Genomic basis of multidrug-resistance, mating, and virulence in *Candida auris* and related emerging species

José F. Muñoz[1], Lalitha Gade[2], Nancy A. Chow[2], Vladimir N. Loparev[3], Phalasy Juieng[3], Rhys A. Farrer[1], Anastasia P. Litvintseva[2]*, Christina A. Cuomo[1]*

[1] Broad Institute of MIT and Harvard, Cambridge, MA USA. [2] Mycotic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, GA USA. [3] Biotechnology Core Facility Branch, Centers for Disease Control and Prevention, Atlanta, GA USA.

*Corresponding authors: Christina A. Cuomo cuomo@broadinstitute.org; Anastasia P. Litvintseva, frq8@cdc.gov

Abstract

Candida auris is an emergent fungal pathogen of rising health concern due to increasing reports of outbreaks in healthcare settings and resistance to multiple classes of antifungal drugs. While distantly related to the more common pathogens *C. albicans* and *C. glabrata*, *C. auris* is closely related to three rarely observed and often multidrug-resistant species, *C. haemulonii, C. duobushaemulonii and C. pseudohaemulonii*. Here, we generated and analyzed near complete genome assemblies and RNA-Seq-guided gene predictions for each of the four major *C. auris* clades and for *C. haemulonii, C. duobushaemulonii and C. pseudohaemulonii*. Our analyses mapped seven chromosomes and revealed chromosomal rearrangements between *C. auris* clades and related species. We found conservation of genes involved in mating and meiosis and identified both *MTL***a** and *MTL*α *C. auris* isolates, strongly suggesting the potential for mating between clades. Gene conservation analysis highlighted that many genes linked to drug resistance and virulence in other pathogenic *Candida* species are conserved in *C. auris* and related species including expanded families of transporters and lipases, as well as mutations and copy number variants in *ERG11* that confer drug resistance. In addition, we found aspects of this emerging clade that likely mediate differences in virulence and drug response, including different cell wall families. To begin to characterize the species-specific genes important for antifungal response, we profiled the gene expression of *C. auris* in response to voriconazole and amphotericin B and found induction of several transporters and metabolic regulators that may play a role in drug resistance. Together, this study provides a comprehensive view of the genomic basis of drug resistance, potential for mating, and virulence in this emerging fungal clade.

Introduction

Candida auris is an emerging fungal pathogen of increasing concern due to high drug resistance and high mortality rates[1,2]. In addition, outbreaks have been reported in hospital settings, suggesting healthcare transmission[2,3]. *C. auris* clinical isolates are typically multidrug-resistant (MDR), with common resistance to fluconazole and variable susceptibility to other azoles, amphotericin B, and echinocandins[2]. *C. auris* causes both bloodstream and invasive infections, similar to a group of rarely observed, phylogenetically related species including *C. haemulonii, C. duobushaemulonii and C. pseudohaemulonii*[4,5]. These species also display MDR, most commonly to amphotericin B and also reduced susceptibility to azoles and echinocandins[4,5]. Together, *C. auris* and these closely related species represent an emerging clade of invasive fungal infections, which are not only difficult to treat, but also difficult to identify using standard laboratory methods[6], generally requiring molecular methods for the proper identification.

Initial whole genome analysis of *C. auris* isolates from Pakistan, India, South Africa, Japan and Venezuela identified four clades that are specific to each geographic region and suggested that each of these clades emerged nearly simultaneously in different regions of the world[2]. Including data from other recent studies, the current representation of each clade is as follows: clade I comprises isolates from India, Pakistan and England[2,3,7], clade II from Japan and South Korea[2,8], clade III from South Africa, and clade IV from Venezuela[2]. Analyses of SNPs identified from whole genome sequence[2,9] and of multilocus sequence typing (MLST)[10,11] found very low genetic variation within *C. auris* clades. Little is known about whether phenotypic differences exist between *C. auris* clades; one noted difference is that isolates from India and South Africa assimilated N-acetylglucosamine in contrast to isolates from Japan and South Korea[2].

Candida auris is distantly related to the more commonly observed species *C. albicans* and *C. glabrata*, and phylogenetic studies supported that the *C. haemulonii* clade is the closest sibling group to *C. auris*[4,5,12,13]. However, these studies have not clearly resolved the relationships of *C. haemulonii, C. duobushaemulonii and C. pseudohaemulonii*[4,5,12,13]. *C. lusitaniae*, a rarely observed cause of infection[14], is a sister clade to this group of emerging species[13,15]. Despite the fact that these genotypic studies found that *C. auris* is highly divergent from other Saccharomycetales yeasts from the CTG clade, including *C. albicans,* most of our limited current knowledge of *C. auris* resistance and virulence had been inferred based on conservation of genes associated with drug resistant and virulence in *C.*

65 *albicans* or *C. glabrata*. Initial comparative analysis of the gene content of one isolate of *C. auris* with *C.*
66 *albicans* found that some orthologs associated with antifungal resistance are present in *C. auris*,
67 including drug transporters, secreted proteases and manosyl transferases[13]. In addition, preliminary
68 genomic studies showed that the targets of several classes of antifungal drugs are conserved in *C.*
69 *auris*, including the azole target lanosterol 14 α-demethylase (*ERG11*), the echinocandin target 1,3-
70 beta-glucan synthase (*FKS1*), and the flucytosine target uracil phosphoribosyl-transferase (*FUR1*)[2,16].
71 Furthermore, point mutations associated with drug resistance in other species are observed in many
72 clinical isolates, with some differences between *C. auris* clades[2,7,16]. However, there have not yet been
73 studies to directly test whether these mutations confer drug resistance in *C. auris* or the role of other
74 genes in drug resistance and virulence in this group of emerging MDR species.
75
76 While efforts to sequence *C. auris* genome provided an initial view of genome content[9,13], available
77 assemblies in GenBank are highly fragmented, inconsistently annotated, and do not provide a complete
78 representation of all *C. auris* clades. The related species *C. haemulonii* and *C. duobushaemulonii* were
79 recently sequenced [17,18]. Here, we generated and annotated highly complete genome assemblies for
80 each of the clades of *C. auris* as well as for the related species *C. haemulonii*, *C. duobushaemulonii*
81 and *C. pseudohaemulonii*. Comparison of these genomes to other sequenced *Candida* species
82 revealed that *C. auris* has notable expansions of genes linked to drug resistance and virulence in *C.*
83 *albicans*, including families of oligopeptide transporters, siderophore-based iron transporters, and
84 secreted lipases. We examined the response to antifungal drugs for two isolates using RNA-Seq and
85 identified transporters and metabolic regulators that have been previously associated with drug
86 resistance in *C. albicans*, but also some specific expanded or unique genes in emergent MDR species.
87 We also found strong evidence that *C. auris* is capable of mating and meiosis, based on the
88 identification of both mating types in the populations, conservation of genes involved in mating and
89 meiosis, and detection of chromosomal rearrangements between two clades. These results revealed
90 fundamental insights into the evolution of drug resistance and pathogenesis in *C. auris* and closely
91 related species and the potential for mating and recombination between the *C. auris* clades.
92
93 **Results**
94 **Genome characteristics of *Candida auris* and closely related species**
95 We generated highly complete genome assemblies for four *Candida auris* isolates and for three
96 phylogenetically related species, *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii*.
97 These four *C. auris* assemblies represent each of the major clades[2], including an updated assembly for

98 B8441 (clade I) and the first representatives for clade II (strain B11220), clade III (strain B11221) and
99 clade IV (strain B11243). As *C. auris* is distantly related to other previously sequenced *Candida*
100 species, we also sequenced genomes of the closely related species *C. haemulonii* (strain B11899), *C.*
101 *duobushaemulonii* (strain B09383), and *C. pseudohaemulonii* (strain B12108) to enable comparative
102 genomic analysis. The genome assemblies of *C. auris* (B8441 and B11221), *C. haemulonii*, and *C.*
103 *duobushaemulonii* were sequenced using PacBio and Illumina, whereas the other two *C. auris* strains
104 were sequenced only with Illumina (**Table 1**; **Methods**). The genome assemblies of *C. auris* range from
105 12.1 Mb in B1120 to 12.7 Mb in B1121. The *C. auris* B8441 and B11221 genome assemblies were
106 organized in 15 and 20 scaffolds, respectively, of which 7 scaffolds included most of the sequenced
107 bases in both strains (**Table 1**). This represents an improvement upon the previously generated
108 genome assembly of *C. auris* strain 6684 (clade I)[13], which consists of 99 scaffolds (759 contigs) that
109 include contig gaps and has an inflated number of predicted genes, based on our analysis (see below;
110 **Methods**). The Illumina only genome assemblies of *C. auris* B11220 (clade II) and B11243 (clade IV)
111 are less contiguous (**Table 1**). The genome assemblies of *C. haemulonii*, *C. duobushaemulonii,* and *C.*
112 *pseudohaemulonii* were also highly contiguous with 11, 7 and 36 scaffolds, respectively (**Table 1**).
113 Comparison of these four *C. auris* genomes and those of the three related species revealed that the
114 genome sizes are very similar, ranging in size from 12.1 Mb in *C. auris* B11220 to 13.3 Mb in *C.*
115 *haemulonii*, similar to that of *C. lusitaniae* (12.1 Mb) and other *Candida* species (**Table S1**). The
116 assemblies of the *C. auris* clades are highly identical, with an average nucleotide identity of 98.7%
117 among clade I, II, III and IV, and 99.3% between clade II and III. By contrast, comparison of nucleotide
118 genome identity between species highlights greater intraspecies genetic divergence: 88% between *C.*
119 *auris* and *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii*, 89% between *C. haemulonii*
120 and *C. duobushaemulonii*, and 92% between *C. duobushaemulonii* and *C. pseudohaemulonii*. We
121 found a very low frequency of candidate heterozygous positions predicted from Illumina data,
122 supporting that all sequenced genomes are haploid.
123
124 As an independent assessment of the genome assembly size and structure, we generated optical maps
125 of the *C. auris* B8441 and B11221 isolates (**Figure S1**). Consistent with the assemblies of these
126 isolates, the maps had seven linkage groups; nearly all of the genome assemblies were anchored to
127 the optical maps (98.8 % in B8441 and 97.2 % of in B11221; **Figure S1; Table 1**). This supports the
128 presence of 7 chromosomes in *C. auris*, consistent with the chromosome number found in previous
129 studies using electrophoretic karyotyping by pulsed-field gel electrophoresis (PFGE)[19]. While the
130 genomes of *C. auris* are highly syntenic, we found evidence of a few large chromosomal

