

1 **Large-Scale Genome Mining Allows Identifying Neutral Polymorphisms and**
2 **Novel Resistance Mutations in Genes Involved in *Candida albicans* Resistance**
3 **to Azoles and Echinocandins**

4

5 *Emilie SITTERLE*^{1,2,3}, *Alix T. COSTE*⁴, *Thomas OBADIA*^{5,6}, *Corinne MAUFRAIS*⁵, *Murielle*
6 *CHAUVEL*¹, *Natacha SERTOOUR*¹, *Dominique SANGLARD*⁴, *Anne PUEL*^{7,8}, *Christophe*
7 *D'ENFERT*¹, *Marie-Elisabeth BOUGNOUX**^{1,3,9}

8

9 ¹Unité Biologie et Pathogénicité Fongiques, Département de Mycologie, Institut Pasteur,
10 USC2019 INRA, Paris, France

11 ²Université Paris Diderot, Sorbonne Paris Cité, Paris, France

12 ³Unité de Parasitologie-Mycologie, Service de Microbiologie clinique, Hôpital Necker-Enfants-
13 Malades, Assistance Publique des Hôpitaux de Paris (APHP), Paris, France

14 ⁴Institut de Microbiologie, Université de Lausanne et Centre Hospitalo-Universitaire, Lausanne,
15 Suisse

16 ⁵Hub de Bioinformatique et Biostatistique, Département Biologie Computationnelle, Institut
17 Pasteur, USR 3756 CNRS, Paris, France

18 ⁶Unité Malaria: parasites et Hôtes, Département Parasites et Insectes Vecteurs, Institut Pasteur,
19 Paris, France

20 ⁷Laboratoire de Génétique humaine des maladies infectieuses, Necker, INSERM U1163, Paris,
21 France

22 ⁸Université Paris Descartes, Institut Imagine, Paris, France

23 ⁹Université de Paris, Paris, France

24

25

26 *Corresponding author:

27 Marie-Elisabeth Bournoux, MD, PhD

28 Institut Pasteur, Unité Biologie et Pathogénicité Fongiques, Département Mycologie,

29 25 rue du Docteur Roux, F-75015 Paris, France ; Phone : +33 (1) 40 61 31 26

30 E-mail : elisabeth.bournoux-andremont@pasteur.fr

31

32 **Short running title:** Antifungal target gene polymorphisms in *Candida albicans*

33

34 **Synopsis**

35 **Background:** The genome of *Candida albicans* displays significant polymorphism. Point
36 mutations in genes involved in resistance to antifungals may either confer phenotypic resistance
37 or be devoid of phenotypic consequences.

38 **Objectives:** Catalog polymorphisms in azole- and echinocandin-resistance genes occurring in
39 susceptible strains in order to rapidly pinpoint relevant mutations in resistant strains.

40 **Methods:** Genome sequences from 151 unrelated *C. albicans* strains susceptible to fluconazole
41 and caspofungin were used to create a catalog of non-synonymous polymorphisms in genes
42 involved in resistance to azoles (*ERG11*, *TAC1*, *MRR1*, *UPC2*) or echinocandins (*FKSI*). The
43 potential of this catalog to reveal putative resistance mutations was tested in 10 azoles-resistant
44 isolates, including 1 intermediate to caspofungin. Selected mutations were analyzed by
45 mutagenesis experiments or mutational prediction effect.

46 **Results:** In the susceptible strains, we identified 126 amino-acid substitutions describing the
47 catalog of phenotypically-neutral polymorphisms. By excluding these neutral substitutions, we
48 identified 22 additional substitutions in the 10 resistant strains. Among these substitutions, 10
49 had already been associated with resistance. The remaining 12 were in Tac1p (n=6), Upc2p
50 (n=2) and Erg11p (n=4). Four out of the 6 homozygous substitutions from Tac1p (H263Y,
51 A790V, H839Y and P971S) conferred increases in azoles MIC, while no effects were observed
52 for those in Upc2p. Additionally, 2 homozygous substitutions (Y64H, P236S) had a predicted
53 conformation effect on Erg11p.

54 **Conclusions:** By establishing a catalog of neutral polymorphisms occurring in genes involved in
55 resistance to antifungal drugs, we provide a useful resource for rapid identification of mutations
56 possibly responsible for phenotypic resistance in *C. albicans*.

57

58 **Introduction**

59 *Candida albicans* is the most frequent human fungal pathogen causing a variety of infections in
60 both immunocompetent and immunocompromized individuals.^{1,2} Azoles and echinocandins are
61 the two main classes of antifungal agents currently used to treat severe *C. albicans* infections.^{3,4}
62 Emergence of azole- and echinocandin-resistant *C. albicans* isolates is a significant concern
63 because the number of systemic antifungal drugs available to fight *C. albicans* infections is
64 limited.⁵⁻⁹ Echinocandin resistance can occur through point mutations within hot-spot regions of
65 *FKSI* encoding the β -(1,3)-D-glucan synthase.^{5,10} Azole resistance can occur through various
66 mechanisms, such as alterations or overexpression of the drug target gene *ERG11* (encoding the
67 14- α -lanosterol demethylase) or increased expression of genes encoding drug efflux transporters:
68 *CDR1*, *CDR2*, or *MDR1*.¹¹⁻¹⁴ The expression of *CDR1/CDR2* or *MDR1* is regulated by the
69 transcription factors Tac1p and Mrr1p, respectively. Gain-of-function (GOF) mutations in the
70 corresponding *TAC1* and *MRR1* genes lead to overproduction of the pumps. Furthermore, up-
71 regulation of *ERG11*, via GOF mutations in the transcriptional factor *UPC2*, has been described
72 in resistant strains of *C. albicans*.¹⁵⁻¹⁷ Finally, azole resistance can arise from large-scale genetic
73 changes through genome rearrangements, including loss-of-heterozygosity events¹⁸⁻²⁰ as well as
74 segmental or whole chromosome aneuploidies leading to copy number variations of relevant
75 antifungal resistance genes.²¹⁻²³ The generation of such variants is facilitated by the high
76 genomic plasticity of *C. albicans*, especially under stress conditions when exposed to, antifungal
77 agents.²⁴⁻²⁸

78 The diploid genome of *C. albicans* displays a high level of natural heterozygosity with an
79 average of 1 heterozygous position every 204 bp.²⁹ When genomes of *C. albicans* isolates are
80 compared to the genome (haplotype A) of reference strain SC5314,^{30,31} homozygous variants are
81 observed every 235 to 1404 bp, depending on the genetic cluster from which isolates
82 originated.³² Therefore, genome resequencing of antifungal-resistant isolates does not allow to

83 rapidly distinguish the polymorphisms that underlie phenotypic resistance from those that may
84 result from natural diversity. Here, we have made the hypothesis that mining the whole genome
85 sequences from a large panel of epidemiologically unrelated antifungal-susceptible *C. albicans*
86 strains ²⁹ will allow establishing a comprehensive catalog of phenotypically-neutral non-
87 synonymous polymorphisms in the key genes involved in resistance to azoles (*ERG11*, *TAC1*,
88 *MRR1* and *UPC2*) and echinocandins (*FKSI*). This catalog could then be used to screen
89 genomes of resistant isolates for mutations linked to antifungal resistance. We addressed this
90 hypothesis through the analysis of 10 uncharacterized drug-resistant isolates and show that their
91 resistance can be rapidly associated to well-known as well as novel resistance mutations.

