h, and then the miltefosine concentration was doubled. The process was repeated until the concentration of miltefosine reached 32 μg/mL (24 days). On day 25, the overnight cultures were plated on YPD with 16 μg/mL miltefosine. Sixteen colonies were picked from each lineage for further analysis, and the genomes of 10 of them were sequenced. (B) Growth of representative isolates evolved from *C. parapsilosis* MSK247 and *C. parapsilosis* MSK795 on 40 μg/mL miltefosine. WT = parental strain.

collision between RNA polymerase II at the *CUP1* promoter and replication from the adjacent origin may lead to fork collapse, which would be repaired by BIR (25, 56). The CNV-B pattern of *RTA3* amplification (in which only a short sequence upstream from the coding sequence is amplified) is particularly intriguing because it results in inversions and triplications of parts of the repeated sequence (Fig. S4A). This structure is reminiscent of the ODIRA model for CNV generation at *SUL1* in *S. cerevisiae*, in which replication errors at short inverted repeats result in the formation of an autonomously-replicating, extrachromosomal "dog bone" structure with inverted regions (31). These
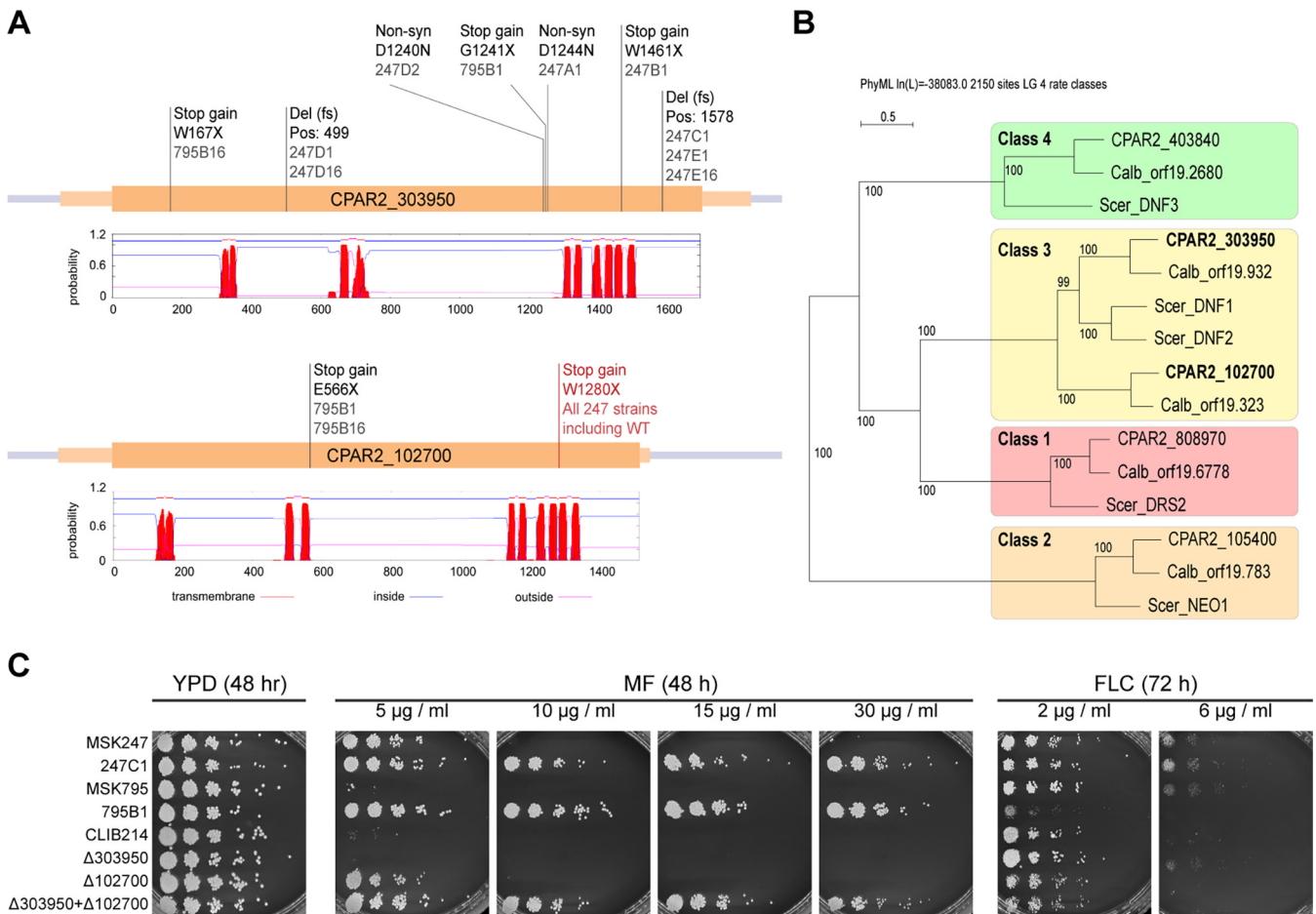
FIG 6 Protein-disrupting mutations in two flippase genes confer resistance to miltefosine in laboratory evolved strains. (A) Schematic showing homozygous mutations in P4-ATPases *CPAR2_102700* and *CPAR2_303950*. Mutations named in black were acquired during the laboratory evolution experiment in the evolved strains named below the mutation name. The mutation named in red is a natural homozygous variant found in the *C. parapsilosis* MSK247 isolate and in all strains derived from it. The predicted transmembrane topology of each protein (from TMHMM [79]) is shown, in which red peaks show predicted transmembrane domains. (B) Phylogeny of P4-ATPases in *C. parapsilosis*, *C. albicans*, and *S. cerevisiae*. The P4-ATPase genes in three yeast species were identified by BLAST and aligned, and a tree was constructed using SeaView (80). Monophyletic groups of genes were assigned to established classes (1–4, 81). Both *CPAR2_102700* and *CPAR2_303950* (in bold) belong to the Class 3 P4-ATPases. (C) Deleting *CPAR2_102700* and *CPAR2_303950* increases resistance to miltefosine but not to fluconazole. *CPAR2_102700* and *CPAR2_303950* were deleted singly or together in the *C. parapsilosis* CLIB214 background, and growth was compared to isolates evolved from *C. parapsilosis* MSK247 (247C1) and *C. parapsilosis* MSK795 (795B1) as in Fig. 5.

are proposed to subsequently reintegrate at *SUL1*, thereby forming a DUP-TRP/INV-DUP CNV in which some regions are duplicated (DUP) and others are triplicated (TRP), surrounding an inverted region. To date, ODIRA has only been observed in *S. cerevisiae*, and it is unclear whether it occurs in other yeasts (31).