131 rearrangements between *C. auris* B8441 and B11221 based on comparison of the assemblies and the
132 optical maps (**Figures 1**, **S1**). We confirmed that the junctions of these rearrangements are well
133 supported in each assembly; these regions show no variation in the depth of PacBio and Illumina
134 aligned reads and there is no evidence of assembly errors across these rearrangement breakpoints.
135 We additionally independently identified structural variants based on the read alignments to the B8441
136 and B11221 assemblies and recovered each of the rearrangements present in these assemblies.
137 These large chromosomal rearrangements included one inversion of 136 kb between B8441 sc01 and
138 B11221 sc01, a 274 kb translocation between B8441 sc08 and B11221 sc03, and a 300 kb
139 translocation between B8441 sc10 and B11221 sc01 (**Figure 1**). The genomes of *C. auris*, *C.*
140 *haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii* showed limited chromosomal
141 rearrangements between each other, mostly intra-chromosomal inversions between *C. auris* and *C.*
142 *haemulonii, C. duobushaemulonii,* and *C. pseudohaemulonii,* and large chromosomal translocations
143 between *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii* (**Figure 1a**). These
144 rearrangements between *C. auris* clades, and between species, could potentially prevent genetic
145 exchange between these groups, since some crossover events will generate missing chromosomal
146 regions or other aneuploidies and may result in nonviable progeny.
147

148 **Evolution of mating-type locus in the emerging multidrug-resistant species including *C. auris***

149 We characterized the mating type locus in *C. auris* and identified representatives of both mating
150 types. The mating locus structure is highly conserved compared to closely related species including *C.*
151 *lusitaniae* and other Saccharomycetales yeasts from the CTG clade *Candida* (**Figure 2**). Many *Candida*
152 species, including diploid asexual species, have heterothallic *MTL* idiomorphs, and mating occurs
153 between cells of opposite mating type, *MTL***a** and *MTL*α[20]. We found that the genes flanking the *MTL*
154 locus in some species from the CTG clade *Candida*, namely the phosphatidylinositol kinase gene
155 (*PIK1*), the oxysterol binding protein gene (*OBP1*) and the poly(A) polymerase gene (*PAP1*) were
156 adjacent in all sequenced genomes of the emerging MDR clade (**Figure 2a**). Previous work reported
157 that the genome of *C. auris* 6684 included the *MTL* flanking genes but did not identify either *MTL***a** or
158 *MTL*α genes at this locus[13]. In the chromosome level genome assemblies of *C. auris* (B8441 and
159 B11221), the *PIK1/OBP1/PAP1* genes are present at a single locus in chromosome 3 in B8441 (sc05)
160 and B11221 (sc03). Notably, we found all *C. auris* isolates contained either the *MTL***a** and *MTL*α
161 idiomorphs at this locus and that gene order was conserved compared to the *C. lusitaniae MTL*[21]
162 (**Figure 2a**; **Table S3**). We found that the *MTL***a** is present in *C. auris* B8441 and 6684 (clade I),

163 B11243 (clade IV), and *C. pseudohaemulonii*, spanning 14.9 kb (**Figure 2a**). By contrast, *C. auris*
164 B11220 (clade II), B11221 (clade III), and *C. haemulonii* and *C. duobushaemulonii* contain *MTLα*,
165 spanning 14.3 kb (**Figure 2a**). Phylogenetic analysis of the non-mating flanking genes
166 (*PIK1/OBP1/PAP1*) supports the inheritance of idiomorphs of these genes with the *MTL**a*** and *MTLα*
167 genes (**Figure 2b**). Upon further examination and manual annotation, we determined that **a**1/**a**2 are
168 present in *MTL***a** isolates, and α1 is present in *MTLα* isolates; as in *C. lusitaniae*, the MTL α locus is
169 missing the α2 gene[21] (**Figure 2a**). RNA-Seq data was used to guide gene prediction of **a**1 and **a**2, and
170 further establishes that both genes are expressed in B8441 (**Figure S2**), supporting the hypothesis that
171 these genes could be functional, and that the MTL locus could be used to classify *C. auris* isolates. To
172 further characterize the evolution of mating type in the population, we examined the *MTL* locus using 50
173 isolates from Lockhart *et al.*[2] and found that all isolates from clade I and IV have *MTL***a**, whereas all
174 isolates from clade II and III had *MTLα* (**Figures 2c**, **S3**). The fact both **a** and α MAT alleles are present
175 and expressed in *C. auris* suggests this species may be capable of mating, similar to that in *C.*
176 *lusitaniae*[21].
177
178 We examined the conservation of genes involved in meiosis to provide additional support for mating in
179 *C. auris*. We found that many of the key meiotic genes are similarly conserved between *C. lusitaniae*,
180 *C. guilliermondii*, *C. auris*, *C. haemulonii*, *C. duobushaemulonii* and *C. pseudohaemulonii*. Some genes
181 involved in meiosis in *S. cerevisiae* that are absent in these species included the recombinase *DMC1*
182 and cofactors (*MEI5* and *SAE3*), synaptonemal-complex proteins (*ZIP1* and *HOP1*), and genes
183 involved in crossover interference (*MSH4* and *MSH5*; **Table S2**). A small number of genes involved in
184 meiosis were present in *C. lusitaniae* but absent in *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and
185 *C. pseudohaemulonii*. In *C. auris*, the DNA recombination and repair genes *RAD55* and *RAD57* appear
186 to be absent, however *RAD55* is widely absent in the CTG *Candida* clade while the *RAD51* paralog of
187 the *RAD55*/*RAD57* complex is present. In *C. lusitaniae*, despite the loss of many of the same meiotic
188 genes, undergoes meiosis during sexual reproduction involving diploid intermediates[21]. The fact that
189 most components of the mating and meiosis pathways are similarly conserved in *C. auris* and closely
190 related species including *C. lusitaniae* suggests these species have the ability to mate and undergo
191 meiosis as observed for *C. lusitaniae*.
192
193 **Phylogenetic position of *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C.***
194 ***pseudohaemulonii***

195 Using the complete genomes, we estimated a strongly supported phylogeny of *C. auris*, *C.*
196 *haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii*, relative to other species from the order
197 Saccharomycetales including *C. lusitaniae*, *C. tropicalis*, *C. albicans* and *C. glabrata*. Based on a
198 concatenated alignment of 2,505 single copy core genes, a well supported maximum likelihood tree
199 placed *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii* as a single clade,
200 confirming the close relationship of these species (100% of bootstrap replicates; **Figure 3a**). The *C.*
201 *auris* clades appear more recently diverged based on short branch lengths. Previous phylogenetic
202 analyses had shown conflicting relationships between *C. auris*, *C. haemulonii*, *C. duobushaemulonii,*
203 and *C. pseudohaemulonii*[4,5,12,13]. Our phylogenetic analysis strongly supports that *C. duobushaemulonii*
204 and *C. pseudohaemulonii* are most closely related to each other, and form a sister group to *C.*
205 *haemulonii*, which appeared as the more basally branching species (**Figure 3a**). The closely related
206 species to this MDR clade is *C. lusitaniae*, which is the more basally branching member of this group
207 (**Figure 3a**).
208
209 **Gene family expansions supported mechanisms of drug resistance and virulence**
210 Gene annotation in *C. auris* genomes was performed using RNA-Seq paired-end reads to
211 improve gene structure predictions (**Methods**). The predicted gene number was highly similar across
212 all *C. auris* genomes as well as in *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii*. In *C.*
213 *auris*, the number of protein-coding genes varied between 5,421 in B8441 and 5,601 in B11243. For *C.*
214 *haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii* the numbers were very similar, ranging
215 from 5,288 to 5,410, predicted genes (**Table 1**; **Figure S4a**). High representation of core eukaryotic
216 genes provides evidence that those genomes are nearly complete; 96-98% of these conserved genes
217 are found in all annotated genome assemblies (**Figure S4b**). By examining orthologous genes in *C.*
218 *auris*, *C. haemulonii*, *C. duobushaemulonii, C. pseudohaemulonii*, and twelve additional
219 Saccharomycetales genomes, including *C. lusitaniae*, *C. tropicalis*, *C. albicans* and *C. glabrata*, we
220 found a total of 2,379 core ortholog clusters had representative genes from all twenty analyzed
221 genomes (**Figure S4**). We found a small number of unique genes in the *C. auris* clades ranging from
222 15 (B8441; clade I) to 54 (B11221; clade III) genes, 203 in *C. haemulonii*, 83 in *C. duobushaemulonii,*
223 and 88 *C. pseudohaemulonii* (**Figure S5**). The unique genes in *C. auris* clades include oligopeptide and
224 ABC transporters; unique glycophosphatidylinositol (GPI)-anchored proteins were identified in *C.*
225 *haemulonii* (**Table S3**).
226
227 To characterize changes in gene content that may play a role in the evolution of multidrug-resistance