92

93 **Material and methods**

94 **Strains collection**

95 Two sets of *C. albicans* clinical strains were used in this study and are listed in supplementary
96 **Tables S1 and S2**. The first one consisted of 151 strains susceptible to fluconazole and
97 caspofungin, selected to represent the natural population diversity within *C. albicans*.²⁹ The
98 second set consisted of 10 azoles-resistant isolates, including 1 intermediate to caspofungin. The
99 antifungal susceptible *C. albicans* reference strain SC5314 was also used.³⁰

100

101 **Culture media**

102 *C. albicans* isolates were sub-cultured from frozen stock in liquid or solid YPD media (1% Yeast
103 extract, 2% Peptone, 2% Dextrose). Nourseothricin (300 µg/mL) was added to YPD agar plates
104 to select strains containing the *SAT1* nourseothricin-resistance gene when required. *Escherichia*
105 *coli* TOP10 (Life Technologies) or XL10-Gold cells (Agilent Technologies) were used as a host
106 for plasmid constructions and propagation. *E. coli* cells were cultured on LB (Luria-Bertani)
107 broth or plates, supplemented with ampicillin (100 mg/L) or kanamycin (50 mg/L) when
108 required.

109

110 **Minimal Inhibitory Concentrations (MICs) determination**

111 Antifungal MICs of *C. albicans* isolates were determined at 24h by using the Etest® method,
112 according to the manufacturer's recommendations and interpreted according to CLSI
113 breakpoints.³³

114

115 **Illumina sequencing and genome analysis**

116 Genomic DNA was extracted using the QIAamp® DNA Mini Kit (Qiagen, Courtaboeuf,
117 France), as described previously.²⁸ Libraries were prepared using the NEXTflex™ PCR-Free
118 DNA Sequencing kit (Illumina), according to the manufacturer's recommendations. Sequencing
119 was performed on a HiSeq2000 or HiSeq2500 platform generating 100 bp paired-end reads
120 (roughly 100x coverage).

121 Sequences and genomic variations were analyzed as described previously²⁹ using BWA v0.7.7³⁴
122 and the *C. albicans* reference strain SC5314 (genome assembly 22 version A22-s06-m01-r01³⁵).

123 The variant calling was performed with GATK v3.1³⁶ according to the best practices
124 recommended annotation for hard filters. In each strain, the entire ORF sequences of 5 genes
125 involved in azoles (*ERG11*, *TAC1*, *MRR1* and *UPC2*) and echinocandins (*FKSI*) resistance were
126 compared to that of reference strain SC5314. Only non-synonymous mutations were considered.

127 At each position, homozygous and heterozygous substitution frequency was calculated from the
128 set of the 151 *C. albicans* susceptible strains. Genotype results are listed in the supplementary

129 **Table S3.**

130

131 **Multilocus Sequence Typing**

132 All strains were typed using Multi Locus Sequence Typing (MLST), as described.³⁷

133

134 **Plasmid and mutant constructions**

135 To assess if the mutations are involved in phenotypic resistance to fluconazole, we have
136 constructed, for each new mutation identified in *TAC1* and *UPC2*, a series of integrative
137 plasmids that were introduced in *C. albicans* strains. The genotypes of the resulting strains are
138 listed in the supplementary **Table S4**. Primers used for plasmids and strains constructions are
139 listed in supplemental **Table S5**.

140 First, for strains harboring a single candidate mutation in *TAC1* (CEC723: 2369G>A
141 corresponding to amino-acid substitution A790V - 2369G>A/A790V) or *UPC2* (CEC 4522:
142 1960C>T/D654N and CEC718: 1970C>T/G657D) as well as SC5314 (wild type strain), the
143 genomic regions harboring *TAC1* or *UPC2* including the promotor and terminator regions were
144 PCR-amplified and cloned into the TOPO-BLUNT cloning vector (Invitrogen). Sequence of the
145 amplified fragments was verified by Sanger sequencing. The resulting plasmids were digested by
146 ad hoc restriction enzymes and the resulting fragments harboring a wild type or mutated *TAC1* or
147 *UPC2* gene were cloned into the CIpSAT plasmid. *E. coli* TOP10 (Invitrogen) competent cells
148 were used for cloning and maintaining the plasmid constructs.

149 Second, for strains harboring more than one candidate mutation in *TAC1*, we introduced each
150 mutation independently by site-directed mutagenesis of the *TAC1* gene obtained from *C.*
151 *albicans* strain SC5314. The plasmid harboring the wild type *TAC1* allele (CIpSAT-*P_{TAC1}-TAC1*)
152 was used as template for PCR amplification to introduce the following mutations
153 (787C>A/H263Y, 2469A>C/N823K, 2515G>A/H839Y, 2573C>T/C858Y, 2911G>A/P971S).
154 Primers were designed using the web-based QuikChange Primer Program available online
155 (www.agilent.com/genomics/qcpd). Reactions and cycling conditions were done according to the
156 manufacturer's instructions with the QuikChange Lightning Site-Directed Mutagenesis kit
157 (Agilent). *E. coli* XL10-Goold (Agilent) Ultracompetent cells were used for plasmids
158 transformation. Complete fragments were verified by Sanger sequencing.

159 Two *C. albicans* strains (DSY4578 or DSY3984), deleted for *TAC1* and *UPC2* (Supporting
160 Information **Table S4**) and susceptible to fluconazole, were used for transformation with the
161 resulting CIPSAT-*P_{TAC1}-TAC1* or CIPSAT-*P_{UPC2}-UPC2* derivatives.

162 CIPSAT derivatives were integrated in the *C. albicans* genome at the *RPS1* locus using the
163 lithium-acetate transformation protocol, as described.³⁸ Transformants were selected on YPD
164 medium supplemented with nourseothricin (300 mg/L). Plasmid integration was checked by
165 PCR with CIPSAT_checkF and CIPSAT_checkR primers (**Table S5**). *TAC1* or *UPC2* ORFs
166 were sequenced to confirm the occurrence of each mutation.

167

168 **Prediction of mutation effect in Erg11p**

169 The impact of mutations in *ERG11* was predicted by using different software and the 5v5z entry
170 ³⁹ from the Protein Data Base (<https://www.rcsb.org/>) as a reference for Erg11p 3D structure.

171 The solvent accessibility of amino-acids within Erg11p and the prediction of protein
172 thermodynamic stability ($\Delta\Delta G$) and thermostability changes (ΔT_m) upon single-site mutations
173 were evaluated using PoPMuSiCv3.1⁴⁰ and HoTMuSiCv0.5.⁴¹ The global effect of mutations
174 were evaluated using SNPmuSiC.⁴² All output files were generated using MuSiCv3.0r3
175 available online at <https://soft.dezyme.com/>. The amino-acid solvent accessibility ranged from 0
176 (buried) to 100 (fully accessible). The mutation effect on protein stability was classified as
177 destabilizing ($\Delta\Delta G > 0$ and $\Delta T_m < 0$) or stabilizing ($\Delta\Delta G < 0$ and $\Delta T_m > 0$). The predicted
178 mutational effect was classified as deleterious or neutral.

179

180 **Statistical analysis**

181 Genetic polymorphism diversity was defined as the number of non-synonymous SNPs that could
182 be identified from the antifungal-susceptible strains. This diversity was explored using
183 accumulation curves showing the number of distinct SNPs as a function of the number of strains.

184 Using a bootstrapping method, the 151 susceptible strains (N_S) were randomly permuted 10,000
185 times and for each permutation the cumulative number of mutations was counted as a function of
186 the increasing number of strains. The rate of new mutations (*i.e.* the number of new SNPs
187 introduced in a pool of strains when adding one new strain) was derived from these resampled
188 sets to assess the representativeness of susceptible strains from our experiment.