Of the 14 CNVs resulting in the tandem array amplification of *RTA3*, only CNV-A and CNV-E have obvious repeat sequences at the CNV endpoints, and these are direct repeats of 19 bp and 10 bp, respectively. The other 12 have either no repeats or repeats of <5 bp (Fig. S5). The longer repeats at the breakpoints of CNV-A may explain why it has originated independently three times (Fig. 1). Short repeats can result in copy number variation from template switching in MMBIR (57). However, we did not find any evidence of a replication origin near *RTA3* or in the amplified sequence of the CNV-B pattern (Fig. S4C). This makes it unlikely that the amplification patterns A, B, or E use MMBIR or similar mechanisms to *CUP1* (fork stalling at repeats) or *SUL1* (ODIRA) in *S. cerevisiae*. No obvious mechanisms explain the CNVs without terminal repeats. However, even in *CUP1*, it is not clear whether the nearby origin of replication fires during a regular S phase (25), and it remains possible that a rarely used origin is present near *RTA3* or *ARR3* in *C. parapsilosis*.

We found that the copy number of *RTA3* correlates with resistance to miltefosine, a PC analog. This suggests that Rta3 controls the localization of PC. There is, however, some debate over whether Rta3-like proteins are transporters or signaling receptors. Rta3 is a member of the Rta1-family, which encodes proteins with 7 transmembrane domains, similar to the structure of G-protein-coupled receptors (GPCRs) (36). Within this family, Rta3 is more closely related to *S. cerevisiae* Rsb1 than to other members. Most of the early evidence suggested that Rsb1 directly flips LCBs in the plasma membrane (35). However, Johnson et al. (37) suggested that rather than acting as a transporter, Rsb1 determines resistance to the LCB phytosphingosine by regulating the endocytosis of the tryptophan transporter Tat2 either by signaling through a G-protein (as a GPCR) or by an arrestin-mediated effect on the targeting of ubiquitin ligases. In addition, Srivistava et al. (38) found that deleting *RTA3* in *C. albicans* resulted in the increased flipping of a fluorescently labeled PC, possibly by controlling the activity of an unknown flippase. An increase in copy number of *RTA3* and *RTA2* in *C. albicans* through chromosome duplication also affects tolerance to tunicamycin, an inducer of endoplasmic reticulum (ER) stress, though the exact mechanism is not known (58, 59).

Several pieces of evidence suggest that Rsb1 is a direct lipid transporter, specifically, a floppase. First, Rsb1 shares no sequence similarity with other yeast GPCRs (36). Second, Makuta et al. (60) showed that Rsb1 (and not other Rta1-family members) regulates LCB transport and that this activity is dependent on a loop region following TMS5 and not on the C terminus, as expected for a GPCR. Rsb1 in *S. cerevisiae* and Rta3 in *C. albicans* are located in the plasma membrane (34, 54). The increased flipping observed by Srivistava et al. (2017) (38) may be due to cross talk between sphingolipids and glycerophospholipids (e.g., PC), as shown in *S. cerevisiae* by Kihara and Igarashi (35). Changes in the distribution of one may be compensated by changing the distribution of the other (36). The association of *RTA3* copy number amplification with miltefosine resistance in *C. parapsilosis* is consistent with a role as a transporter. Our results also suggest that *C. parapsilosis CPAR2_102700* and *CPAR2_303950*, members of the class 3 family of P4-ATPases, are the direct flippases of PC. However, our data cannot rule out the alternative model of Johnson et al. (37), in which Rta3 regulates the functions of *CPAR2_102700* and *CPAR2_303950* rather than directly acting as a transporter.

Our results strongly suggest that there is selective pressure driving the amplification of *RTA3* and *ARR3* in *C. parapsilosis*. This conclusion is based on our observation that amplification occurred in many isolates that were from diverse genetic and environmental backgrounds and were distributed across the *C. parapsilosis* phylogeny (Fig. 1). We also found that there have been multiple independent events of amplification with at least 16 unique endpoints for *RTA3* and 8 for *ARR3*. The CNVs sometimes encompass parts of the adjacent genes, but *RTA3* and *ARR3* are the only complete genes amplified in all of them. There are some unusual features in *RTA3*. For example, in CNV-B, the likely promoter is amplified, and the 3′ is slightly truncated in the amplified copies in CNV-H. West et al. (47) previously described *RTA3* amplifications in 4 *C. parapsilosis* isolates, including one from the New York subway. Those amplifications also have different endpoints, but because the data come from metagenomics analyses, we cannot determine if their endpoints differ from those of the 16 CNVs we describe.

Most similar CNVs that have been described in other species occur during experimental adaptation to environmental conditions, including amplifications of *CUP1*, *SUL1*, *HXT6*, *GAP1*, and *DUR3* in *S. cerevisiae* (5, 7–9). From this group, only *CUP1* amplifications have been described in natural isolates. Other amplifications in natural isolates (including *ARR3*) occur in subtelomeric regions (45, 46). The majority of the *C. parapsilosis* isolates in this study are associated with humans (either clinical or from healthy donors), although one with an *RTA3* amplification (*C. parapsilosis* UCD321) and two with *ARR3* amplifications (CBS1954 and UCD321) have environmental origins (soil and olive tree) (Table S1). Whereas the *ARR3* amplifications may be induced by the presence of arsenate (45, 46), it is unlikely that the *RTA3* amplifications were driven by exposure to miltefosine because it is not a commonly used antifungal drug. In addition,

as we have shown, exposure to miltefosine is likely to result in loss-of-function mutations in flippase genes. At present, we do not know what kind of selective pressure led to the widespread amplifications of *RTA3* in *C. parapsilosis*.