228 and virulence in *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii*, we searched
229 for expansions or contractions in functionally classified genes compared to other related species (**Table**
230 **S1**). We identified PFAM domains that were significantly enriched or depleted (**Methods**; **Figures S4**).
231 Domains associated with transmembrane transporters (OPT, MFS) and secreted lipases (LIP) were
232 enriched in *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii* compared to other
233 genomes (q-value < 0.05, Fisher's exact test; **Figure 3b**). We therefore further classified
234 transmembrane transporters using the Transporter Classification Database (TCDB) and found that the
235 higher copy number of transporters in the emergent MDR clade could be attributed to oligopeptide
236 transporters (*OPT*) and siderophore iron transporters (*SIT*). In *C. albicans*, a family of *OPT* transporters
237 enables uptake of small peptides; the expression of some *OPT* transporters are up-regulated by azole
238 drugs[22,23]. While most of the fourteen *OPT* genes found in *C. auris* had orthologs in *C. albicans* (*OPT1-*
239 *8*), a subset of these transporters appear recently duplicated in the MDR emergent clade, including an
240 expansion of three *OPT1*-like transporters, and five transporters similar to *OPT2*, *OPT3*, and *OPT4*
241 (**Figures 4a**, **S6**). Most of the *OPT* genes (up to 8) are located in a conserved locus among emerging
242 MDR species encompassing 296 kb of chromosome 6; this gene family appears to have expanded by
243 tandem duplication (**Figures 4c**). In *C. albicans*, iron transporters include the siderophore transporter
244 *SIT1* and the iron permeases *FTR1* and *FTR2*; a subset of these transporters is uniquely expanded in
245 *C. auris* and closely related species, including the expansion of fourteen ortholog groups in *C. auris*
246 related to *C. albicans SIT1* (**Figures 4b**, **S6**). Secreted lipases are also expanded in the genomes of *C.*
247 *auris* and closely related species (q-value < 0.05, Fisher's exact test; **Figure 3b**). The *C. auris* clade
248 has similar counts of lipases relative to *C. albicans* and *C. dubliniensis*, however, these proteins are
249 expanded relative to more closely related human pathogenic species, including *C. lusitaniae*, *C.*
250 *guilliermondii*, *C. krusei*, and *C. glabrata* (**Figure 3b**). Phylogenetic analysis suggested independent
251 evolutionary trajectories of secreted lipases in emerging *C. auris* and related species, where the most
252 recent ortholog family of lipases includes *C. albicans LIP4*, *5*, *8* and *9* (**Figure S7**). The secretion of
253 lipases may be important during infection for nutrient acquisition, adaptation, virulence and immune
254 evasion[24]; we identified a predicted secretion signal in all lipases encoded by *C. auris* and related
255 species, supporting an extracellular role in these emergent MDR species.
256
257 **Conservation of known drug resistance and pathogenesis-associated genes**
258 Most of the genes previously associated with drug resistance and pathogenesis in *C. albicans*
259 are conserved in all the emergent multidrug resistant species. We identified orthologs of genes noted to
260 confer drug resistance in *C. albicans*, either by acquiring point mutations, increasing transcription, or

261 copy number variation. The annotated genome assemblies of *C. auris*, *C. haemulonii*, *C.*
262 *duobushaemulonii,* and *C. pseudohaemulonii* contain a single copy of the *ERG11* azole target and the
263 *UPC2* transcription factor that regulates expression of genes in the ergosterol pathway. Several of the
264 sites in *ERG11* subject to drug resistant mutations in *C. albicans* are similarly mutated in drug resistant
265 *C. auris* isolates (at positions Y132, K143, and F126, as reported previously[2]). Analysis of the
266 annotated genome assemblies of *C. auris* isolates agrees with prior SNP analysis[2], with the exception
267 of F126L mutation in B11221 (clade III) previously reported as F126T for this isolate. The F126L
268 mutation is also observed in a separate genomic study of *C. auris*[7] and in drug resistant *C. albicans*[25]
269 (**Figure S8**). The Y132F mutation that is observed in two *C. auris* clades is also found in *C.*
270 *pseudohaemulonii*, suggesting this site could contribute to azole resistance of this *C. pseudohaemulonii*
271 isolate. Otherwise, we found that sites in *ERG11* subject to drug resistant mutations in *C. albicans* are
272 not predicted to have drug resistant mutations *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C.*
273 *pseudohaemulonii* (**Figure S8**). In addition, we found that not all *C. auris* isolates have a single copy of
274 the *ERG11*. By copy number variation (CNV) analysis of the normalized read density distribution across
275 the genome we determined that two isolates (B11227 and B11229) from clade III had a duplication of a
276 140 kb region that includes *ERG11* (**Figure S9**). Long read assemblies of these isolates could be used
277 to examine the chromosomal context of this duplication.
278
279 Additionally, we identified orthologs of transporters from the ATP binding cassette (*ABC*) and major
280 facilitator superfamily (*MFS*) classes of efflux proteins that are involved in clinical antifungal resistance
281 in *C. albicans*, by the overexpression of *CDR* genes members of the ABC family and *MDR1* member of
282 the MFS transporters[26,27]. We identified a single copy of the multidrug efflux pump *MDR1* in all
283 sequenced isolates. For the candidate multidrug transporters such *CDR1*, *SNQ2* and related genes we
284 identified 5 copies in *C. auris* B8441, B11220, B11243, *C. haemulonii*, *C. pseudohaemulonii*, 6 copies
285 in *C. auris* B11221, and 4 copies in *C. duobushaemulonii* (**Table 2**; **Figure S10**). Phylogenetic analysis
286 of these ABC transporters showed that one of the genes in *C. auris* is related to *CDR1*/*CDR2*/*CDR11*,
287 two genes related to *CDR4*, and two genes related to *SNQ2*; *C. auris* B11221 has an additional copy of
288 this gene (**Figure S10**). The *TAC1* transcription factor that regulates expression of *CDR1* and *CDR2* in
289 *C. albicans* is present in two copies in *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C.*
290 *pseudohaemulonii* (**Table 2**).
291
292 While many gene families involved in pathogenesis in *C. albicans* are present in similar numbers in *C.*
293 *auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii*, there are some notable

294 differences such as of cell wall and transmembrane proteins. We identified similar numbers of the
295 secreted aspartyl proteases, lipases and oligopeptide transporters (*OPT*), and only one copy of the *ALS*
296 cell surface family of *C. albicans* (**Table 2**). We also examined whether other genes involved in the *C.*
297 *albicans* core filamentation response were conserved in the emerging multidrug-resistant species.
298 While most of these genes are conserved in *C. auris* and closely related species, two genes are
299 absent, candidalysin (*ECE1*) and the hyphal cell wall protein (*HWP1*), both of which are highly
300 expressed during the pathogenic phase of *C. albicans* and essential for hyphae formation (**Table 2**).
301 Thus, we additionally assessed if any other cell surface families of proteins are enriched in *C. auris* and
302 closely related species. We found a total of 75 genes with a predicted GPI anchor, including genes that
303 were found only in the emerging multidrug resistant clade, including one unique family expanded in *C.*
304 *auris* (**Table S3**). The most represented protein family domains in these genes encompassed the N-
305 terminal cell wall domain, the aspartyl protease domain, the fungal specific cysteine rich (CFEM)
306 domain, the flocculin domain, the cell-wall agglutinin ALS domain, the lysophospholipase catalytic
307 domain, and the glycosyl hydrolase family 16 domain (**Table S3**). The shared profile of these genes
308 across *C. auris* and other MDR species suggests that the more rarely observed species are primed to
309 become more common human pathogens.
310

311 **Transcriptional analysis to voriconazole and amphotericin B in *C. auris* susceptible and**
312 **resistant strains**

313 To investigate which *C. auris* genes are involved in multidrug-resistance, we carried out RNA-
314 Seq of *C. auris* strains B8441 and B11210 (clade I) to profile gene expression changes after exposure
315 to two antifungals, voriconazole (VCZ) and amphotericin B (AMB). B11210 is highly resistant to both
316 VCZ and AMB, while B8441 displays moderate resistance to VCZ and is susceptible to AMB (**Table**
317 **S4**; [2]). We identified differentially expressed genes (DEGs) in B8441 and B11210 (fold change (FC) > 2;
318 false discovery rate (FDR) < 0.05) after 2 and 4 hours of drug exposure (**Methods**). The response of
319 B8441 to both drugs only involved changes in expression of small sets of genes. In response to AMB, a
320 total of 39 genes were differentially expressed in B8441 across the 4-hour time course. These genes
321 were enriched in small molecule biosynthetic process and iron transport (enriched GO terms corrected-
322 $P < 0.05$, hypergeometric distribution with Bonferroni correction; **Table S4**), and include genes involved
323 in the *C. albicans* transcriptional response to AMB[28] such as argininosuccinate synthase (*ARG1*), a
324 putative ornithine carbamoyltransferase (*ARG3*), an aromatic decarboxylase (*ARO10*), the C-14 sterol
325 reductase (*ERG24*) involved in ergosterol biosynthesis, fatty-acid synthases (*FAS1*/*FAS2*),
326 sulfhydrylase (*MET15*), and several iron transporters (class *FTH1* and *SIT1*; **Table S4**). A set of 21

327 genes were differentially expressed in B8441 in response to VCZ; these included a subset of 14 genes
328 also involved in the response to AMB and were enriched in transmembrane transport and iron transport
329 categories (enriched GO terms corrected-$P < 0.05$, hypergeometric distribution with Bonferroni
330 correction; **Table S4**), including ferric reductase (*FRP1*), high affinity iron transporter (*FTH1*), glucose
331 transporter (*HGT7*), N-acetylglucosamine transporter (*NGT1*) and oligopeptide transporter (*OPT1*). The
332 oligopeptide transporter *PTR22* was only induced with VCZ (**Table S4**). We also examined expression
333 differences in B11210; as a large set of genes was identified as differentially expressed compared to
334 the control sample, we focused on the most highly induced or repressed genes. A total of 7 of the
335 genes that were differentially expressed in B8441 were similarly up-regulated in B11210, including
336 *ARG1*, *CSA1* and *MET15* (**Table S4**). To further examine differences in the response to each drug, we
337 identified differentially expressed genes across treatments at each time point. We identified 2 genes
338 that varied between AMB and VCZ in B8441, a small heat shock factor (*HSP20*) that was highly
339 induced with AMB and the *HGT17* transporter that was highly induced with VCZ. We also identified a
340 set of 37 DEGs that varied between these treatments in B11210. Eight genes were highly induced in
341 response to AMB treatment but not VCZ, including homologs of *CR_01630C_A*, *FTR1*, and *ZRT2*,
342 previously noted to have roles during *C. albicans* response to AMB[28]. In addition, twenty-nine DEGs
343 were highly induced in response to VCZ treatment but not AMB included the well-known multidrug
344 efflux pump (*MDR1*)[26] and other genes involved in *C. albicans* response to azoles[28], including a high-
345 affinity phosphate transporter (*PHO84*), the small subunit processome complex (*SAS10*), and an
346 uncharacterized protein induced in azole-resistant strains that overexpress *MDR1* (C2_01450C_A).
347
348 Since *C. auris* isolates B8441 and B11210 (clade I) had distinct resistant phenotypes predominantly in
349 response to AMB, we further examined expression changes between these two strains. While the drug
350 response was very similar in *C. auris* B8441 and B11210, we identified a small set of differentially
351 expressed genes (**Table S4**). Only two DEGs were found induced in B11210 but not B8441, including
352 an uncharacterized transmembrane protein (B9J08_004887) induced by AMB, and an uncharacterized
353 cell surface glycoprotein (B9J08_005574) induced by VCZ. Other genes induced in B8441 but not
354 B11210 were previously noted to be involved in the *C. albicans* transcriptional response to drug
355 exposure, including D-xylulose reductase (*XYL2*)[28], Cell surface mannoprotein (*MP65*)[29], NAD-aldehyde
356 dehydrogenase (*ALD5*)[30], Phosphoenolpyruvate carboxykinase (*PCK1*)[22,28], among other genes that
357 are important to the stress adaptation, including genes associated with oxidation reduction process and
358 fatty acid oxidation (**Table S4**).
359 To characterize the underlying emergence of multidrug resistant phenotype in *C. auris*, we examined