189 To assess the contribution of the 10 antifungal-resistant strains (N_R) to genetic diversity, the
190 above method was adapted to replace the 10 last susceptible strains from the random
191 permutations with random permutations of the N_R strains. This resulted in two sets of
192 bootstrapped distributions: the null with the N_S strains and the alternate with $N_S - N_R$ susceptible
193 strains and last N_R resistant strains. Finally, the number of new mutations identified in the last
194 N_R strains from the alternate distribution (that is the number of mutations from the pool of
195 resistant strains that cannot be identified in susceptible strains) was compared to the distribution
196 of the number of mutations in the null bootstrap using percentiles as credible intervals.

197

198 **Results**

199 **Constitution of a catalog of phenotypically-neutral non-synonymous polymorphisms in** 200 **genes involved in azole- and echinocandin-resistance.**

201 We took advantage of the availability of whole genome sequences of 151 susceptible *C. albicans*
202 strains to extensively investigate the sequence variations of genes involved in resistance to
203 azoles (*ERG11*, *TAC1*, *MRR1* and *UPC2*) and echinocandins (*FKS1*). Comparison of these
204 whole genome sequences with that of SC5314 led to the identification of 126 distinct non-
205 synonymous polymorphisms in these 5 genes (**Table 1**). As shown in **Figure 1**, we identified 38,
206 20, 15, 33 and 20 amino-acid substitutions in Tac1p, Erg11p, Upc2p, Mrr1p and Fks1p,
207 respectively. Sequencing statistics for the 151 susceptible *C. albicans* strains as well as their
208 MICs distributions are summarized in the supplementary **Table S1 and Figure 2**. For each
209 identified non-synonymous polymorphism, the genotypes obtained for all susceptible strains are
210 provided in **Table S3**.

211 Among the 126 non-synonymous mutations, 41% (52/126) were found both at the heterozygous
212 and homozygous state while 54% (69/126) were only found in the heterozygous state and 4%
213 (5/126) in the homozygous state. For each of these 126 mutations, we could determine the
214 homozygous and heterozygous mutation frequencies observed from the population of the 151
215 susceptible strains (**Table 1**). As shown in **Table 1**, a low frequency of nucleotide mutation
216 (<1%, corresponding to only one strain) was observed for 63 out of the 126 mutations (48
217 heterozygous and 15 homozygous).

218 In order to assess whether the 126 non-synonymous mutations identified using the set of 151
219 susceptible strains provided an exhaustive catalog of phenotypically-neutral polymorphisms in
220 the 5 selected genes, we quantified the number of new mutations identified with the increasing
221 number of strains analyzed (**Figure 3**). The number of newly identified mutations by addition of
222 each new strain from the “set of susceptible strains” quickly dropped below 1 after accounting

223 for the first ~20 strains, reaching a plateau and thus confirming that natural genetic diversity is
224 well captured by the chosen strain catalog (**Figure 3**).

225

226 **Assessment of the capacity to detect resistance mutations in resistant *C. albicans* isolates**
227 **using the catalog of phenotypically-neutral non-synonymous polymorphisms as an**
228 **exclusion set.**

229 A set of 10 clinical isolates exhibiting resistance to fluconazole and/or caspofungin were used to
230 test the capacity of the strain catalog to facilitate the identification of resistance mutations in
231 genes involved in antifungal resistance through exclusion of phenotypically-neutral
232 polymorphisms. The characteristics of the resistant isolates used here and results of *in vitro*
233 susceptibility testing are detailed in **Table 2** and in supplementary **Table S2**. All strains were
234 resistant to fluconazole and itraconazole (MIC \geq 8 mg/L and \geq 1 mg/L, respectively) with 7
235 exhibiting high-MIC levels to fluconazole (MIC $>$ 256 mg/L).³³ One strain (CEC4524) displayed
236 caspofungin MIC value of 0.5 mg/L and was classified as “intermediate”.³³

237 Comparison of the whole genome sequences of the 10 resistant isolates (average sequencing
238 depth of 123x, **Table S2**) to the genome sequence of the SC5314 reference strain led to the
239 identification of 45 non-synonymous mutations across the 5 genes (**Table S3**). Among these
240 mutations, 23 were present in the catalog of phenotypically-neutral polymorphisms (**Table 1**),
241 leaving 22 amino-acid changes that could be responsible for antifungal resistance. These
242 polymorphisms impact Erg11p (n=11), Tac1p (n=7), Upc2p (n=3) and Fks1p (n=1) and are listed
243 in supplementary **Table S6**. Mrr1p did not exhibit novel substitution.

244 Using a bootstrapping procedure, replacing the last 10 susceptible strains by the 10 resistant ones
245 led to a significant increase in the number of cumulative mutations, by comparison to that
246 observed in the catalog of susceptible strains only (**Figure 4A**). As shown in **Figure 4B**,
247 distribution of the number of novel mutations in the last 10 susceptible strains was significantly

248 lower than that observed for the 10 resistant strains, suggesting the appearance of additional
249 mutations to those expected only from natural non-synonymous polymorphisms (**Figure 4B**;
250 $p < 0.0001$).

251 Among the 22 substitutions predicted to be responsible for antifungal resistance based on their
252 absence from the catalog of phenotypically-neutral polymorphisms, 10 had already been
253 associated to azole or echinocandin resistance (**Table S6**). These include one mutation in Tac1p
254 (A736T),⁴³ seven in Erg11p (T123I, Y132H, R157K, S405F, G448E, G450E, R467I),^{44–58} one in
255 Upc2p (G648D^h; h=heterozygous state),⁵⁹ and one in Fks1p (R647G).⁵ The strain distribution
256 of these known mutations as well as the 12 remaining mutations selected by our approach is
257 summarized in **Table 2**.

258 Each of the 10 fluconazole-resistant strains harbored at least one known mutation that could
259 explain phenotypic resistance. However, the high level of resistance of some strains (CEC4504
260 and CEC4503) could not be explained by the unique presence of one heterozygous or
261 homozygous substitution in Erg11p, suggesting the presence of additional mutations involved in
262 resistance. This hypothesis was supported by a new comparison of the mutation accumulation
263 curves of the two sets of strains. We removed the 10 mutations already known to be involved in
264 resistance so that differences between susceptible and resistant strains are only restricted to the
265 newly-identified mutations (n=12). Again, the introduction of the 10 resistant strains caused an
266 increase in the number of cumulative mutations (**Figures 4C**). The number of new mutations
267 arising in the 10 resistant strains was still significantly different to the distribution of the number
268 of new mutations arising in 10 susceptible strains (**Figure 4D**; $p = 0.0014$). We predicted that at
269 least 4 out the 12 new mutations were distinct from the occurrence of phenotypically-neutral
270 polymorphisms.

271

272

273 **A subset of the newly identified mutations in *TAC1* and *UPC2* confer an increase azole**
274 **MICs.**

275 To explore the possibility that the identified mutations contributed to the high level of resistance
276 to fluconazole in our strains, we constructed mutant strains harboring each of the 6 homozygous
277 mutations in *TAC1* and 2 in *UPC2* (see Materials and Methods). No change in fluconazole MIC
278 was observed for the *UPC2* mutants (CEC5218 and CEC5220) by comparison to the *C. albicans*
279 strain transformed with the *UPC2* wild-type allele (CEC5216) (**Figure 5B, Table S4**).
280 Introduction of 4 out of the 6 *TAC1* mutations (CEC5222: 2369G>A/A790V; CEC5228:
281 787G>A/H263Y; CEC5230: 2515G>A/H839Y, and CEC5232: 2911G>A/P971S) resulted in an
282 increase of the fluconazole MIC (7.8-, 8-, 85- and 5.3-fold changes, respectively) and
283 itraconazole MIC (7.7-, 8-, 16- and 2.7-fold changes, respectively) compared to the *C. albicans*
284 strain transformed with the *TAC1* wild-type allele (CEC5221 or CEC5226) (**Table S4, S7 and**
285 **Figure 5A**).