Miltefosine has potential as an antifungal drug (61–63), and it has recently been designated an orphan drug for the treatment of invasive candidiasis (https://www.accessdata.fda.gov/scripts/opdlisting/oopd/detailedIndex.cfm?cfgridkey=843921). However, our analyses suggest that it would be a poor choice, at least for invasive *Candida parapsilosis* infections. Many isolates are naturally resistant because of amplifications of *RTA3*, and in others, resistance rapidly arises due to loss-of-function mutations in the flippase genes. The resistance of *Leishmania donovanai* to miltefosine is also associated with mutations in flippases (64). Although mutations in two flippase genes are required for resistance in *C. parapsilosis*, we note that many of the *C. parapsilosis* samples that we examined (56 out of 170) have predicted loss-of-function variants in *CPAR2_102700*, meaning that acquiring a mutation in *CPAR2_303950* would be sufficient to render them highly resistant to miltefosine.

MATERIALS AND METHODS

Strains and growth. The isolates used are listed in Table S1. The isolates were maintained on YPD agar (1% Bacto Yeast Extract [212750, Sigma], 2% Bacto Peptone [211677, Sigma], 2% Bacto Agar [214010, Sigma], 2% D-[+]-Glucose [G8270, Sigma]), and liquid cultures were grown in 5 mL YPD broth without agar at 30°C and 200 rpm shaking overnight. For serial dilutions, 0.5 mL of the overnight cultures were harvested at 13,000 rpm at room temperature for 1 min, washed twice in 0.5 mL of phosphate-buffered saline (PBS) buffer (BR0014G, Thermo Fisher), resuspended in 0.5 mL PBS, and diluted to A_{600} = 0.0625 (approximately 6.25×10^5 cells/mL) in PBS. Five-fold serial dilutions were made in PBS and transferred with a pinner to YPD agar plates containing miltefosine (M5571, Sigma-Aldrich) or fluconazole (F8929, Sigma-Aldrich) at the indicated concentrations. The pinned plates were incubated at 30°C for the indicated times and photographed using a Singer PhenoBooth.

Gene deletions/disruptions. The entire open reading frames of *CPAR2_104610* (*RTA3*) and the flippases *CPAR2_303950* and *CPAR2_102700* were deleted using CRISPR-Cas9 with the pCP-tRNA system, as described in Lombardi et al. (55). All of the primers used for the gRNA and repair template synthesis are listed in Table S3.

Microevolution of miltefosine resistant isolates. The microevolution method was modified from Papp et al. (65) and Ene et al. (66). Three colonies from *C. parapsilosis* MSK247 and *C. parapsilosis* MSK795 were originally chosen. Subsequent analysis showed that 5 evolved lineages were derived from *C. parapsilosis* MSK247 (247A to E) and that one was derived from *C. parapsilosis* MSK795 (795B). The colonies were incubated for 10 h in 5 mL YPD at 200 rpm and 30°C. Miltefosine was added to a final concentration of 1.0 μg/mL, and the cultures were incubated for a further 14 h. The cultures were diluted (1:100) into fresh media containing the same concentration of miltefosine every 24 h for 3 days. Overnight cultures were then diluted to A_{600} = 0.1 and incubated for 10 h, and the miltefosine concentration was doubled. This 5-day cycle was repeated until a concentration of 32 μg/mL of miltefosine was reached (Fig. 4A). The cultures were diluted and plated on YPD agar plates containing 16 μg/mL miltefosine, yielding 100 to 200 colonies on each plate, and were incubated at 30°C for 48 h. The genomes of 10 resistant isolates (247A1, 247B1, 247C1, 247D1, 247D2, 247D16, 247E1, 247E16, 797B1, and 795B16), together with the parental strains, were sequenced by the Beijing Genomic Institute via DNBseq.

Illumina sequencing. Illumina sequencing of the MSK isolates was carried out as described in Zhai et al. (33). 45 *C. parapsilosis* isolates from Centre Hospitalier Universitaire de Nantes, France were screened for resistance to miltefosine, and 16 isolates were chosen for sequencing. Genomic DNA was isolated from 5 mL overnight cultures in YPD at 30°C using phenol-chloroform-isoamyl alcohol extraction. Cell pellets were resuspended in 200 μL of extraction buffer (Triton X-100 2% m/v, NaCl 100 mM, Tris 10 mM [pH 7.4], EDTA 1 mM, SDS 1% m/v) and transferred to screw cap tubes, and then approximately 0.3 g of acid-washed beads and 200 μL of phenol-chloroform-isoamyl alcohol (25:24:1) was added. Cells were lysed using a 1600 MiniG bead beater from Spex SamplePrep for 6 $\times$ 30 s with 30 s pauses in between and then centrifuged at 14,000 rpm for 10 min at room temperature. The supernatant was extracted twice more with 200 μL of TE, 200 μL of phenol-chloroform-isoamyl alcohol, and one 30 s agitation in the bead beater. DNA was precipitated using 80 μL of ammonium acetate (7.5 M) and 1 mL of 100% isopropanol, washed with 1 mL of 70% ethanol, and air-dried. The pellets were resuspended in 400 μL of TE with 1 μL of RNase A (100 mg/mL) and incubated overnight at 37°C. The DNA was precipitated again and resuspended in 150 μL water. Illumina sequencing was performed by the UCD Conway Genomics Core using a NextSeq 500. 1 ng of gDNA was tagmented (fragmented and tagged with adapter sequences) using the Nextera kit transposome. Dual-indexed paired-end libraries were prepared using the Nextera XT DNA Library Prep Kit. An Illumina NextSeq500 mid output 300 cycle sequencing kit was used to prepare and run the flowcell (HVGWJAFX2).