360 the conservation of DEGs across sequenced genomes in addition to predicted function. Some genes
361 induced by AMB and VCZ in *C. auris* B8441 were species-specific in the emerging MDR species,
362 including five ortholog families unique to *C. auris*, comprising three putative GPI-anchored cell wall
363 proteins, two genes similar to *IFF6* (a putative GPI-anchored adhesin-like protein), one gene homolog
364 of *PGA54*, and an aspartyl protease similar to *SAP8* (**Table S4**). In addition, this set included gene
365 families that are either expanded in *C. auris* and closely related species (**Figures 3b** and **4**). This
366 includes four *SIT*-like, one *FTR*-like, and two *OPT*-like class transporters, cell wall adhesins, and
367 several predicted secreted proteins (**Table S4**). Notably, while other Saccharomycetales species,
368 including *C. albicans* only have one copy of *SIT1* that is induced during AMB treatment, *C. auris* and
369 closely related species had up to 11 *SIT1*-like genes, and 4 of them are induced during AMB treatment
370 **Figures 3b, 4, S6)**. Together, these transcriptional changes highlighted shared and *C. auris*-specific
371 genes that might contribute to the MDR phenotype observed in *C. auris* isolates and provided candidate
372 genes to further investigate *C. auris* multidrug-resistance.
373
374
375 **Discussion**
376 As a recently emerging pathogen, *Candida auris* is not well studied to date, highlighting the
377 need for rapidly closing this knowledge gap to respond to the increasing number of fatal infections.
378 There is also a limitation on how much of the biology of *C. auris* we can infer from related *Candida*
379 species; *C. auris* is distantly related to the two most commonly observed pathogenic *Candida* species,
380 *C. albicans* and *C. glabrata*, as well as other sequenced species. Our comparative genomic analyses,
381 incorporating new genomic data for more closely related, multidrug-resistant species, revealed the
382 recent evolution of this group of emerging pathogens including shared properties that underlie
383 antifungal resistance and virulence. In addition to *C. auris* clades, we generated annotated genomes for
384 three other closely related species that are rarely observed to date as infecting humans. Building off
385 prior studies of individual loci[4,5,12,13], the phylogenetic relationship of these species was more clearly
386 resolved by whole genome comparisons; *C. haemulonii*, *C. duobushaemulonii*, and *C.*
387 *pseudohaemulonii* are more closely related to each other than to *C. auris*, with the closest relationship
388 between *C. duobushaemulonii*, and *C. pseudohaemulonii*. Our phylogenetic analysis integrating the
389 genomes of these species with other *Candida* highlight the distant relationship of this group to other
390 pathogenic *Candida* species and the placement of these species within the CTG clade.
391

392 To characterize mechanisms that may contribute to virulence and drug resistance, we compared the
393 gene content between the emerging multidrug resistant species and other related *Candida*. Recent
394 work found that virulence in *C. auris* appears similar to *C. albicans* and *C. glabrata*[31], suggesting that
395 shared gene content could play a role. *C. auris* shares some notable gene family expansions described
396 in *C. albicans* and related pathogens[20], including of transporters and secreted lipases. While an
397 expansion of transporters is shared, species-specific expansions have contributed to the diversification
398 of transporters in *C. auris* and closely related species. Similarly, the expansion of lipases suggests this
399 could be part of a shared mechanism of virulence; however, the roles of specific genes will need to be
400 investigated. By contrast, expansions of cell wall families detected in *C. albicans* and related pathogens
401 are not found in *C. auris* and related species. For example, the *ALS* family is represented by two to four
402 copies in *C. auris* and related MDR species, compared to the 8 copies present in *C. albicans*.
403 Examining the predicted cell wall proteins in *C. auris* did not reveal any highly expanded families;
404 perhaps the longer history of pathogenic interactions drives this diversification in the more commonly
405 observed pathogenic *Candida* species.
406
407 Drug resistance in *C. auris* likely involves mechanisms previously described in *C. albicans*, however the
408 specific transporters involved in drug response are less well conserved. All four species are resistant to
409 antifungal drugs, and previously described sites of point mutation in *C. auris* in *ERG11*, the target of
410 azole drugs, are conserved in all species. In addition, we find evidence of increased copy number of
411 *ERG11* in two *C. auris* isolates, but little evidence of large copy number variation, and no evidence of
412 aneuploidy. This suggests that increased copy number of *ERG11* may be a mechanism of drug
413 resistance in *C. auris*, as has previously been described in *C. albicans*[32]. However other mechanisms of
414 drug resistance may vary between species and strains. For example, we found that efflux proteins such
415 as the *CDR* family were not induced during amphotericin B or voriconazole treatment, instead other set
416 of transporters were up-regulated, including those that are either expanded or unique in *C. auris* and
417 closely multidrug emergent species, highlighting that different molecular mechanisms are likely involved
418 in the drug response.
419
420 Our analysis identified representative *C. auris* isolates for each of the two mating types, suggesting that
421 this species can undergo mating and meiosis. Each of the four described *C. auris* clades consists of
422 isolates that are all *MTL**a*** or *MTL*α*;* as the clades are geographically restricted, this suggests that there
423 is a geographic barrier to opposite sex mating. This expectation would change if wider sampling
424 demonstrated the presence of both mating types within a geographic area. The potential for mating

425 within this species is supported by the conservation of genes involved in mating and meiosis; these
426 patterns are similar to that of the related species *C. lusitaniae*, for which mating and recombination
427 have been demonstrated[21]. Isolates from clades of opposite mating types could be directly tested for
428 mating and production of progeny. One potential barrier to mating between clades is the presence of
429 chromosomal rearrangements, as some recombination events may result in inviable progeny.
430
431 Analyses of these genomes have revealed fundamental aspects of these emerging multidrug resistant
432 fungi. The role of specific genes in mating, drug resistance or pathogenesis needs to be directly tested,
433 utilizing gene deletion technologies recently adapted for *C. auris* (e.g. [33]). Further analysis of this data
434 will not only advance our understanding of the basis of drug resistance and virulence in this pathogen
435 but can also inform development of fungal diagnostics for accurate tracking of these emerging
436 pathogens.
437
438
439 **Methods**
440 **Selected isolates and genome sequencing**
441 Strains used in this study were described previously[2,17,18]; brief details are presented in **Table 1**.
442 Isolates were grown on Sabouard Dextrose media supplemented with chloramphenicol and gentamycin
443 and incubated for 24-48 hours at 37°C. For Illumina sequencing, genomic DNA was extracted using the
444 *Quick*-DNA™ (ZR) Fungal/Bacterial Miniprep Kit (Zymo Research, Irvine, CA, USA). Genomic libraries
445 were constructed and barcoded using the NEBNext Ultra DNA Library Prep kit (New England Biolabs,
446 Ipswich, MA, USA) by following manufacturer's instructions. Genomic libraries were sequenced using
447 either Illumina HiSeq 2500 with HiSeq Rapid SBS Kit v2 or Illumina MiSeq platform using MiSeq
448 Reagent Kit v2 (Illumina, San Diego, CA, USA). For PacBio sequencing, DNA was extracted using
449 MasterPure™ Yeast DNA Purification Kit (Epicenter, Madison, WI, USA). Single-molecule real-time
450 (SMRT) sequencing was done using the PacBio RS II SMRT DNA sequencing system (Pacific
451 Biosciences, Menlo Park, CA, USA). Specifically, 20-kb libraries were generated with the SMRTbell
452 Template Prep Kit 1.0 (Pacific Biosciences). Libraries were bound to polymerase using the
453 DNA/Polymerase Binding Kit P6v2 (Pacific Biosciences), loaded on two SMRTcells (Pacific
454 Biosciences), and sequenced with C4v2 chemistry (Pacific Biosciences) for 360 min movies.
455
456 **Genome assemblies**

457 The genomes of B8441 and B11221 *C. auris* isolates were assembled as described[2]. *C.*
458 *haemulonii* and *C. duobushaemulonii* genomes were assembled using Canu v1.6[34]. The resultant
459 contigs were checked for further joins and circularity using Circlator v1.5[35]. The final contigs were
460 polished using Quiver, part of SmrtAnalysis suite v2.3 (Pacific Biosciences)[36]. The sequence order for
461 the chromosomes was verified using restriction enzyme AflII Whole Genome Mapping (OpGen,
462 Gaithersburg, MA).
463
464 The sequenced Illumina reads of *C. auris* (B11220 and B11243), and *C. pseudohaemulonii* (strain
465 B12108) were assembled using the SPAdes assembler v3.1.1[37]. Next, Pilon v1.16[38] was used to polish
466 the best assembly of each isolate, resolving single nucleotide errors (SNPs), artifactual indels and local
467 mis-assemblies. All genome assemblies were evaluated using the GAEMR package
468 (http://software.broadinstitute.org/software/gaemr/), which revealed no aberrant regions of coverage,
469 GC content or contigs with sequence similarity suggestive of contamination. Scaffolds representing the
470 mitochondrial genome were separated out from the nuclear assembly. All genome assemblies have
471 been deposited at deposited at DDBJ/EMBL/GenBank (see ***Data availability statement***). To address if
472 any strain representing each *C. auris* clade could be uniformly diploid, we examined candidate
473 heterozygous positions predicted by Pilon v1.12[38] using mapped Illumina data. The low frequency and
474 absence of such positions supported that all sequenced genomes are homozygous haploids.
475
476 **Optical mapping**
477 Two strains of *C. auris* (B11221 and B8441) were compared by the OpGen optical mapping
478 platform (OpGen, Inc., Gaithersburg, Maryland). High molecular weight genomic DNA from overnight
479 grown cells were purified with Argus HMW DNA Isolation Kit (OpGen, Inc.) and examined for quality
480 and concentration using the ARGUS QCards(OpGen, Inc.). The software program Enzyme Chooser
481 (OpGen, Inc.) identified *BamHI* restriction endonuclease to be optimal for optical map production,
482 because its cleavage of reference genomes would result in fragments that average 6–12 kbp in size,
483 with no fragments larger than 80 kbp. Single genomic DNA fragments were loaded onto a glass surface
484 of a MapCard (OpGen, Inc.) using the microfluidic device, washed and then digested with *BamHI*
485 restriction enzyme, and stained with JOJO-1 dye through the ARGUS MapCard Processor (OpGen,
486 Inc.). Map cards were scanned and analyzed by automated fluorescent microscopy using the ARGUS
487 Whole Genome Mapper (OpGen, Inc.). The single molecule restriction map collections were then tiled
488 according to overlapping fragment patterns to produce a consensus whole genome map. This map was
489 imported into MapSolver (OpGen, Inc.) along with predicted *in silico* maps of contigs derived from