286

287 **Newly identified mutations impact Erg11p structure.**

288 In order to evaluate the putative effect of newly identified amino-acid substitutions in Erg11p,
289 we explored the effect of each mutation using structural prediction with the newly described *C.*
290 *albicans* CYP51/Erg11 structural model protein 5v5z as reference.³⁹ The impact of the 3 newly
291 identified amino-acid substitutions (P30S^h, Y64H, P236S) identified in Erg11p was evaluated
292 and compared with the results obtained for mutations already known to be involved in
293 phenotypic fluconazole resistance. The PoPMuSiCv3.1 and HoTMuSiCv0.5 software were used
294 to predict $\Delta\Delta G$, ΔT_m and solvent accessibility of amino-acids and the SNPMuSiC software was
295 used to predict the overall effect of mutations. The heterozygous substitution D9V^h could not be
296 evaluated due to absence of the starting positions information in the protein 5v5z entry.

297 The heterozygous substitution P30S^h was predicted to have a neutral impact by SNPMuSiC and a

298 putative protein stabilizing effect ($\Delta\Delta G = -0.27$ and $\Delta T_m = 0.35$) using HoTMuSiC and
299 PoPMuSiC. These results were not in favor of a mutation involved in resistance. However, the 2
300 homozygous substitutions (Y64H and P236S) showed a low level of amino-acid accessibility <
301 20% (5.85 and 17.94) and an assumed protein destabilizing effect ($\Delta\Delta G = 2.03$ and 1.41 ; $\Delta T_m = -$
302 2.48 and -1.6 for Y64H and P236S, respectively). These results were corroborated using the
303 SNPMuSiC software that predicted a deleterious effect for both mutations. For comparison, the
304 mean values obtained for substitutions already known to be involved in fluconazole phenotypic
305 resistance and found in this study (T123I, Y132H, R157K, S405F, G448E, G450E, R467I) were
306 $\Delta\Delta G = 1.0$ and $\Delta T_m = -1.3$ with an average amino-acid accessibility level of 14.37%.

307

308 **Discussion**

309 In this study, we have taken advantage of a collection of 151 whole genomes of *C. albicans*
310 strains²⁹ susceptible to fluconazole and caspofungin in order to establish a comprehensive
311 catalog of phenotypically-neutral polymorphisms in the five main genes involved in azoles
312 (*ERG11*, *TAC1*, *UPC2* and *MRR1*) and echinocandins (*FKSI*) resistance in *C. albicans*. This
313 catalog includes 126 distinct non-synonymous substitutions and depicts with high confidence the
314 global polymorphisms existing in these genes within the natural population of *C. albicans*.
315 Indeed, this set of 151 *C. albicans* strains is representative of the global diversity of the *C.*
316 *albicans* population as these strains are distributed across the known genetic clusters and have
317 been isolated from various cases of infections (invasive or superficial) or during commensalism,
318 but also from various geographical areas.²⁹ Furthermore, these strains were selected for their
319 susceptibility to antifungals drugs. The distributions of their MICs to fluconazole and to
320 caspofungin were similar to the epidemiological cut-off (ECOFF) described for *C. albicans*
321 species (<https://mic.eucast.org>). Finally, the observation that sequencing of additional susceptible

322 isolates rarely leads to the identification of novel substitutions (**Figure 3**) strongly reinforced the
323 comprehensiveness of the established catalog of 126 natural non-synonymous substitutions.

324 We demonstrated the usefulness of the catalog to select putative resistance mutations among the
325 polymorphisms identified in genes involved in antifungal resistance for 10 azole- or
326 caspofungin-resistant strains. Indeed, using this catalog as a prediction guide, 23 non-
327 synonymous mutations identified in resistant isolates could be excluded from further analysis as
328 since they were also present in susceptible strains. By contrast, 22 non-synonymous mutations
329 were classified as putative resistance mutations, among which 10 were already involved in azole
330 or caspofungin resistance. Four newly described mutations in *TAC1* were experimentally
331 confirmed as contributing to azole resistance while two were predicted to enable azole resistance
332 by their significant impact on Erg11 protein structure. In contrast, we could not confirm that the
333 mutations identified in *UPC2* confer fluconazole resistance, suggesting that the strains harboring
334 these mutations have other resistance mechanisms.^{60,61} Despite these limitations, our approach
335 was highly efficient to highlight novel resistant mutations. Other groups have conducted
336 descriptions of phenotypically-neutral polymorphisms or mutations involved in azole resistance.
337 However, these descriptions were often restricted to a single gene, or on a smaller number of
338 clinical strains.⁶²⁻⁶⁵ For example, in the study performed by Spettel *et al.*, the amino-acid
339 substitution S937L in Tac1p, identified as a potentially causal resistance mutation, could have
340 been ruled out using the neutral polymorphisms catalog.⁶²

341 In our study, two amino-acid substitutions identified in Erg11p (Y64H, P236S) were predicted
342 with high confidence as impacting Erg11p protein structure. Interestingly, these positions were
343 already predicted to be vital for various antifungal interactions with the active site of Erg11p
344 from *C. albicans* or *Saccharomyces cerevisiae*.⁶⁶⁻⁶⁸ This suggests a likely role of these
345 substitutions in increasing azole resistance in our strains. Obviously, significant protein
346 structural changes in the Erg11p active site domain will occur if the amino-acids responsible for

347 key molecular interactions (with its substrate or inhibitors) are substituted.⁶⁶ However, the
348 specific contributions of individual *ERG11* mutations to azole resistance in *C. albicans* have not
349 been widely explored and very few mutations were demonstrated to be directly associated with
350 resistance.⁶⁹ The impact of a non-synonymous mutation on azole resistance could be
351 demonstrated by heterologous expression of the *ERG11* allele in different models, such as
352 *Saccharomyces cerevisiae*^{44,45} or, recently, in genetically engineered strain of *Candida*
353 *lusitaniae*.⁷⁰ The CRISPR-Cas9 system could also be used.⁷¹ These systems could be used to
354 formally confirm the involvement of the Y64H and P236S mutations in fluconazole resistance.
355 In conclusion, we have established a nearly comprehensive catalog that includes the
356 phenotypically-neutral non-synonymous polymorphisms in genes involved in azole- and
357 echinocandin-resistance in *C. albicans*. This catalog is now available to the community working
358 with this pathogen. Interestingly, it could be used to evaluate the possible role of neutral-
359 polymorphisms on MICs variability when resistance mutations were present in isolates. This
360 catalog is a powerful tool to prioritize mutations putatively involved in antifungal resistance and
361 to describe new mutations involved in azoles or caspofungin MIC increase in *C. albicans*.

362

363 **Acknowledgements**

364 **Funding**

365 E.S. was the recipient of a Ph.D. grant from DIM Malinf -Région Ile-de-France (DIM Malinf-
366 Région Ile-de-France 2014 to M-E.B. and A.P.).

367 This work has been supported by the French Government's Investissement d'Avenir program
368 (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases, ANR-10-
369 LABX-62-IBEID to C.d'E).

370 **Transparency declarations**

371 The authors declare no competing interests.

372 **Data availability**

373 Raw reads have been deposited at the NCBI Sequence Read Archive under BioProject ID
374 PRJNA432884 and PRJNA563993.

375

376 **References**

377 1. Noble SM, Johnson AD. Genetics of *Candida albicans*, a diploid human fungal pathogen.
378 *Annu Rev Genet* 2007; **41**: 193–211.