Oxford Nanopore sequencing. Strains were grown overnight in 50 mL of YPD broth, and genomic DNA was extracted from approximately 4×10^9 cells using a Qiagen Genomic-tip 100/G kit (10223, Qiagen) with minor modifications. The lyticase incubation was extended to 2 h, and the proteinase K incubation was extended to overnight ($\sim$15 h). DNA libraries were prepared using three different kits as per the manufacturers' instructions. Libraries from *C. parapsilosis* MSK812 and UCD321 were prepared

using a Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore), using 1 μg of DNA per strain. DNA was repaired using a NEBNext FFPE DNA Repair Mix (M6630, New England Biolabs) and NEBNext Ultra II End repair/dA-tailing Module (E7546, New England Biolabs). Adapters were ligated using a NEBNext Quick Ligation Module (E6056, New England Biolabs). Libraries were sequenced on a MinION Mk1C device. Priming, loading and washing were performed using the EXP-FLP002, SQK-LSK109, and EXP-WSH002 Oxford Nanopore kits, respectively, as per the manufacturers' instructions. The genomes were sequenced in parallel on one flow cell. The first isolate (*C. parapsilosis* MSK812) was sequenced for 24 h, and the flow cell was washed, reprimed, and loaded with the *C. parapsilosis* UCD321 isolate for an additional ~48 h. Libraries from *C. parapsilosis* MSK802 and MSK803 were prepared using a Rapid Barcode Sequencing Kit (SQK-RBK004, Oxford Nanopore), using 400 ng of DNA per sample. Samples were multiplexed using barcodes R04 and R05, respectively. Libraries were sequenced on an original MinION device for 72 h. Priming and loading were performed using the EXP-FLP002 and SQK-RBK004 kits, respectively. The *C. parapsilosis* MSK478 library was prepared using a Rapid Sequencing Kit (SQK-RAD004, Oxford Nanopore), using 400 ng of DNA, and sequenced for 72 h using an original MinION device. Basecalling was performed for all samples using guppy_basecaller with the following parameters: "–input_path fast5 –save_path fastq –c dna_r9.4.1_450bps_fast.cfg –verbose_logs –cpu_threads_per_caller 5 –num_callers 10". Guppy v.4.2.2+effbaf8 was used for the MSK812 and UCD321 samples, and Guppy v.3.6 was used for the MSK802 and MSK803 samples. For the multiplexed samples MSK802 and MSK803, demultiplexing was performed using guppy_barcoder with the following parameters: "–barcode_kits SQK-RBK004 –t 30 –verbose_logs –trim_barcodes". The reads were filtered using NanoFilt with the following parameters: "-l 1000 -q 7" (67).

Sequence analysis. The Illumina reads were trimmed with Skewer version 0.2.2 using tags "-m pe -t 4 -l 35 -q 30 -Q 30" (68). The trimmed reads were aligned to the *C. parapsilosis* reference genome using bwa-mem version 0.7.12. The resulting BAM files were sorted, and duplicate reads were marked using the GenomeAnalysisToolkit (GATK version 4.0.1.2) SortSam and MarkDuplicates tools, respectively. Variants were called using GATK HaplotypeCaller with the tag "–genotyping_mode DISCOVERY," combined using GATK CombineGVCFs, and joint-genotyped using GATK GenotypeGVCFs. Variant files were filtered for read depth (<15) and genotype quality (<40) using GATK VariantFiltration. Additionally, clusters of SNPs (5 SNPs in a 100 bp window) were filtered using GATK VariantFiltration. A custom script was used to remove variants that were flanked on either side by a long string of mononucleotide or dinucleotide repeats and by variants that were called as heterozygous but had an allele depth ratio <0.25 or >0.75 (https://github.com/CMOTsean/milt_variant_filtration). Additionally, for tree construction, indels were excluded using GATK SelectVariants with the tag "–select-type-to-include SNP". For the analysis of the evolved strains, a custom script was used to filter out variants in the evolved strains that were also present in the respective parent strain (https://github.com/CMOTsean/milt_variant_filtration).

SIFT4G analysis. A SIFT prediction database was created for *C. parapsilosis* using the SIFT4G algorithm and the recommended Uniref90 database as a reference for the protein sequences (69). The *C. parapsilosis* prediction database was used to annotate variants from the evolved strains with whether they are likely to be deleterious to protein function.

For each annotated gene in *C. parapsilosis*, the number of evolved strains which carried a variant predicted by SIFT to be protein function-affecting in that gene was tallied. Variants were also visualized using Integrative Genomics Viewer (IGV) to manually check results (70).

Phylogeny construction. Called SNPS were concatenated, and heterozygous sites were resolved randomly to either allele by 1,000 iterations of random repeated haplotype sampling (71). SNP trees were then constructed from each of the 1,000 haploid inputs using RAxML (v8.2.12) with the GTRGAMMA model of nucleotide substitution and the random number seed "-p 12345" (72). The tree with the highest maximum likelihood score was chosen, and the remaining 999 trees were used to generate branch support values.

Estimating copy numbers. Six strains were removed from the aneuploidy step because of uneven sequence coverage (full list used is in Table S2). The mean coverage of each chromosome (except the rDNA on Chromosome 7) for each of the remaining 163 strains was calculated using BEDTools and was divided by the average coverage of the genome to identify chromosome copy numbers. Chromosomes with copy numbers >2.5 were called as aneuploid.

CNVs were identified using DELLY (73) with default CNV length = 1,000 bp. CNVs within 20 kb of the telomeric regions, the rDNA, and the mitochondrial genome were removed (Table S2). Deletions which were over 1.25× coverage and duplications which were under 2.75× coverage were also removed. CNVs were merged if the start points or endpoints were within 1,000 bp. The lengths and coverages were averaged between all strains with the merged CNV (Table S2). Several strains with coverage problems were excluded from the merging step (FM05, FM06, FM07, FM10, FM14, FM32, FM43, 611, GA1_ERR246510, J931058, J931845, J950218, Kw1590-18, Kw2006-15, 103) because of uneven sequence coverage. For each strain with a CNV at *RTA3* or *ARR3*, the average coverage across the ORF in each isolate was found using BEDTools coverage (v2.29.2) (74). This value was divided by the average genome coverage (found with BEDTools genomecov) and multiplied by two to adjust for ploidy in order to calculate an estimate for the copy number. For CNVs that did not cover the entire *RTA3* ORF, the average coverage of a representative section of the CNV was used instead.