490 WGS, using the same restriction enzyme for ordering and orientation of contigs during genome
491 circularization. *In-silico* predicted optical maps of complete genomes were scaled according to the size
492 of sequenced genomes to show identity with Optical maps (**Figures S1**).
493
494 **RNA-Seq of B8441 and B11220 during drug treatment**
495 For RNA extraction *C. auris* cells were grown in YPD broth medium (Difco Laboratories, Sparks,
496 MD) at 30°C in a shaking incubator at 300 rpm. After 18 hours, the stationary phase cells were diluted
497 with the equal volume of fresh YPD broth and incubated for two hours at 37° to induce growth. After
498 that, the cells were treated with Amphotericin B at the final concentration of 0.25 μg/mL or Voriconazole
499 at the final concentration of 1 μg/mL (Sigma-Aldrich, St. Lois, MO) and incubated at 30°C for additional
500 2 and 4 hours in the presence of drug. Cells were centrifuged for 2 minutes at 12000xg, pellets were
501 flash frozen in dry ice/ethanol bath and stored at -80°C. RNA was isolated using RiboPure™-Yeast
502 rapid RNA isolation kit (Life technologies, Carlsbad, CA) using the manufacturer's protocol. RNA was
503 adapted for sequencing using the RNAtag-Seq approach[39], with the modification that the yeast
504 RiboZero reagent was used for rRNA depletion. For each condition, two biological replicates were
505 performed, and the read counts per transcript were highly correlated between replicates (R> 0.90).
506
507 **Gene annotation**
508 Gene annotation in *C. auris* was performed using RNA-Seq paired-end reads to improve gene
509 calling and structure predictions. Briefly, we mapped RNA-Seq reads to the genome assembly using
510 Tophat2, and use the alignments to predict genes using BRAKER1[40], that combines GeneMark-ET[41]
511 and AUGUSTUS[42], incorporating RNA-Seq data into unsupervised training and subsequently generates
512 *ab initio* gene predictions. Additionally, we re-annotated the genome of the 6684 strain[13] improving its
513 gene set and predicted gene structures. tRNAs were predicted using tRNAscan[43] and rRNAs predicted
514 using RNAmmer[44]. Genes containing PFAM domains found in repetitive elements or overlapping
515 tRNA/rRNA features were removed. Genes were named and numbered sequentially. For the protein
516 coding-gene name assignment we combined HMMER PFAM/TIGRFAM, Swissprot and Kegg products.
517 For comparative analysis genes were functionally annotated by assigning PFAM domains, GO terms,
518 and KEGG classification. HMMER3[45] was used to identify PFAM domains using release 27. GO terms
519 were assigned using Blast2GO[46], with a minimum e-value of $1x10^{-10}$. Protein kinases were identified
520 using Kinannote[47] and transporter families using TCDB version 01-05-2017[48]. To evaluate the
521 completeness of predicted gene sets, the representation of core eukaryotic genes was analyzed using
522 CEGMA genes[49] and BUSCO[50].

523

524 **Comparative genomics and phylogenomic analysis**

525 To examine the phylogenetic relationship of the emerging multidrug-resistance clade, including
526 *C. auris*, *C. haemulonii*, *C. duobushaemulonii,* and *C. pseudohaemulonii* we identified single copy
527 orthologs in these sequenced genomes and twelve related species using OrthoMCL v1.4[51] (Markov
528 index 1.5; maximum e-value 1e-5). Protein sequences of 20 genomes were aligned using MUSCLE,
529 and a phylogeny was estimated from the concatenated alignments using RAxML v7.7.8[52] with model
530 PROTCATWAG with a total of 1,000 bootstrap replicates. To compare gene family expansion and
531 contractions, we used orthologous gene clusters we classified as core, auxiliary and unique. We then
532 searched for expansions or contractions in functionally classified genes by assigning PFAM domains,
533 GO terms, and KEGG classification. Using a matrix of gene class counts for each classification type,
534 we identified enrichment comparing the emerging multidrug-resistance clade with all the other related
535 species using Fisher's exact test. Fisher's exact test was used to detect enrichment of PFAM, KEGG,
536 GO terms, and transporter families between groups of interest, and p-values were corrected for multiple
537 comparisons[53]. Significant (corrected *P*-value < 0.05) gene class expansions or depletions were
538 examined for different comparisons.

539

540 **Transcriptional analysis of *C. auris* RNA-Seq**

541 RNA-Seq reads were aligned to the transcript sequences of *C. auris* B8441 or B11221 using
542 Bowtie2[54]. Transcript abundance was estimated using RSEM (RNA-Seq by expectation maximization;
543 v.1.2.21) as transcripts per million (TPM). TPM-normalized 'transcripts per million transcripts' (TPM) for
544 each transcript were calculated, and differentially expressed transcripts were identified using edgeR[55],
545 all as implemented in the Trinity package version 2.1.1[56]. Genes were considered differentially
546 expressed only if they had a 2-fold change difference (> 2 FC) in TPM values and a false discovery rate
547 below or equal to 0.05 (FDR < 0.05), unless specified otherwise. To determine major patterns of
548 antifungal-response specific we clustered gene expression patterns by *k-means*. To identify functional
549 enrichment of differentially expressed genes, we used functional gene assignations from PFAM, GO
550 terms and KEGG (see ***Gene annotation***), and then performed comparisons with Fisher's exact test. To
551 identify possible functions of the gene products of significantly differentially expressed drug resistance
552 genes, protein homologs were assigned based on orthology, functional assignments (GO, PFAM,
553 TIGRFAM) and experimental evidence from *Candida* genome database
554 (http://www.candidagenome.org).

555

556 **Data availability statement**

557 All genome assemblies and gene annotations have been deposited at DDBJ/EMBL/GenBank
558 under the following accession numbers: *C. auris* B8441 PEKT00000000; *C. auris* B11221
559 PGLS00000000, *C. auris* B11220 PYFR00000000, *C. auris* B11243 PYGM00000000, *C. haemulonii*
560 B11899 PKFO00000000, *C. duobushaemulonii* B09383 PKFP00000000, *C. pseudohaemulonii* B12108
561 PYFQ00000000. The RNA-Seq data from *C. auris* has been deposited at GenBank under Bioproject
562 PRJNA445471.

563

564 **Acknowledgements**

565 We thank Sarah Young and Paul Cao for their assistance with genome annotation, the Broad
566 Technology labs for RNAtag-Seq library construction, and the Broad Genomics Platform for Illumina
567 sequencing. This project has been funded in whole or in part with Federal funds from the National
568 Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and
569 Human Services, under award N°: U19AI110818. The use of product names in this manuscript does not
570 imply their endorsement by the US Department of Health and Human Services. The finding and
571 conclusions in this article are those of the authors and do not necessarily represent the views of the
572 CDC.

573

574 **Author contributions**

575 Conceived and designed the study: APL CAC. Performed experiments: LG NAC VNL PJ.
576 Performed the assembly and annotation: JFM. Analyzed the data: JFM VNL PJ RAF CAC. Wrote the
577 paper: JFM CAC.

578

579 **Additional information**

580 Competing interests: The authors declare no competing interests.

581

582 **Tables and Figures**

583

584 **Table 1.** Genome assembly statistics of *Candida auris* and closely related species*

Species	*Candida auris*				*C. hae*	*C. duo*	*C. pseu*
Strain	B8441	B11221	B11220	B11243	B11899	B09383	B12108
Clade	I	III	II	IV	.	.	.
Country of origin	Pakistan	South Africa	Japan	Venezuela	Israel	US	Venezuela
Total assembly size (Mb)	12.4	12.7	12.1	12.3	13.3	12,3	12.6
Chromosomes	7	7	.	.	.	.	.
Assembly anchored (%)	98.8	97.2	.	.	.	.	.
Scaffolds	15	20	324	240	11	7	36
Contigs	18	23	324	265	11	7	41
Scaffold N50 (Mb)	1.1	2.4	0.06	0.09	1.7	3.3	0.64
Scaffold N90 (kb)	777	949	19.5	27.1	952	788	227
GC content (%)	45.2	45.3	45.0	45.0	45.3	46.9	47.2
Protein coding genes	5,421	5,527	5,546	5,601	5,410	5,331	5,288

585 **C. hae* = *C. haemulonii*; *C. duo* = *C. duobushaemulonii*; *C. pseu* = *C. pseudohaemulonii*

586

587 **Table 2.** Conservation of genes involved in pathogenesis and drug resistance*