379 2. Kullberg BJ, Arendrup MC. Invasive Candidiasis. *N Engl J Med* 2015; **373**: 1445–56.

380 3. Cornely OA, Bassetti M, Calandra T, *et al.* ESCMID* guideline for the diagnosis and
381 management of *Candida* diseases 2012: non-neutropenic adult patients. *Clin Microbiol*
382 *Infect Off Publ Eur Soc Clin Microbiol Infect Dis* 2012; **18 Suppl 7**: 19–37.

383 4. Pappas PG, Kauffman CA, Andes DR, *et al.* Clinical Practice Guideline for the Management
384 of Candidiasis: 2016 Update by the Infectious Diseases Society of America. *Clin Infect Dis Off*
385 *Publ Infect Dis Soc Am* 2016; **62**: e1-50.

386 5. Lackner M, Tscherner M, Schaller M, *et al.* Positions and Numbers of *FKS* Mutations in
387 *Candida albicans* Selectively Influence *In Vitro* and *In Vivo* Susceptibilities to Echinocandin
388 Treatment. *Antimicrob Agents Chemother* 2014; **58**: 3626–35.

389 6. Al-Dorzi HM, Sakkijha H, Khan R, *et al.* Invasive Candidiasis in Critically Ill Patients: A
390 Prospective Cohort Study in Two Tertiary Care Centers. *J Intensive Care Med* 2018;
391 088506661876783.

392 7. Fisher MC, Hawkins NJ, Sanglard D, *et al.* Worldwide emergence of resistance to
393 antifungal drugs challenges human health and food security. *Science* 2018; **360**: 739–42.

394 8. Pfaller MA, Messer SA, Woosley LN, *et al.* Echinocandin and triazole antifungal
395 susceptibility profiles for clinical opportunistic yeast and mold isolates collected from 2010

396 to 2011: application of new CLSI clinical breakpoints and epidemiological cutoff values for
397 characterization of geographic and temporal trends of antifungal resistance. *J Clin Microbiol*
398 2013; **51**: 2571–81.

399 9. Whaley SG, Berkow EL, Rybak JM, *et al.* Azole Antifungal Resistance in *Candida albicans*
400 and Emerging Non-albicans *Candida* Species. *Front Microbiol* 2017; **7**. Available at:
401 <http://journal.frontiersin.org/article/10.3389/fmicb.2016.02173/full>.

402 10. Balashov SV, Park S, Perlin DS. Assessing resistance to the echinocandin antifungal drug
403 caspofungin in *Candida albicans* by profiling mutations in FKS1. *Antimicrob Agents*
404 *Chemother* 2006; **50**: 2058–63.

405 11. Lohberger A, Coste AT, Sanglard D. Distinct roles of *Candida albicans* drug resistance
406 transcription factors TAC1, MRR1, and UPC2 in virulence. *Eukaryot Cell* 2014; **13**: 127–42.

407 12. Sanglard D, Odds FC. Resistance of *Candida* species to antifungal agents: molecular
408 mechanisms and clinical consequences. *Lancet Infect Dis* 2002; **2**: 73–85.

409 13. Coste AT, Karababa M, Ischer F, *et al.* TAC1, transcriptional activator of CDR genes, is a
410 new transcription factor involved in the regulation of *Candida albicans* ABC transporters
411 CDR1 and CDR2. *Eukaryot Cell* 2004; **3**: 1639–52.

412 14. Morio F, Jensen RH, Le Pape P, *et al.* Molecular basis of antifungal drug resistance in
413 yeasts. *Int J Antimicrob Agents* 2017; **50**: 599–606.

414 15. Silver PM, Oliver BG, White TC. Role of *Candida albicans* transcription factor Upc2p in
415 drug resistance and sterol metabolism. *Eukaryot Cell* 2004; **3**: 1391–7.

416 16. Oliver BG, Song JL, Choiniere JH, *et al.* cis-Acting elements within the *Candida albicans*
417 ERG11 promoter mediate the azole response through transcription factor Upc2p. *Eukaryot*
418 *Cell* 2007; **6**: 2231–9.

- 419 17. MacPherson S, Akache B, Weber S, *et al.* *Candida albicans* zinc cluster protein Upc2p
420 confers resistance to antifungal drugs and is an activator of ergosterol biosynthetic genes.
421 *Antimicrob Agents Chemother* 2005; **49**: 1745–52.
- 422 18. Coste A, Turner V, Ischer F, *et al.* A mutation in Tac1p, a transcription factor regulating
423 CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate
424 antifungal resistance in *Candida albicans*. *Genetics* 2006; **172**: 2139–56.
- 425 19. Dunkel N, Blaß J, Rogers PD, *et al.* Mutations in the multi-drug resistance regulator
426 *MRR1*, followed by loss of heterozygosity, are the main cause of *MDR1* overexpression in
427 fluconazole-resistant *Candida albicans* strains. *Mol Microbiol* 2008; **69**: 827–40.
- 428 20. Dunkel N, Morschhäuser J. Loss of Heterozygosity at an Unlinked Genomic Locus Is
429 Responsible for the Phenotype of a *Candida albicans sap4* Δ *sap5* Δ *sap6* Δ Mutant. *Eukaryot*
430 *Cell* 2011; **10**: 54–62.
- 431 21. Selmecki A, Forche A, Berman J. Aneuploidy and isochromosome formation in drug-
432 resistant *Candida albicans*. *Science* 2006; **313**: 367–70.
- 433 22. Selmecki A, Gerami-Nejad M, Paulson C, *et al.* An isochromosome confers drug
434 resistance in vivo by amplification of two genes, ERG11 and TAC1. *Mol Microbiol* 2008; **68**:
435 624–41.
- 436 23. Selmecki AM, Dulmage K, Cowen LE, *et al.* Acquisition of aneuploidy provides increased
437 fitness during the evolution of antifungal drug resistance. *PLoS Genet* 2009; **5**: e1000705.
- 438 24. Bennett RJ, Forche A, Berman J. Rapid mechanisms for generating genome diversity:
439 whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harb Perspect Med*
440 2014; **4**: a019604
- 441 25. Wang JM, Bennett RJ, Anderson MZ. The Genome of the Human Pathogen *Candida*
442 *albicans* Is Shaped by Mutation and Cryptic Sexual Recombination. *mBio* 2018; **9**. e01205-
443 18

- 444 26. Ford CB, Funt JM, Abbey D, *et al.* The evolution of drug resistance in clinical isolates of
445 *Candida albicans*. *eLife* 2015; **4**: e00662.
- 446 27. Ene IV, Farrer RA, Hirakawa MP, *et al.* Global analysis of mutations driving
447 microevolution of a heterozygous diploid fungal pathogen. *Proc Natl Acad Sci U S A* 2018;
448 **115**: E8688–97.
- 449 28. Sitterlé E, Maufrais C, Sertour N, *et al.* Within-Host Genomic Diversity of *Candida*
450 *albicans* in Healthy Carriers. *Sci Rep* 2019; **9**: 2563.
- 451 29. Ropars J, Maufrais C, Diogo D, *et al.* Gene flow contributes to diversification of the major
452 fungal pathogen *Candida albicans*. *Nat Commun* 2018; **9**: 2253.
- 453 30. Gillum AM, Tsay EY, Kirsch DR. Isolation of the *Candida albicans* gene for orotidine-5'-
454 phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF*
455 mutations. *Mol Gen Genet MGG* 1984; **198**: 179–82.
- 456 31. van het Hoog M, Rast TJ, Martchenko M, *et al.* Assembly of the *Candida albicans* genome
457 into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol* 2007; **8**: R52.
- 458 32. Hirakawa MP, Martinez DA, Sakthikumar S, *et al.* Genetic and phenotypic intra-species
459 variation in *Candida albicans*. *Genome Res* 2015; **25**: 413–25.
- 460 33. CLSI. Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts-
461 Fourth Edition: M27-S4 and M27-A3. 2012.
- 462 34. Li H, Handsaker B, Wysoker A, *et al.* The Sequence Alignment/Map format and
463 SAMtools. *Bioinformatics* 2009; **25**: 2078–9.
- 464 35. Binkley J, Arnaud MB, Inglis DO, *et al.* The *Candida* Genome Database: the new
465 homology information page highlights protein similarity and phylogeny. *Nucleic Acids Res*
466 2014; **42**: D711-716.