MinION read sequences were used to identify the exact *RTA3* copy number for a set of strains. The respective CNV sequences plus 1 kb flanking sequences on either side, were searched against the set of MinION reads for strains MSK478, MSK802, MSK803, MSK812, and UCD321 using BLASTN (v2.10.0). The search outputs were parsed with RECON-EBB (https://github.com/CMOTsean/recon-ebb) to estimate and visualize the copy number from reads which included hits for multiple copies of the CNV sequence

and both regions of the flanking sequence (i.e., the reads which covered the entirety of the repeat region). The same method was used to characterize the *ARR3* amplification in UCD321.

Replication timing profiles. Relative replication time was determined by SORT-seq as described previously (42). Briefly, replicating (S phase) and nonreplicating (G_2 phase) cells were enriched from an asynchronously growing culture by FACS based on DNA content. In each sample, genomic DNA was extracted and subjected to Illumina sequencing to measure the relative DNA copy number. Replication timing profiles were generated by normalizing the replicating (S phase) sample read count to the non-replicating (G_2) sample read count in 1 kb windows.

Quantitative RT-PCR. Cell harvesting and RNA extraction methods were adapted from Cravener and Mitchell (75). Cells were inoculated from overnight cultures to an A_{600} of 0.2 in 25 mL of prewarmed YPD broth and were incubated at 30°C for 6 h at 200 rpm using an orbital shaker. Cells were then harvested via vacuum filtration using MicroPlus-21 Sterile 0.45 μm filters (10407713, Whatman) and stored at $-80°C$ for at least 24 h prior to RNA extraction. Cells were lysed by mechanical disruption as recommended in the Qiagen RNeasy Minikit (74104, Qiagen) with some modifications (75). Cells were lysed in RLT lysis buffer (Qiagen) and phenol-chloroform-isoamyl alcohol (25:24:1) (P3803, Sigma) in a 1:1 ratio. Lysis was performed using a 1600 MiniG from Spex SamplePrep using 30 s lysis followed by 30 s of chilling on ice for a total of 6 min. The RNA was extracted using a Qiagen RNeasy Minikit followed by two rounds of DNase digestion: one on-column using the Qiagen RNase-Free DNase Set (79254, Qiagen) and one off-column using Invitrogen's TURBO DNA-free Kit (AM1907, Thermo Fisher). The cDNA was synthesized from 1 μg RNA using M-MLV reverse transcriptase (9PIM170, Promega) and Oligo(dT)15 primers (C110A, Promega), following the manufacturers' instructions. Quantitative PCR was performed in 20 μL reactions using 50 ng cDNA using FastStart Universal SYBR green Master (Rox) (4913850001, Sigma) as per the manufacturers' instructions on an Agilent Technologies Stratagene Mx3005p machine using default "two-step" settings. All primers are listed in Table S2. Relative quantification was performed using the 2(-Delta Delta C[T]) method by comparing the expression to *ACT1* and using *C. parapsilosis* CLIB214 as the calibrator strain. Calculations and statistics were performed in R using the pcr package (76).

Data availability. All sequencing data are deposited at NCBI under BioProject numbers PRJNA795920 and PRJNA748054 (SRP328964).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, PDF file, 0.5 MB.
FIG S2, PDF file, 0.3 MB.
FIG S3, PDF file, 0.5 MB.
FIG S4, PDF file, 1.4 MB.
FIG S5, PDF file, 0.4 MB.
TABLE S1, DOCX file, 0.1 MB.
TABLE S2, XLSX file, 0.2 MB.
TABLE S3, DOCX file, 0.01 MB.

REFERENCES

1. Katju V, Bergthorsson U. 2013. Copy-number changes in evolution: rates, fitness effects and adaptive significance. Front Genet 4:273. https://doi.org/10.3389/fgene.2013.00273.

2. Lauer S, Gresham D. 2019. An evolving view of copy number variants. Curr Genet 65:1287–1295. https://doi.org/10.1007/s00294-019-00980-0.

3. Mirzaei G, Petreaca RC. 2022. Distribution of copy number variations and rearrangement endpoints in human cancers with a review of literature. Mutat Res 824:111773. https://doi.org/10.1016/j.mrfmmm.2021.111773.

4. Welch JW, Fogel S, Cathala G, Karin M. 1983. Industrial yeasts display tandem gene iteration at the *CUP1* region. Mol Cell Biol 3:1353–1361. https://doi.org/10.1128/mcb.3.8.1353-1361.1983.

5. Adamo GM, Lotti M, Tamas MJ, Brocca S. 2012. Amplification of the *CUP1* gene is associated with evolution of copper tolerance in *Saccharomyces cerevisiae*. Microbiology (Reading) 158:2325–2335. https://doi.org/10.1099/mic.0.058024-0.

6. Fogel S, Welch JW. 1982. Tandem gene amplification mediates copper resistance in yeast. Proc Natl Acad Sci U S A 79:5342–5346. https://doi.org/10.1073/pnas.79.17.5342.

7. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet 4:e1000303. https://doi.org/10.1371/journal.pgen.1000303.

8. Gresham D, Usaite R, Germann SM, Lisby M, Botstein D, Regenberg B. 2010. Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the *GAP1* locus. Proc Natl Acad Sci U S A 107:18551–18556. https://doi.org/10.1073/pnas.1014023107.

9. Payen C, Di Rienzi SC, Ong GT, Pogachar JL, Sanchez JC, Sunshine AB, Raghuraman MK, Brewer BJ, Dunham MJ. 2014. The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. G3 (Bethesda) 4:399–409. https://doi.org/10.1534/g3.113.009365.

10. Todd RT, Selmecki A. 2020. Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. Elife 9:e58349. https://doi.org/10.7554/eLife.58349.

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

12. Selmecki A, Forche A, Berman J. 2010. Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryot Cell 9:991–1008. https://doi.org/10.1128/EC.00060-10.

13. Peter J, De Chiara M, Friedrich A, Yue JX, Pflieger D, Bergstrom A, Sigwalt A, Barre B, Freel K, Llored A, Cruaud C, Labadie K, Aury JM, Istace B, Lebrigand K, Barbry P, Engelen S, Lemainque A, Wincker P, Liti G, Schacherer J. 2018. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature 556:339–344. https://doi.org/10.1038/s41586-018-0030-5.

14. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, Zeng Q, Zisson E, Wang JM, Greenberg JM, Berman J, Bennett RJ, Cuomo CA. 2015. Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res 25:413–425. https://doi.org/10.1101/gr.174623.114.

15. Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, Agrafioti I, Arnaud MB, Bates S, Brown AJ, Brunke S, Costanzo MC, Fitzpatrick DA, de Groot PW, Harris D, Hoyer LL, Hube B, Klis FM, Kodira C, Lennard N, Logue ME, Martin R, Neiman AM, Nikolaou E, Quail MA, Quinn J, Santos MC, Schmitzberger FF, Sherlock G, Shah P, Silverstein KA, Skrzypek MS, Soll D, Staggs R, Stansfield I, Stumpf MP, Sudbery PE, Srikantha T, Zeng Q, Berman J, Berriman M, Heitman J, Gow NA, Lorenz MC, Birren BW, Kellis M, et al. 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida genomes*. Nature 459:657–662. https://doi.org/10.1038/nature08064.

16. Guo X, Zhang R, Li Y, Wang Z, Ishchuk OP, Ahmad KM, Wee J, Piskur J, Shapiro JA, Gu Z. 2020. Understand the genomic diversity and evolution of fungal pathogen *Candida glabrata* by genome-wide analysis of genetic variations. Methods 176:82–90. https://doi.org/10.1016/j.ymeth.2019.05.002.

17. Dunn B, Richter C, Kvitek DJ, Pugh T, Sherlock G. 2012. Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. Genome Res 22:908–924. https://doi.org/10.1101/gr.130310.111.

18. Anderson MZ, Baller JA, Dulmage K, Wigen L, Berman J. 2012. The three clades of the telomere-associated *TLO* gene family of *Candida albicans* have different splicing, localization, and expression features. Eukaryot Cell 11:1268–1275. https://doi.org/10.1128/EC.00230-12.

19. Gallone B, Steensels J, Prahl T, Soriaga L, Saels V, Herrera-Malaver B, Merlevede A, Roncoroni M, Voordeckers K, Miraglia L, Teiling C, Steffy B, Taylor M, Schwartz A, Richardson T, White C, Baele G, Maere S, Verstrepen KJ. 2016. Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. Cell 166:1397–1410e16. https://doi.org/10.1016/j.cell.2016.08.020.

20. Steenwyk J, Rokas A. 2017. Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. G3 (Bethesda) 7:1475–1485. https://doi.org/10.1534/g3.117.040105.

21. Warringer J, Zorgo E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt SW, Louis EJ, Liti G, Moses A, Blomberg A. 2011. Trait variation in yeast is defined by population history. PLoS Genet 7:e1002111. https://doi.org/10.1371/journal.pgen.1002111.

22. Zhao Y, Strope PK, Kozmin SG, McCusker JH, Dietrich FS, Kokoska RJ, Petes TD. 2014. Structures of naturally evolved *CUP1* tandem arrays in yeast indicate that these arrays are generated by unequal nonhomologous recombination. G3 (Bethesda) 4:2259–2269. https://doi.org/10.1534/g3.114.012922.

23. Crosato G, Nadai C, Carlot M, Garavaglia J, Ziegler DR, Rossi RC, De Castilhos J, Campanaro S, Treu L, Giacomini A, Corich V. 2020. The impact of *CUP1* gene copy-number and XVI-VIII/XV-XVI translocations on copper and sulfite tolerance in vineyard *Saccharomyces cerevisiae* strain populations. FEMS Yeast Res 20. https://doi.org/10.1093/femsyr/foaa028.

24. Hull RM, Cruz C, Jack CV, Houseley J. 2017. Environmental change drives accelerated adaptation through stimulated copy number variation. PLoS Biol 15:e2001333. https://doi.org/10.1371/journal.pbio.2001333.

25. Whale AJ, King M, Hull RM, Krueger F, Houseley J. 2022. Stimulation of adaptive gene amplification by origin firing under replication fork constraint. Nucleic Acids Res 50:915–936. https://doi.org/10.1093/nar/gkab1257.

26. Gu W, Zhang F, Lupski JR. 2008. Mechanisms for human genomic rearrangements. Pathogenetics 1:4. https://doi.org/10.1186/1755-8417-1-4.

27. Croll D, Zala M, McDonald BA. 2013. Breakage-fusion-bridge cycles and large insertions contribute to the rapid evolution of accessory chromosomes in a fungal pathogen. PLoS Genet 9:e1003567. https://doi.org/10.1371/journal.pgen.1003567.

28. Kramara J, Osia B, Malkova A. 2018. Break-induced replication: the where, the why, and the how. Trends Genet 34:518–531. https://doi.org/10.1016/j.tig.2018.04.002.

29. Hastings PJ, Ira G, Lupski JR. 2009. A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet 5:e1000327. https://doi.org/10.1371/journal.pgen.1000327.

30. Carvalho CMB, Zhang F, Liu P, Patel A, Sahoo T, Bacino CA, Shaw C, Peacock S, Pursley A, Tavyev YJ, Ramocki MB, Nawara M, Obersztyn E, Vianna-Morgante AM, Stankiewicz P, Zoghbi HY, Cheung SW, Lupski JR. 2009. Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching. Hum Mol Genet 18:2188–2203. https://doi.org/10.1093/hmg/ddp151.

31. Brewer BJ, Payen C, Di Rienzi SC, Higgins MM, Ong G, Dunham MJ, Raghuraman M. 2015. Origin-dependent inverted-repeat amplification: tests of a model for inverted DNA amplification. PLoS Genet 11:e1005699. https://doi.org/10.1371/journal.pgen.1005699.

32. Pryszcz LP, Nemeth T, Gacser A, Gabaldon T. 2013. Unexpected genomic variability in clinical and environmental strains of the pathogenic yeast *Candida parapsilosis*. Genome Biol Evol 5:2382–2392. https://doi.org/10.1093/gbe/evt185.

33. Zhai B, Ola M, Rolling T, Tosini NL, Joshowitz S, Littmann ER, Amoretti LA, Fontana E, Wright RJ, Miranda E, Veelken CA, Morjaria SM, Peled JU, van den Brink MRM, Babady NE, Butler G, Taur Y, Hohl TM. 2020. High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. Nat Med 26:59–64. https://doi.org/10.1038/s41591-019-0709-7.