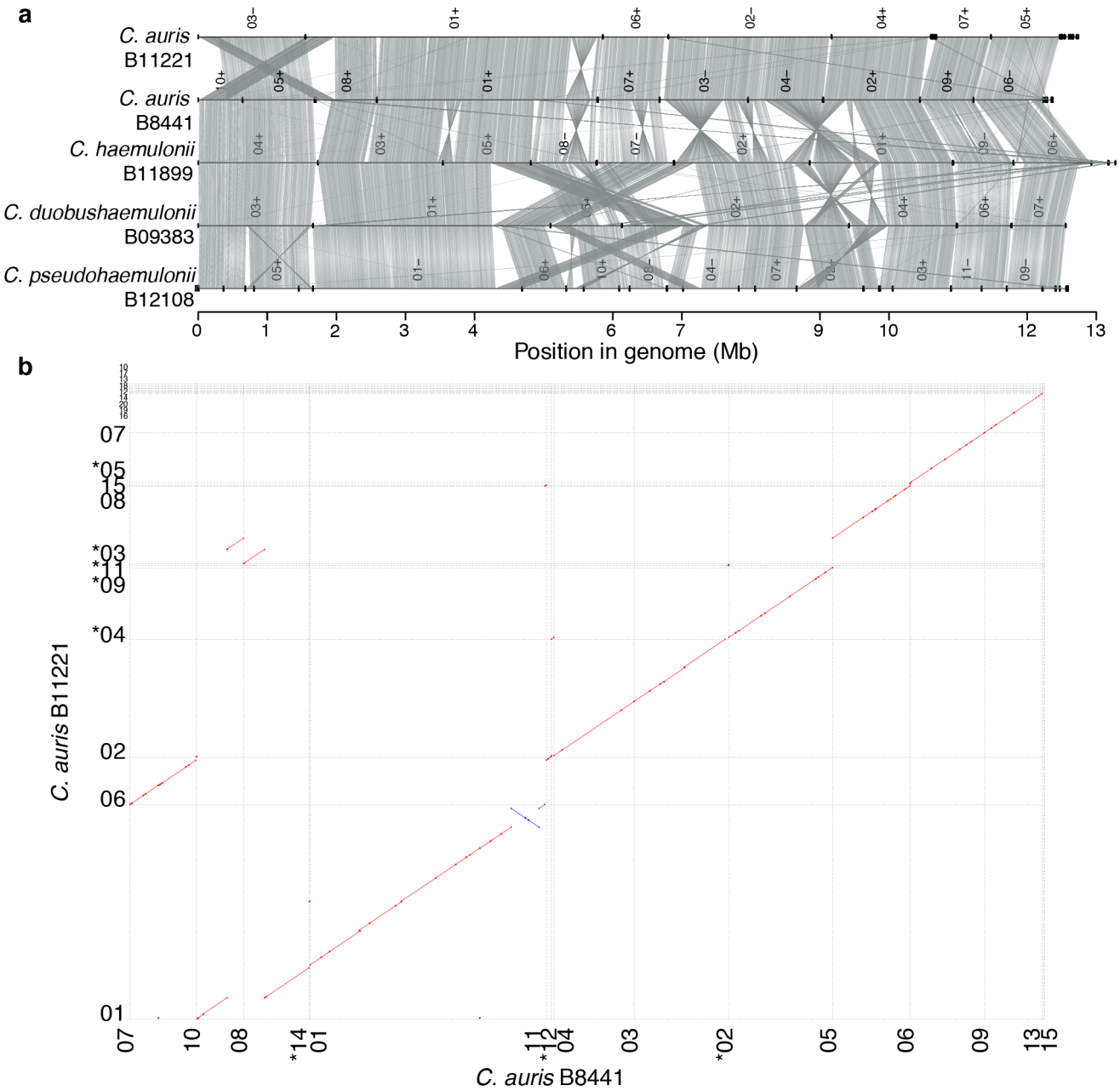
Category	*C. albicans*	*Candida auris* (B8441 - I)	*Candida auris* (B11221 - III)	*Candida auris* (B11220 - II)	*Candida auris* (B11243 - IV)	*Candida auris* (6684 - I)	*Candida haemulonii*	*Candida pseudohaemulonii*	*Candida duobushaeumulonii*	*Candida lusitaniae*	*Debaryomyces hansenii*	*Candida guilliermondii*	*Candida tropicalis*	*Candida albicans*	*Candida dubliniensis*	*Candida orthopsilosis*	*Candida parapsilosis*	*Lodderomyces elongisporus*	*Candida krusei*	*Candida glabrata*	*Saccharomyces cerevisiae*
Drug resistance	ERG11	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	TAC1	••	••	••	••	••	••	••	••	••	••	•	•	•	•	•	•	•	•		
	UPC2	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	••	••
	MDR1	•	•	•	•	•	•	•	•	•	•••	••• •••	••	•	•	•••	••	••• •		•	•
	SNQ2 CDR4 CDR2 CDR11 CDR1	••• ••	••• •••	••• ••	••• ••	••• ••	••• ••	••• ••	••• •	•••	••• •••	••• ••• ••	••	••• ••	••• ••	•••	••• •••	••• ••	•••	••• ••	••• ••
Secreted aspartyl proteinases	SAP1 SAP2 SAP8 SAP3	•••	•••	••	•••	•••	••	••	••	••		••• •	•••	••• •	••• •	•••	••	•			
	SAP9	•	•	•	•	•	•	•	•	•	••	•	•	•	•	••	••• ••	•	••	•	•••
Secreted lipases	LIP4, LIP9 LIP5, LIP8 LIP2, LIP1 LIP10, LIP6 LIP3	••• ••• ••	••• ••• ••	••• ••• ••	••• ••• ••	••• ••• ••	••• ••• •••	••• ••• •••	••• ••• ••		•	•	••	••• ••• •••	••• ••• •••	••	••	•			
Cell wall adhesins	ALS2, ALS5 ALS1, ALS9 ALS4, ALS3 ALS7, ALS6	•	•	•	•	•	••	•••	•••	•		••	••• •	••• ••• ••	••• •••	•	••				
	ALS-like	••	••	•	••	••		•	•												

588

589 *Each dot represents a gene count. Roman numerals represent *C. auris* clades.

590

591



592

593 **Figure 1.** Whole genome conservation, structure and synteny. **(a)** Genome wide synteny among
594 *Candida auris*, *C. haemulonii*, *C. duobushaemulonii*, and *C. pseudohaemulonii*. Shared synteny regions
595 are based on orthologs. Isolate names are shown to the right of their genomes, which are represented
596 by lines, with vertical lines indicating scaffold borders, and their identifiers listed above
597 (+/−=orientation). **(b)** Shared synteny regions based on whole genome alignments between *C. auris*
598 B8441 and B11221.
599

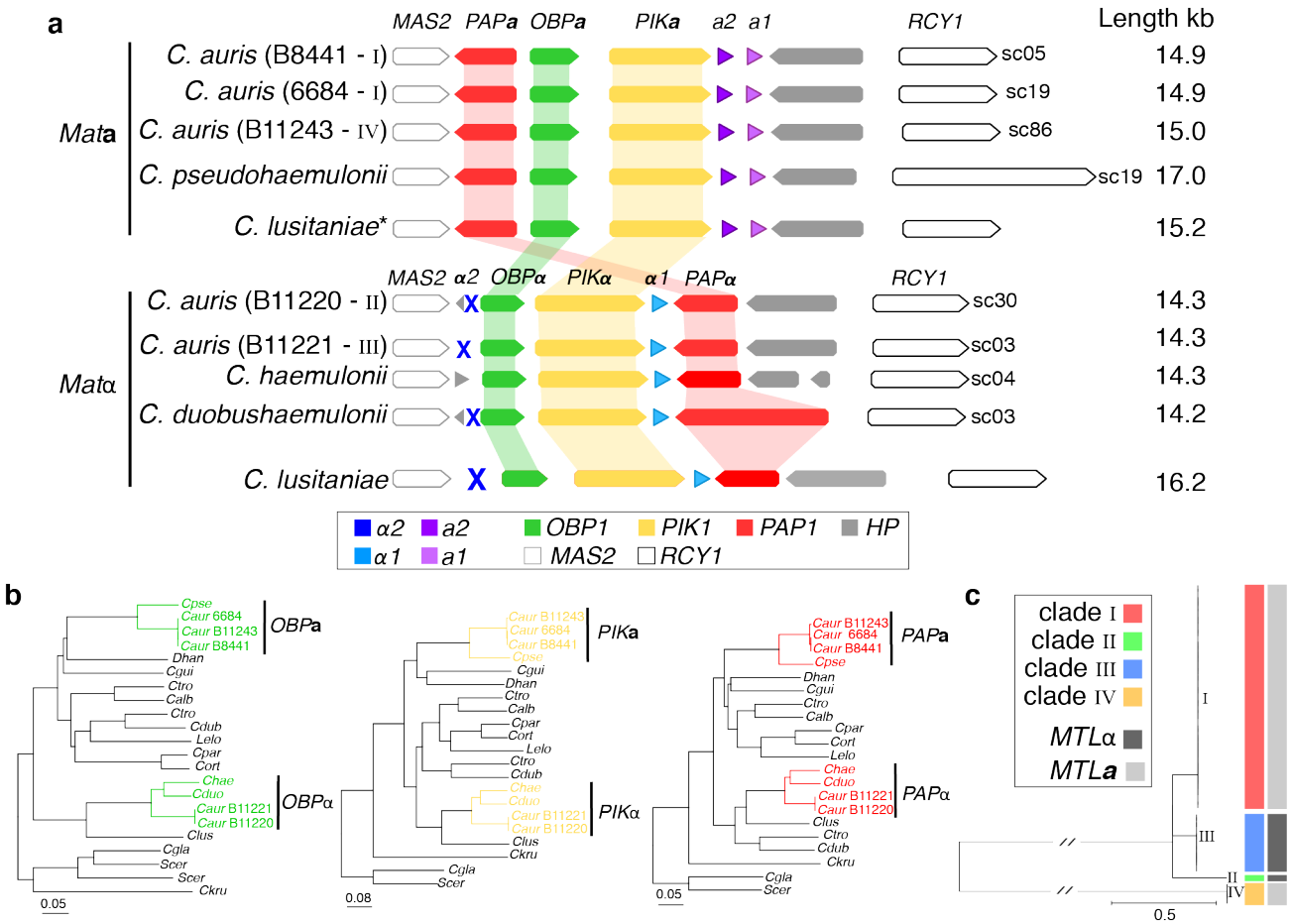


Figure 2. Identification of Mating-type loci (*MTL*) in *Candida auris* and closely related species. **(a)** Synteny schema depicting orientation and conservation of the color-coded *MTL* idiomorphs and genes adjacent to the *MTL*. The putative MTL in *C. auris*, *C. haemulonii*, *C. duobushaemulonii*, and *C. pseudohaemulonii* are shown in comparison with the *MTL**a*** and *MTL*α idiomorphs from *C. lusitaniae*. **(b)** Phylogenetic analysis of the non-mating flanking genes (*PIK1/OBP1/PAP1*) showing the inheritance of idiomorphs of these genes within the *MTL**a*** and *MTL*α loci. Branch lengths indicate the mean number of changes per site. **(c)** Phylogenetic tree of *C. auris* isolates from Lockhart et al. 2016. Isolates are color-coded according the clades (I, II, III and IV) and mating type (*MTL**a*** and *MTL*α). Figure S3 shows isolates, origin and the normalized depth read coverage of mapped positions for all isolates aligned to B8441 (*MTL**a***) and B11221 (*MTL*α), supporting the classification into *MTL**a*** or *MTL*α.