467 36. McKenna A, Hanna M, Banks E, *et al.* The Genome Analysis Toolkit: a MapReduce
468 framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**:
469 1297–303.

470 37. Bounoux M-E, Tavanti A, Bouchier C, *et al.* Collaborative Consensus for Optimized
471 Multilocus Sequence Typing of *Candida albicans*. *J Clin Microbiol* 2003; **41**: 5265–6.

472 38. Walther A, Wendland J. An improved transformation protocol for the human fungal
473 pathogen *Candida albicans*. *Curr Genet* 2003; **42**: 339–43.

474 39. Keniya MV, Sabherwal M, Wilson RK, *et al.* Structure of CYP51 from the pathogen
475 *Candida albicans*. 2017. Available at:
476 <ftp://ftp.wwpdb.org/pub/pdb/data/structures/divided/pdb/v5/pdb5v5z.ent.gz>.

477 40. Dehouck Y, Kwasigroch J, Gilis D, *et al.* PoPMuSiC 2.1: a web server for the estimation of
478 protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*
479 2011; **12**: 151.

480 41. Pucci F, Bourgeas R, Rooman M. Predicting protein thermal stability changes upon point
481 mutations using statistical potentials: Introducing HoTMuSiC. *Sci Rep* 2016; **6**. Available at:
482 <http://www.nature.com/articles/srep23257>.

483 42. Ancien F, Pucci F, Godfroid M, *et al.* Prediction and interpretation of deleterious coding
484 variants in terms of protein structural stability. *Sci Rep* 2018; **8**. Available at:
485 <http://www.nature.com/articles/s41598-018-22531-2>.

486 43. Coste AT, Crittin J, Bauser C, *et al.* Functional Analysis of cis- and trans-Acting Elements
487 of the *Candida albicans* CDR2 Promoter with a Novel Promoter Reporter System. *Eukaryot*
488 *Cell* 2009; **8**: 1250–67.

489 44. Lamb DC, Kelly DE, White TC, *et al.* The R467K amino acid substitution in *Candida*
490 *albicans* sterol 14alpha-demethylase causes drug resistance through reduced affinity.
491 *Antimicrob Agents Chemother* 2000; **44**: 63–7.

492 45. Sanglard D, Ischer F, Koymans L, *et al.* Amino acid substitutions in the cytochrome P-
493 450 lanosterol 14 α -demethylase (CYP51A1) from azole-resistant *Candida albicans*
494 clinical isolates contribute to resistance to azole antifungal agents. *Antimicrob Agents*
495 *Chemother* 1998; **42**: 241–53.

496 46. Marichal P, Koymans L, Willemsens S, *et al.* Contribution of mutations in the
497 cytochrome P450 14 α -demethylase (Erg11p, Cyp51p) to azole resistance in *Candida*
498 *albicans*. *Microbiol Read Engl* 1999; **145**: 2701–13.

499 47. Chau AS, Mendrick CA, Sabatelli FJ, *et al.* Application of Real-Time Quantitative PCR to
500 Molecular Analysis of *Candida albicans* Strains Exhibiting Reduced Susceptibility to Azoles.
501 *Antimicrob Agents Chemother* 2004; **48**: 2124–31.

502 48. Favre B, Ryder NS, Didmon M. Multiple amino acid substitutions in lanosterol 14 α -
503 demethylase contribute to azole resistance in *Candida albicans*. *Microbiology* 1999; **145**:
504 2715–25.

505 49. Goldman GH, da Silva Ferreira ME, dos Reis Marques E, *et al.* Evaluation of fluconazole
506 resistance mechanisms in *Candida albicans* clinical isolates from HIV-infected patients in
507 Brazil. *Diagn Microbiol Infect Dis* 2004; **50**: 25–32.

508 50. Kakeya H, Miyazaki Y, Miyazaki H, *et al.* Genetic analysis of azole resistance in the
509 Darlington strain of *Candida albicans*. *Antimicrob Agents Chemother* 2000; **44**: 2985–90.

510 51. Lee M-K, Williams LE, Warnock DW, *et al.* Drug resistance genes and trailing growth in
511 *Candida albicans* isolates. *J Antimicrob Chemother* 2004; **53**: 217–24.

512 52. Löffler J, Kelly SL, Hebart H, S *et al.* Molecular analysis of cyp51 from fluconazole-
513 resistant *Candida albicans* strains. *FEMS Microbiol Lett* 1997; **151**: 263–8.

514 53. Perea S, Lopez-Ribot JL, Kirkpatrick WR, *et al.* Prevalence of Molecular Mechanisms of
515 Resistance to Azole Antifungal Agents in *Candida albicans* Strains Displaying High-Level

516 Fluconazole Resistance Isolated from Human Immunodeficiency Virus-Infected Patients.
517 *Antimicrob Agents Chemother* 2001; **45**: 2676–84.

518 54. Sanglard D, Bille J. Action of and resistance to antifungal agents. In: *Candida and*
519 *candidiasis*. In: Washington DC: American Society For Microbiology. Calderone RA, 2002;
520 370.

521 55. Wang H, Kong F, Sorrell TC, *et al*. Rapid detection of ERG11 gene mutations in clinical
522 *Candida albicans* isolates with reduced susceptibility to fluconazole by rolling circle
523 amplification and DNA sequencing. *BMC Microbiol* 2009; **9**: 167.

524 56. White TC. The presence of an R467K amino acid substitution and loss of allelic variation
525 correlate with an azole-resistant lanosterol 14 α demethylase in *Candida albicans*.
526 *Antimicrob Agents Chemother* 1997; **41**: 1488–94.

527 57. Xu Y, Chen L, Li C. Susceptibility of clinical isolates of *Candida* species to fluconazole and
528 detection of *Candida albicans* ERG11 mutations. *J Antimicrob Chemother* 2008; **61**: 798–
529 804.

530 58. Wu Y, Gao N, Li C, *et al*. A newly identified amino acid substitution T123I in the 14 α -
531 demethylase (Erg11p) of *Candida albicans* confers azole resistance. *FEMS Yeast Res* 2017;
532 **17**. Available at: [https://academic.oup.com/femsyr/article-](https://academic.oup.com/femsyr/article-lookup/doi/10.1093/femsyr/fox012)
533 [lookup/doi/10.1093/femsyr/fox012](https://academic.oup.com/femsyr/article-lookup/doi/10.1093/femsyr/fox012).

534 59. Flowers SA, Barker KS, Berkow EL, *et al*. Gain-of-function mutations in UPC2 are a
535 frequent cause of ERG11 upregulation in azole-resistant clinical isolates of *Candida*
536 *albicans*. *Eukaryot Cell* 2012; **11**: 1289–99.

537 60. Heilmann CJ, Schneider S, Barker KS, R *et al*. An A643T Mutation in the Transcription
538 Factor Upc2p Causes Constitutive ERG11 Upregulation and Increased Fluconazole
539 Resistance in *Candida albicans*. *Antimicrob Agents Chemother* 2010; **54**: 353–9.