34. Kihara A, Igarashi Y. 2002. Identification and characterization of a *Saccharomyces cerevisiae* gene, *RSB1*, involved in sphingoid long-chain base release. J Biol Chem 277:30048–30054. https://doi.org/10.1074/jbc.M203385200.

35. Kihara A, Igarashi Y. 2004. Cross talk between sphingolipids and glycerophospholipids in the establishment of plasma membrane asymmetry. Mol Biol Cell 15:4949–4959. https://doi.org/10.1091/mbc.e04-06-0458.

36. Manente M, Ghislain M. 2009. The lipid-translocating exporter family and membrane phospholipid homeostasis in yeast. FEMS Yeast Res 9:673–687. https://doi.org/10.1111/j.1567-1364.2009.00513.x.

37. Johnson SS, Hanson PK, Manoharlal R, Brice SE, Cowart LA, Moye-Rowley WS. 2010. Regulation of yeast nutrient permease endocytosis by ATP-binding cassette transporters and a seven-transmembrane protein, RSB1. J Biol Chem 285:35792–35802. https://doi.org/10.1074/jbc.M110.162883.

38. Srivastava A, Sircaik S, Husain F, Thomas E, Ror S, Rastogi S, Alim D, Bapat P, Andes DR, Nobile CJ, Panwar SL. 2017. Distinct roles of the 7-transmembrane receptor protein Rta3 in regulating the asymmetric distribution of phosphatidylcholine across the plasma membrane and biofilm formation in *Candida albicans*. Cell Microbiol 19. https://doi.org/10.1111/cmi.12767.

39. Iacano AJ, Lewis H, Hazen JE, Andro H, Smith JD, Gulshan K. 2019. Miltefosine increases macrophage cholesterol release and inhibits NLRP3-inflammasome

assembly and IL-1β release. Sci Rep 9:1–12. https://doi.org/10.1038/s41598-019-47610-w.

40. Asadzadeh M, Dashti M, Ahmad S, Alfouzan W, Alameer A. 2021. Whole-genome and targeted-amplicon sequencing of fluconazole-susceptible and -resistant *Candida parapsilosis* isolates from Kuwait reveals a previously undescribed N1132D polymorphism in CDR1. Antimicrob Agents Chemother 65:e01633-20. https://doi.org/10.1128/AAC.01633-20.

41. Brown CA, Murray AW, Verstrepen KJ. 2010. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr Biol 20:895–903. https://doi.org/10.1016/j.cub.2010.04.027.

42. Batrakou DG, Müller CA, Wilson RHC, Nieduszynski CA. 2020. DNA copy-number measurement of genome replication dynamics by high-throughput sequencing: the sort-seq, sync-seq and MFA-seq family. Nat Protoc 15:1255–1284. https://doi.org/10.1038/s41596-019-0287-7.

43. Wysocki R, Bobrowicz P, Ułaszewski S. 1997. The *Saccharomyces cerevisiae* ACR3 gene encodes a putative membrane protein involved in arsenite transport. J Biol Chem 272:30061–30066. https://doi.org/10.1074/jbc.272.48.30061.

44. Opulente DA, Langdon QK, Buh KV, Haase MAB, Sylvester K, Moriarty RV, Jarzyna M, Considine SL, Schneider RM, Hittinger CT. 2019. Pathogenic budding yeasts isolated outside of clinical settings. FEMS Yeast Res 19. https://doi.org/10.1093/femsyr/foz032.

45. Bergstrom A, Simpson JT, Salinas F, Barre B, Parts L, Zia A, Nguyen Ba AN, Moses AM, Louis EJ, Mustonen V, Warringer J, Durbin R, Liti G. 2014. A high-definition view of functional genetic variation from natural yeast genomes. Mol Biol Evol 31:872–888. https://doi.org/10.1093/molbev/msu037.

46. Chow EW, Morrow CA, Djordjevic JT, Wood IA, Fraser JA. 2012. Microevolution of *Cryptococcus neoformans* driven by massive tandem gene amplification. Mol Biol Evol 29:1987–2000. https://doi.org/10.1093/molbev/mss066.

47. West PT, Peters SL, Olm MR, Yu FB, Gause H, Lou YC, Firek BA, Baker R, Johnson AD, Morowitz MJ, Hettich RL, Banfield JF. 2021. Genetic and behavioral adaptation of *Candida parapsilosis* to the microbiome of hospitalized infants revealed by *in situ* genomics, transcriptomics, and proteomics. Microbiome 9:1–17. https://doi.org/10.1186/s40168-021-01085-y.

48. Moore IK, Martin MP, Dorsey MJ, Paquin CE. 2000. Formation of circular amplifications in *Saccharomyces cerevisiae* by a breakage-fusion-bridge mechanism. Environ Mol Mutagen 36:113–120. https://doi.org/10.1002/1098-2280(2000)36:2%3C113::AID-EM5%3E3.0.CO;2-T.

49. Müller CA, Hawkins M, Retkute R, Malla S, Wilson R, Blythe MJ, Nakato R, Komata M, Shirahige K, de Moura APS, Nieduszynski CA. 2014. The dynamics of genome replication using deep sequencing. Nucleic Acids Res 42:e3. https://doi.org/10.1093/nar/gkt878.

50. Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, Sertour N, Mosca K, Permal E, Laval G, Bouchier C, Ma L, Schwartz K, Voelz K, May RC, Poulain J, Battail C, Wincker P, Borman AM, Chowdhary A, Fan S, Kim SH, Le Pape P, Romeo O, Shin JH, Gabaldón T, Sherlock G, Bougnoux M-E, d'Enfert C. 2018. Gene flow contributes to diversification of the major fungal pathogen *Candida albicans*. Nat Commun 9:2253. https://doi.org/10.1038/s41467-018-04787-4.

51. Sitterlé E, Maufrais C, Sertour N, Palayret M, d'Enfert C, Bougnoux ME. 2019. Within-host genomic diversity of *Candida albicans* in healthy carriers. Sci Rep 9:2563. https://doi.org/10.1038/s41598-019-38768-4.

52. 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. https://doi.org/10.1371/journal.pgen.1006404.