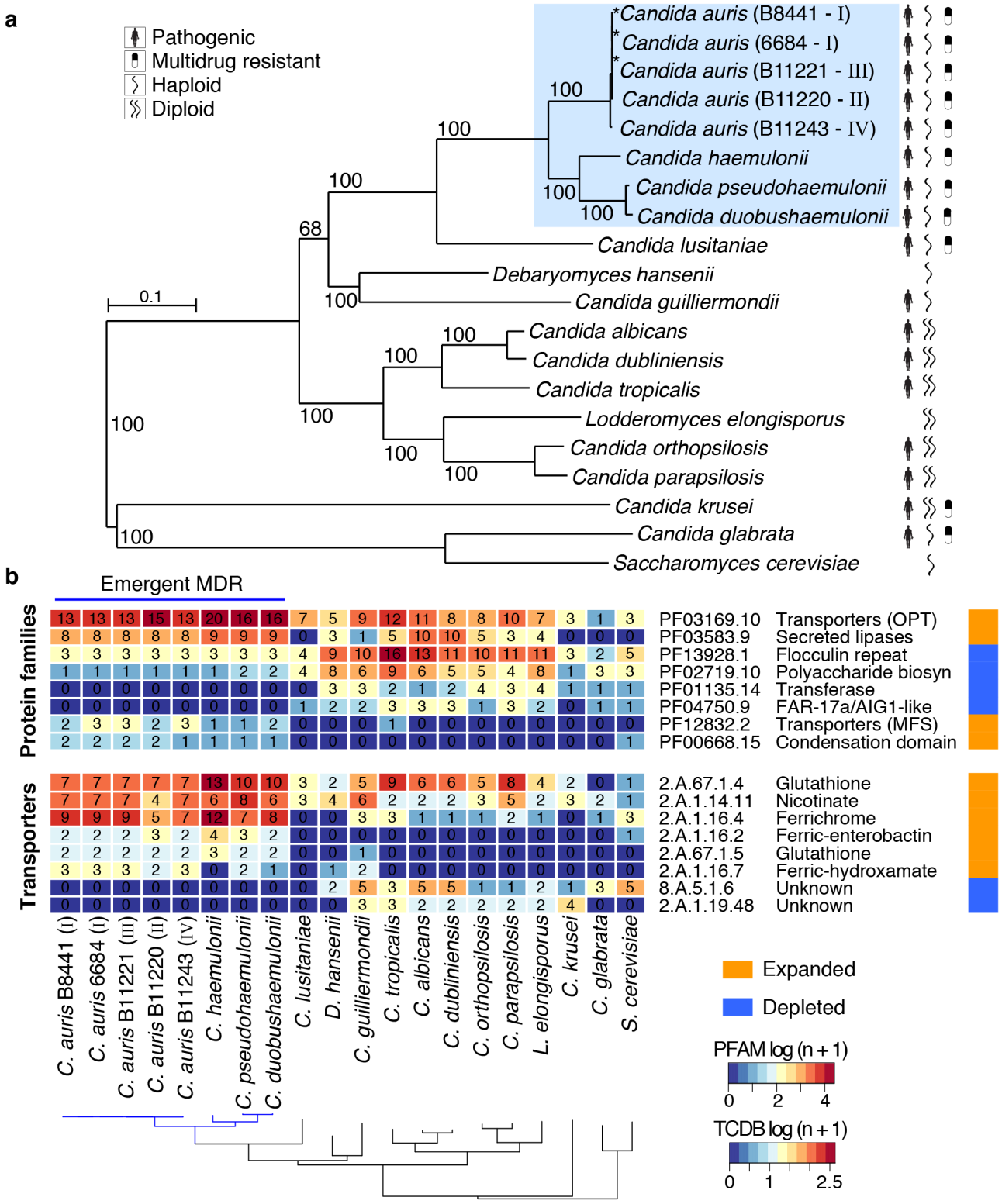


Figure 3. Phylogenomic and gene family changes across *Candida auris* and closely related species. **(a)** Maximum likelihood phylogeny using 1,570 core genes based on 1,000 replicates, among 20 annotated genome assemblies, including *Candida auris, C. haemulonii* (B11899), *C. duobushaemulonii* (B09383), and *C. pseudohaemulonii* (B12108), and closely related species. Branch lengths indicate the

617 mean number of changes per site. **(b)** Heatmap depicting results of protein family enrichment analysis
618 (PFAM domains; corrected *p-value* < 0.05) comparing the gene content of *C. auris* strains representing
619 each clade, *C. haemulonii*, *C. duobushaemulonii*, and *C. pseudohaemulonii*, and other closely related
620 species, including *C. lusitaniae*, *C. albicans*, *C. krusei* and *C. glabrata*. Values are colored along a blue
621 (low counts) to red (high counts) color scale, with color scaling relative to the low and high values of
622 each row. Each protein family domain has a color code (right) indicating whether expanded or depleted.

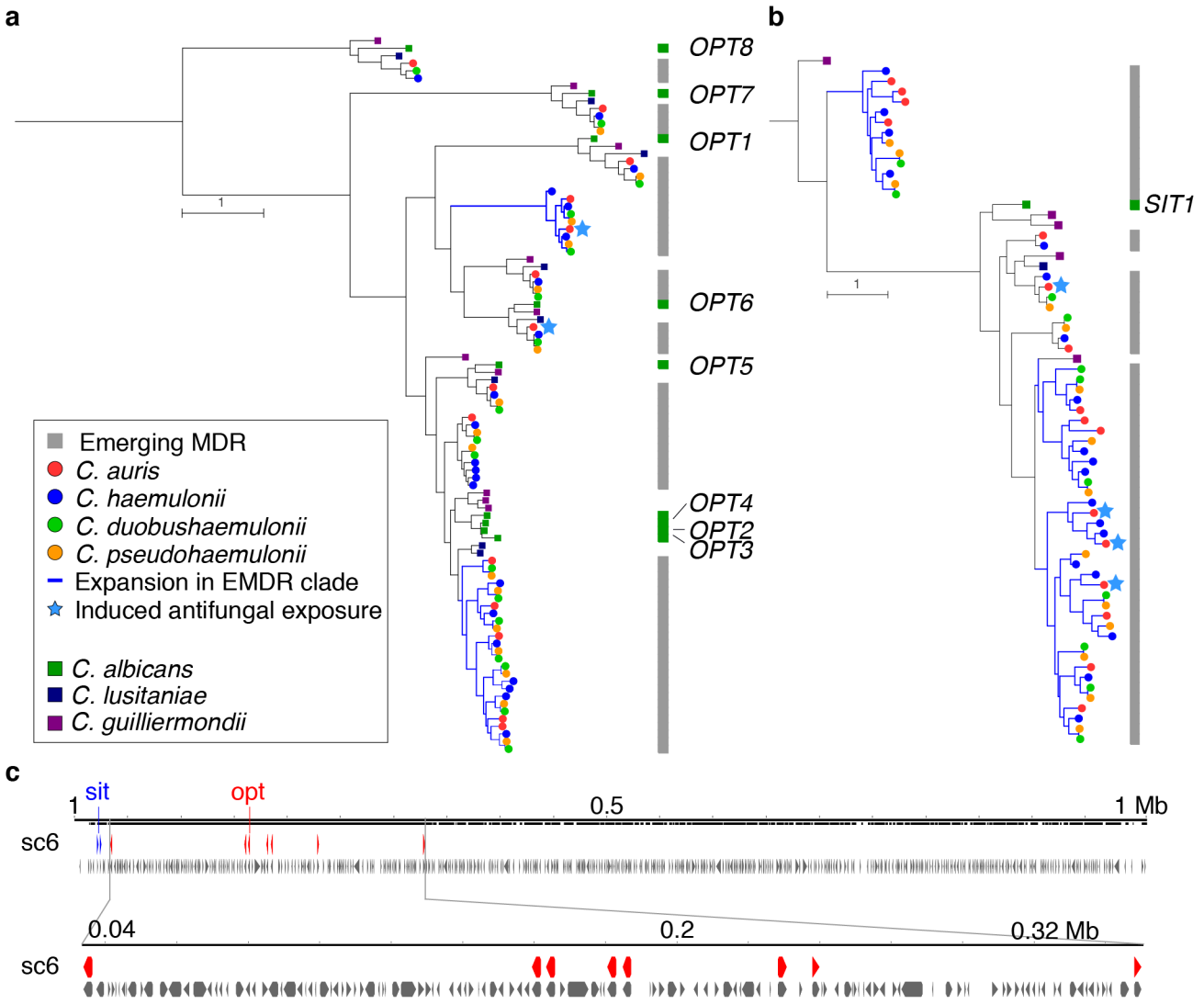


Figure 4. Phylogenetic relationships of oligopeptide transporters (*OPT*) and siderophore iron transporters (*SIT*) families. Phylogenetic trees estimated by maximum likelihood with RAxML showing expansion of *OPT* **(a)** and *SIT* **(b)** transporter families in *C. auris* and *C. haemulonii*, *C. duobushaemulonii* and *C. pseudohaemulonii*. Each species has a color code and lineage-specific expansions (blue branches) can be seen in *C. auris* and closely related species relative to the close ancestor *C. lusitaniae* and *C. albicans*. Orthologs of *OPT* and *SIT* transporters in *C. albicans* are depicted alongside each tree. **(c)** Chromosome view depicting genes and orientation located in chromosome 6 (B8441 scaffold05). This region highlights expansion and tandem duplication of eight *OPT* class transporters (in red).