- 540 61. Jensen RH, Astvad KMT, Silva LV, *et al.* Stepwise emergence of azole, echinocandin and
541 amphotericin B multidrug resistance *in vivo* in *Candida albicans* orchestrated by multiple
542 genetic alterations. *J Antimicrob Chemother* 2015; **70**: 2551–5.
- 543 62. Spettel K, Barousch W, Makristathis A, *et al.* Analysis of antifungal resistance genes in
544 *Candida albicans* and *Candida glabrata* using next generation sequencing. *PloS One* 2019;
545 **14**: e0210397.
- 546 63. Garnaud C, Botterel F, Sertour N, *et al.* Next-generation sequencing offers new insights
547 into the resistance of *Candida* spp. to echinocandins and azoles. *J Antimicrob Chemother*
548 2015; **70**: 2556–65.
- 549 64. Morio F, Pagniez F, Besse M, *et al.* Deciphering azole resistance mechanisms with a
550 focus on transcription factor-encoding genes TAC1, MRR1 and UPC2 in a set of fluconazole-
551 resistant clinical isolates of *Candida albicans*. *Int J Antimicrob Agents* 2013; **42**: 410–5.
- 552 65. Morio F, Loge C, Besse B, *et al.* Screening for amino acid substitutions in the *Candida*
553 *albicans* Erg11 protein of azole-susceptible and azole-resistant clinical isolates: new
554 substitutions and a review of the literature. *Diagn Microbiol Infect Dis* 2010; **66**: 373–84.
- 555 66. Debnath S, Addya S. Structural basis for heterogeneous phenotype of ERG11 dependent
556 Azole resistance in *C. albicans* clinical isolates. *SpringerPlus* 2014; **3**: 660.
- 557 67. Hargrove TY, Friggeri L, Wawrzak Z, *et al.* Structural analyses of *Candida albicans* sterol
558 14 α -demethylase complexed with azole drugs address the molecular basis of azole-
559 mediated inhibition of fungal sterol biosynthesis. *J Biol Chem* 2017; **292**: 6728–43.
- 560 68. Sagatova AA, Keniya MV, Wilson RK, *et al.* Structural Insights into Binding of the
561 Antifungal Drug Fluconazole to *Saccharomyces cerevisiae* Lanosterol 14 α -Demethylase.
562 *Antimicrob Agents Chemother* 2015; **59**: 4982–9.

563 69. Flowers SA, Colón B, Whaley SG, *et al.* Contribution of Clinically Derived Mutations in
564 *ERG11* to Azole Resistance in *Candida albicans*. *Antimicrob Agents Chemother* 2015; **59**:
565 450–60.

566 70. Accoceberry I, Rougeron A, Biteau N, *et al.* A CTG Clade *Candida* Yeast Genetically
567 Engineered for the Genotype-Phenotype Characterization of Azole Antifungal Resistance in
568 Human-Pathogenic Yeasts. *Antimicrob Agents Chemother* 2017; **62**. Available at:
569 <http://aac.asm.org/lookup/doi/10.1128/AAC.01483-17>.

570 71. Vyas VK, Barrasa MI, Fink GR. A *Candida albicans* CRISPR system permits genetic
571 engineering of essential genes and gene families. *Sci Adv* 2015; **1**: e1500248.

572

573 **Figure and table legends**

574 **Table 1: Characteristics of phenotypically-neutral non-synonymous polymorphisms in**
575 **genes involved in resistance to azoles and echinocandins.** The catalog includes 126 non-
576 synonymous phenotypically-neutral polymorphisms (ORF position and amino-acid
577 correspondence) detected in the 5 genes involved in resistance to azoles (*ERG11*, *TAC1*, *MRR1*
578 and *UPC2*) and echinocandins (*FKS1*) using genome sequences of 151 susceptible strains. For
579 each strain, the genotype at each position is compared to that of the reference strain SC5314 and
580 summarized by the homozygous and heterozygous substitution frequency. h=substitution found
581 only at the heterozygous state, *=indicate a stop codon.

582

583 **Table 2: Results of *in vitro* fluconazole, itraconazole, voriconazole and caspofungin**
584 **susceptibility testing and amino-acid substitutions detected for the 10 resistant strains of *C.***
585 ***albicans*.** Only amino-acid substitutions prioritized by our approach (n=22) as possibly being
586 involved in antifungal resistance are shown. These include resistance mutations previously
587 described in the literature and new amino-acid substitutions (boldface) in the protein Tac1p,
588 Upc2p, Erg11p and Fks1p. New resistant mutations validated in this study are marked with an
589 asterisk (*). P=patient; h=heterozygous mutation found in a single allele; FLC=fluconazole;
590 ITC=itraconazole; VOR=voriconazole; CAS=caspofungin.

591 *C. albicans* susceptible isolates to fluconazole, itraconazole, voriconazole and caspofungin were
592 defined as MIC \leq 2 mg/L, MIC \leq 0.125 mg/L MIC \leq 0.125 mg/L, MIC \leq 0.25 mg/L,
593 respectively.³³

594

595

596 **Figure 1: Schematic mapping of the phenotypically-neutral amino-acid substitutions in**
597 **Erg11p, Tac1p, Upc2, Mrr1p and Fks1p obtained from data of the 151 *C. albicans***
598 **fluconazole and caspofungin susceptible clinical strains.**

599 This figure appears in colour in the online version of *JAC* and in black and white in the printed
600 version of *JAC*.

601

602 **Figure 2: Distribution of fluconazole (A) and caspofungin (B) MICs for the 151 susceptible**
603 **clinical strains of *C. albicans*, determined by the Etest® method.**

604

605 **Figure 3: Rate of new mutations detected in the set of 151 susceptible strains.** The solid line
606 shows the mean number of new mutations that could be identified for each additional susceptible
607 strain considered (bootstrapping method). The grey envelopes correspond to 95% and 90%
608 credible intervals. Dashed line shows the threshold of 1 new mutation/strain.

609

610 **Figure 4: Contribution of resistant strains to genetic diversity of *C. albicans*.** Accumulation
611 curves of mutations as a function of the random selection of clinical strains of *C. albicans*,
612 accounting for all mutations present in resistant strains (A) or restricting those to undocumented
613 mutations only (C). Solid lines present the mean accumulation number of mutations with
614 increasing number of strains (black: random resampling of all susceptible strains ($N_S=151$);
615 dashed line: random samples of 141 susceptible strains (N_S) with final introduction of the 10
616 resistant strains (N_R)). Grey envelopes correspond to 95% and 90% credible intervals.

617 Bar plots (B, D) show the bootstrapped, expected distribution of number of new mutations in the
618 last 10 susceptible strains (N_S), with black diamond identifying the observed value when
619 considering resistant strains (N_R) instead under scenarios used for panels A and C, respectively.

620

621 **Figure 5: Fluconazole MICs determination of the constructed mutant strains using the**
622 **Etest® method. A. Mutants obtained for *TAC1*. B. Mutants obtained for *UPC2*.**

623 This figure appears in colour in the online version of *JAC* and in black and white in the printed
624 version of *JAC*.