53. 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 6:1069–1078. https://doi.org/10.1093/gbe/evu082.

54. Whaley SG, Tsao S, Weber S, Zhang Q, Barker KS, Raymond M, Rogers PD. 2016. The RTA3 gene, encoding a putative lipid translocase, influences the susceptibility of *Candida albicans* to fluconazole. Antimicrob Agents Chemother 60:6060–6066. https://doi.org/10.1128/AAC.00732-16.

55. Lombardi L, Oliveira-Pacheco J, Butler G. 2019. Plasmid-based CRISPR-Cas9 gene editing in multiple *Candida* species. mSphere 4:e00125-19. https://doi.org/10.1128/mSphere.00125-19.

56. Hull RM, King M, Pizza G, Krueger F, Vergara X, Houseley J. 2019. Transcription-induced formation of extrachromosomal DNA during yeast ageing. PLoS Biol 17:e3000471. https://doi.org/10.1371/journal.pbio.3000471.

57. Sakofsky CJ, Ayyar S, Deem AK, Chung W-H, Ira G, Malkova A. 2015. Translesion polymerases drive microhomology-mediated break-induced replication leading to complex chromosomal rearrangements. Mol Cell 60:860–872. https://doi.org/10.1016/j.molcel.2015.10.041.

58. Thomas E, Sircaik S, Roman E, Brunel JM, Johri AK, Pla J, Panwar SL. 2015. The activity of RTA2, a downstream effector of the calcineurin pathway, is required during tunicamycin-induced ER stress response in *Candida albicans*. FEMS Yeast Res 15.

59. Yang F, Gritsenko V, Slor Futterman Y, Gao L, Zhen C, Lu H, Jiang Y-y, Berman J. 2021. Tunicamycin potentiates antifungal drug tolerance via aneuploidy in *Candida albicans*. mBio 12:e02272-21. https://doi.org/10.1128/mBio.02272-21.

60. Makuta H, Obara K, Kihara A. 2017. Loop 5 region is important for the activity of the long-chain base transporter Rsb1. J Biochem 161:207–213. https://doi.org/10.1093/jb/mvw059.

61. Barreto TL, Rossato L, de Freitas ALD, Meis JF, Lopes LB, Colombo AL, Ishida K. 2020. Miltefosine as an alternative strategy in the treatment of the emerging fungus *Candida auris*. Int J Antimicrob Agents 56:106049. https://doi.org/10.1016/j.ijantimicag.2020.106049.

62. Wu Y, Wu M, Gao J, Ying C. 2020. Antifungal activity and mode of action of miltefosine against clinical isolates of *Candida krusei*. Front Microbiol 11:854. https://doi.org/10.3389/fmicb.2020.00854.

63. Vila T, Ishida K, Seabra SH, Rozental S. 2016. Miltefosine inhibits *Candida albicans* and non-*albicans Candida* spp. biofilms and impairs the dispersion of infectious cells. Int J Antimicrob Agents 48:512–520. https://doi.org/10.1016/j.ijantimicag.2016.07.022.

64. Mondelaers A, Sanchez-Cañete MP, Hendrickx S, Eberhardt E, Garcia-Hernandez R, Lachaud L, Cotton J, Sanders M, Cuypers B, Imamura H, Dujardin J-C, Delputte P, Cos P, Caljon G, Gamarro F, Castanys S, Maes L. 2016. Genomic and molecular characterization of miltefosine resistance in *Leishmania infantum* strains with either natural or acquired resistance through experimental selection of intracellular amastigotes. PLoS One 11:e0154101. https://doi.org/10.1371/journal.pone.0154101.

65. Papp C, Kocsis K, Tóth R, Bodai L, Willis JR, Ksiezopolska E, Lozoya-Pérez NE, Vágvölgyi C, Mora Montes H, Gabaldón T, Nosanchuk JD, Gácser A. 2018. Echinocandin-induced microevolution of *Candida parapsilosis* influences virulence and abiotic stress tolerance. Msphere 3:e00547-18. https://doi.org/10.1128/mSphere.00547-18.

66. Ene IV, Farrer RA, Hirakawa MP, Agwamba K, Cuomo CA, Bennett RJ. 2018. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. Proc Natl Acad Sci U S A 115:E8688–E8697. https://doi.org/10.1073/pnas.1806002115.

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. https://doi.org/10.1093/bioinformatics/bty149.

68. Jiang H, Lei R, Ding S-W, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics 15:182–112. https://doi.org/10.1186/1471-2105-15-182.

69. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. 2016. SIFT missense predictions for genomes. Nat Protoc 11:1–9. https://doi.org/10.1038/nprot.2015.123.

70. Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. https://doi.org/10.1093/bib/bbs017.

71. Lischer HE, Excoffier L, Heckel G. 2014. Ignoring heterozygous sites biases phylogenomic estimates of divergence times: implications for the evolutionary history of *Microtus* voles. Mol Biol Evol 31:817–831. https://doi.org/10.1093/molbev/mst271.

72. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

73. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28:i333–i339. https://doi.org/10.1093/bioinformatics/bts378.

74. Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842. https://doi.org/10.1093/bioinformatics/btq033.

75. Cravener MV, Mitchell AP. 2020. *Candida albicans* culture, cell harvesting, and total RNA extraction. Bio Protoc 10:e3803.

76. Ahmed M, Kim DR. 2018. pcr: an R package for quality assessment, analysis and testing of qPCR data. PeerJ 6:e4473. https://doi.org/10.7717/peerj.4473.

77. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110.

78. Lopez-Delisle L, Rabbani L, Wolff J, Bhardwaj V, Backofen R, Grüning B, Ramírez F, Manke T. 2021. pyGenomeTracks: reproducible plots for multivariate genomic datasets. Bioinformatics 37:422–423. https://doi.org/10.1093/bioinformatics/btaa692.

79. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580. https://doi.org/10.1006/jmbi.2000.4315.

80. 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. https://doi.org/10.1093/molbev/msp259.

81. Van der Mark VA, Elferink RP, Paulusma CC. 2013. P4 ATPases: flippases in health and disease. Int J Mol Sci 14:7897–7922. https://doi.org/10.3390/ijms14047897.