634 **References**

635 1. Clancy, C. J. & Nguyen, M. H. Emergence of *Candida auris*: An International Call to Arms. *Clin.*
636 *Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* **64,** 141–143 (2017).
637 2. Lockhart, S. R. *et al.* Simultaneous Emergence of Multidrug-Resistant *Candida auris* on 3
638 Continents Confirmed by Whole-Genome Sequencing and Epidemiological Analyses. *Clin Infect Dis*
639 (2016).
640 3. Schelenz, S. *et al.* First hospital outbreak of the globally emerging *Candida auris* in a European
641 hospital. *Antimicrob. Resist. Infect. Control* **5,** (2016).
642 4. Cendejas-Bueno, E. *et al.* Reclassification of the *Candida haemulonii* Complex as *Candida*
643 *haemulonii* (*C. haemulonii* Group I), *C. duobushaemulonii* sp. nov. (*C. haemulonii* Group II), and *C.*
644 *haemulonii var. vulnera var. nov.*: Three Multiresistant Human Pathogenic Yeasts. *J. Clin. Microbiol.*
645 **50,** 3641–3651 (2012).
646 5. Kumar, A. *et al. Candida haemulonii* species complex: an emerging species in India and its genetic
647 diversity assessed with multilocus sequence and amplified fragment-length polymorphism analyses.
648 *Emerg. Microbes Infect.* **5,** e49 (2016).
649 6. Kathuria, S. *et al.* Multidrug-Resistant *Candida auris* Misidentified as *Candida haemulonii*:
650 Characterization by Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry
651 and DNA Sequencing and Its Antifungal Susceptibility Profile Variability by Vitek 2, CLSI Broth
652 Microdilution, and Etest Method. *J. Clin. Microbiol.* **53,** 1823–1830 (2015).
653 7. Rhodes, J. *et al.* Genomic epidemiology of the UK outbreak of the emerging human fungal
654 pathogen *Candida auris*. *Emerg. Microbes Infect.* **7,** 43 (2018).
655 8. Borman, A. M., Szekely, A. & Johnson, E. M. Isolates of the emerging pathogen *Candida auris*
656 present in the UK have several geographic origins. *Med. Mycol.* **55,** 563–567 (2017).
657 9. Sharma, C., Kumar, N., Pandey, R., Meis, J. F. & Chowdhary, A. Whole genome sequencing of
658 emerging multidrug resistant *Candida auris* isolates in India demonstrates low genetic variation.
659 *New Microbes New Infect.* **13,** 77–82 (2016).
660 10. Chowdhary, A. *et al.* Multidrug-resistant endemic clonal strain of *Candida auris* in India. *Eur. J. Clin.*
661 *Microbiol. Infect. Dis.* **33,** 919–926 (2014).
662 11. Prakash, A. *et al.* Evidence of genotypic diversity among *Candida auris* isolates by multilocus
663 sequence typing, matrix-assisted laser desorption ionization time-of-flight mass spectrometry and
664 amplified fragment length polymorphism. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol.*
665 *Infect. Dis.* **22,** 277.e1-9 (2016).
666 12. Ben-Ami, R. *et al.* Multidrug-Resistant *Candida haemulonii* and *C. auris*, Tel Aviv, Israel. *Emerg.*
667 *Infect. Dis.* **23,** 195–203 (2017).
668 13. Chatterjee, S. *et al.* Draft genome of a commonly misdiagnosed multidrug resistant pathogen
669 *Candida auris*. *BMC Genomics* **16,** 686 (2015).
670 14. Gargeya, I. B., Pruitt, W. R., Simmons, R. B., Meyer, S. A. & Ahearn, D. G. Occurrence of
671 *Clavispora lusitaniae, the teleomorph of Candida lusitaniae*, among clinical isolates. *J. Clin.*
672 *Microbiol.* **28,** 2224–2227 (1990).
673 15. Gabaldón, T., Naranjo-Ortíz, M. A. & Marcet-Houben, M. Evolutionary genomics of yeast pathogens
674 in the Saccharomycotina. *FEMS Yeast Res.* **16,** (2016).
675 16. Chowdhary, A. *et al.* A multicentre study of antifungal susceptibility patterns among 350 *Candida*
676 *auris* isolates (2009–17) in India: role of the ERG11 and FKS1 genes in azole and echinocandin
677 resistance. *J. Antimicrob. Chemother.* (2018).
678 17. Chow, N. A. *et al.* Genome Sequence of a Multidrug-Resistant *Candida haemulonii* Isolate from a
679 Patient with Chronic Leg Ulcers in Israel. *Genome Announc.* **(in press),** (2018).
680 18. Chow, N. A. *et al.* Genome Sequence of the Amphotericin B-Resistant *Candida duobushaemulonii*
681 Strain, B09383. *Genome Announc.* **(in press),** (2018).

682 19. Oh, B. J. *et al.* Biofilm formation and genotyping of *Candida haemulonii*, *Candida*
683 *pseudohaemulonii*, and a proposed new species (*Candida auris*) isolates from Korea. *Med. Mycol.*
684 **49,** 98–102 (2011).
685 20. Butler, G. *et al.* Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes.
686 *Nature* **459,** 657–62 (2009).
687 21. Reedy, J. L., Floyd, A. M. & Heitman, J. Mechanistic Plasticity of Sexual Reproduction and Meiosis
688 in the *Candida* Pathogenic Species Complex. *Curr. Biol.* **19,** 891–899 (2009).
689 22. Copping, V. M. S. *et al.* Exposure of *Candida albicans* to antifungal agents affects expression of
690 SAP2 and SAP9 secreted proteinase genes. *J. Antimicrob. Chemother.* **55,** 645–654 (2005).
691 23. Reuß, O. & Morschhäuser, J. A family of oligopeptide transporters is required for growth of *Candida*
692 *albicans* on proteins. *Mol. Microbiol.* **60,** 795–812 (2006).
693 24. Hube, B. *et al.* Secreted lipases of *Candida albicans*: cloning, characterisation and expression
694 analysis of a new gene family with at least ten members. *Arch. Microbiol.* **174,** 362–374 (2000).
695 25. Perea, S. *et al.* Prevalence of Molecular Mechanisms of Resistance to Azole Antifungal Agents in
696 Candida albicans Strains Displaying High-Level Fluconazole Resistance Isolated from Human
697 Immunodeficiency Virus-Infected Patients. *Antimicrob. Agents Chemother.* **45,** 2676–2684 (2001).
698 26. Gaur, M. *et al.* MFS transportome of the human pathogenic yeast *Candida albicans*. *BMC*
699 *Genomics* **9,** 579 (2008).
700 27. Prasad, R. & Goffeau, A. Yeast ATP-binding cassette transporters conferring multidrug resistance.
701 *Annu. Rev. Microbiol.* **66,** 39–63 (2012).
702 28. Liu, T. T. *et al.* Genome-wide expression profiling of the response to azole, polyene, echinocandin,
703 and pyrimidine antifungal agents in Candida albicans. *Antimicrob. Agents Chemother.* **49,** 2226–
704 2236 (2005).
705 29. Sorgo, A. G. *et al.* Effects of fluconazole on the secretome, the wall proteome, and wall integrity of
706 the clinical fungus Candida albicans. *Eukaryot. Cell* **10,** 1071–1081 (2011).
707 30. Rogers, P. D. & Barker, K. S. Evaluation of differential gene expression in fluconazole-susceptible
708 and -resistant isolates of Candida albicans by cDNA microarray analysis. *Antimicrob. Agents*
709 *Chemother.* **46,** 3412–3417 (2002).
710 31. Fakhim, H. *et al.* Comparative virulence of *Candida auris* with *Candida haemulonii*, *Candida*
711 *glabrata* and *Candida albicans* in a murine model. *Mycoses* (2018). doi:10.1111/myc.12754
712 32. Selmecki A., Gerami-Nejad M., Paulson C., Forche A. & Berman J. An isochromosome confers
713 drug resistance in vivo by amplification of two genes, ERG11 and TAC1. *Mol. Microbiol.* **68,** 624–
714 641 (2008).
715 33. Grahl, N., Demers, E. G., Crocker, A. W. & Hogan, D. A. Use of RNA-Protein Complexes for
716 Genome Editing in Non-*albicans Candida* Species. *mSphere* **2,** e00218-17 (2017).
717 34. Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptivek-mer weighting and
718 repeat separation. *Genome Res.* **27,** 722–736 (2017).
719 35. Hunt, M. *et al.* Circlator: automated circularization of genome assemblies using long sequencing
720 reads. *Genome Biol.* **16,** 294 (2015).
721 36. Chin, C.-S. *et al.* Nonhybrid, finished microbial genome assemblies from long-read SMRT
722 sequencing data. *Nat. Methods* **10,** 563–569 (2013).
723 37. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell
724 sequencing. *J Comput Biol* **19,** 455–77 (2012).
725 38. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and
726 genome assembly improvement. *PLoS One* **9,** e112963 (2014).
727 39. Shishkin, A. A. *et al.* Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat*
728 *Methods* **12,** 323–5 (2015).

40. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32,** 767–769 (2016).
41. Lomsadze, A., Burns, P. D. & Borodovsky, M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42,** e119–e119 (2014).
42. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24,** 637–644 (2008).
43. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25,** 955–964 (1997).
44. Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35,** 3100–3108 (2007).
45. Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput Biol* **7,** e1002195 (2011).
46. Conesa, A. *et al.* Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21,** 3674–6 (2005).
47. Goldberg, J. M. *et al.* Kinannote, a computer program to identify and classify members of the eukaryotic protein kinase superfamily. *Bioinformatics* **29,** 2387–94 (2013).
48. Saier, M. H., Jr., Tran, C. V. & Barabote, R. D. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34,** D181-6 (2006).
49. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **23,** 1061–7 (2007).
50. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).
51. Li, L., Stoeckert, C. J., Jr. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* **13,** 2178–89 (2003).
52. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22,** 2688–90 (2006).
53. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U A* **100,** 9440–5 (2003).
54. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
55. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26,** 139–140 (2010).
56. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* **8,** 1494–512 (2013).