625

626

627 **Table 1:**

Gene	ORF position	SC5314 reference strain genotype	Nucleotide substitution state/frequency (%)				Amino-acid substitution
			Heterozygous		Homozygous		
<i>TAC1</i>	261	TT	TG	8.4	GG	0.65	K87N
<i>TAC1</i>	268	CC	CT	0.65			A90T ^h
<i>TAC1</i>	310	AC			AA/CC	48/2.6	F104V
<i>TAC1</i>	323	CC	CT	16.1	TT	3.2	S108N
<i>TAC1</i>	391	GG	GT	23.2	TT	1.3	L131I
<i>TAC1</i>	445	AA	AT	1.3			L149I ^h
<i>TAC1</i>	508	TT	TC	20.65	CC	3.2	M170V
<i>TAC1</i>	510	CC	CA	9.0	AA	0.65	M170I
<i>TAC1</i>	520	TT	TC	9.0	CC	0.65	N174D
<i>TAC1</i>	566	AA	AG	28.8	GG	9.2	F189S
<i>TAC1</i>	593	TT	TC	1.3			H198R ^h
<i>TAC1</i>	596	CT			TT/CC	66.05/0.65	S199N
<i>TAC1</i>	617	CT			TT/CC	61.9/2.0	R206H
<i>TAC1</i>	620	AG			GG/AA	61.4/2.0	V207A
<i>TAC1</i>	628	CC	CA	0.65			A210S ^h
<i>TAC1</i>	1121	GG	GA	0.65			T374I ^h
<i>TAC1</i>	1130	GG	GA	12.3			A377V ^h
<i>TAC1</i>	1187	TC			CC/TT	65.6/1.3	N396S
<i>TAC1</i>	1924	TT	TC	1.3			T642A ^h
<i>TAC1</i>	1994	AA	AG	0.65			F665S ^h
<i>TAC1</i>	2083	GG	GA	13.5	AA	1.3	L695F
<i>TAC1</i>	2123	TT	TG	0.65			K708T ^h
<i>TAC1</i>	2140	TT	TC	0.65			S714G ^h
<i>TAC1</i>	2276	AA	AG	0.65			L759S ^h
<i>TAC1</i>	2291	CC	CT	2.0			S764N ^h
<i>TAC1</i>	2298	TT	TA	1.3			K766N ^h
<i>TAC1</i>	2316	AT			AATT	43.8/5.2	N772K
<i>TAC1</i>	2326	CT			TT/CC	67.75/0.65	D776N
<i>TAC1</i>	2344	TT	TC	0.65			T782A ^h
<i>TAC1</i>	2485	CG			GG/CC	47.4/9.2	E829Q
<i>TAC1</i>	2495	AA	AG	0.65			F832S ^h
<i>TAC1</i>	2606	CC	CT	2.6			R869Q ^h
<i>TAC1</i>	2684	AA	AG	14.2			I895T
<i>TAC1</i>	2687	TT	TC	49.4	CC	4.5	N896S
<i>TAC1</i>	2723	CC	CT	1.3	TT	0.65	C908Y
<i>TAC1</i>	2804	GA			GG/AA	74.8/1.3	S935L

629 **Table 1:** (Continued)

Gene	ORF position	SC5314 reference strain genotype	Nucleotide substitution state/frequency (%)			Amino-acid substitution	
			Heterozygous	Homozygous			
<i>TAC1</i>	2810	GG	GA	8.4		S937L ^h	
<i>TAC1</i>	2822	AG			GG/AA	20.8/5.8	S941P
<i>ERG11</i>	49	GG	GA	0.65			L17F ^h
<i>ERG11</i>	125	GG	GA	0.65			S42L ^h
<i>ERG11</i>	348	AT			AATT	60/7.1	D116E
<i>ERG11</i>	383	TG			TT/GG	78.6/2.0	K128T
<i>ERG11</i>	386	CC	CG	2.0			G129A ^h
<i>ERG11</i>	433	AA	AT	0.65			F145I ^h
<i>ERG11</i>	459	AA	AC	0.65	CC	3.2	D153E
<i>ERG11</i>	475	CC	CT	0.65			V159I ^h
<i>ERG11</i>	488	CC	CT	0.65			R163K ^h
<i>ERG11</i>	632	CC			TT	0.65	R211K
<i>ERG11</i>	798	TT	TG	12.5	GG	29.7	E266D
<i>ERG11</i>	886	CC	CA	1.3			E296* ^h
<i>ERG11</i>	982	GG	GA	0.65	AA	0.65	H328Y
<i>ERG11</i>	1007	TT	TC	1.3			E336G ^h
<i>ERG11</i>	1008	TT	TG	0.65			E336D ^h
<i>ERG11</i>	1217	GG	GA	0.65			P406L ^h
<i>ERG11</i>	1309	CC	CT	10.9	TT	4.5	V437I
<i>ERG11</i>	1325	GG	GA	1.3			S442F ^h
<i>ERG11</i>	1447	TT	TC	0.65			I483V ^h
<i>ERG11</i>	1462	CC	CT	24.6	TT	2.6	V488I
<i>UPC2</i>	73	GG	GA	0.65			L25F ^h
<i>UPC2</i>	76	TT	TC	0.65	CC	0.65	K26E
<i>UPC2</i>	122	GG	GT	0.65			T41N ^h
<i>UPC2</i>	203	CC	CT	31.2	TT	2.6	R68K
<i>UPC2</i>	425	AA	AC	48.6	CC	12.2	I142S
<i>UPC2</i>	431	GG	GA	1.3			T144I ^h
<i>UPC2</i>	569	CC	CT	22.2	TT	0.65	S190N
<i>UPC2</i>	661	CC			GG	0.65	A221P
<i>UPC2</i>	683	CC	CT	30.9	TT	3.9	S228N
<i>UPC2</i>	749	TT	TC	0.65			N250S ^h
<i>UPC2</i>	817	TT	TC	5.7	CC	1.3	T273A
<i>UPC2</i>	961	CC	CT	1.3			G321S ^h
<i>UPC2</i>	1024	GG	GC	1.3			L342V ^h

631 **Table 1:** (Continued)

Gene	ORF position	SC5314 reference strain genotype	Nucleotide substitution state/frequency (%)				Amino-acid substitution
			Heterozygous	Homozygous			
<i>UPC2</i>	1519	CC	CT	1.3			V507I ^h
<i>UPC2</i>	1904	AA	AC	0.65			L635W ^h
<i>MRR1</i>	47	GG	GT	3.6	TT	1.3	S16I
<i>MRR1</i>	56	CC	CT	2.0	TT	18.7	P19L
<i>MRR1</i>	79	GG	GA	27.1	AA	3.9	V27I
<i>MRR1</i>	179	TT	TC	1.3	CC	0.65	V60A
<i>MRR1</i>	218	CC	CA	3.6	AA	1.3	T73N
<i>MRR1</i>	223	GG	GA	2.0	AA	16.0	G75R
<i>MRR1</i>	304	GG	GA	2.0			A102T ^h
<i>MRR1</i>	387	TT	TG	0.65			N129K ^h
<i>MRR1</i>	394	CC	CA	0.65			Q132K ^h
<i>MRR1</i>	511	CC	CT	0.65			S171L ^h
<i>MRR1</i>	512	TT	TC	37.2	CC	18.2	S171P
<i>MRR1</i>	539	CC	CT	0.65			P180L ^h
<i>MRR1</i>	626	AA	AC	0.65			Q209P ^h
<i>MRR1</i>	742	TT	TG	28.0	GG	3.9	L248V
<i>MRR1</i>	1022	TT	TA	32.3	AA	11.8	V341E
<i>MRR1</i>	1226	GG	GA	1.3			G409E ^h
<i>MRR1</i>	1432	AA	AT	1.3			I478L ^h
<i>MRR1</i>	1675	AA	AT	1.3			T559S ^h
<i>MRR1</i>	1774	CC	CT	30.7	TT	4.5	L592F
<i>MRR1</i>	1925	CC	CT	0.65			A642V ^h
<i>MRR1</i>	1951	GG	GA	2.0			D651N ^h
<i>MRR1</i>	2128	GG	GA	0.65			E710K ^h
<i>MRR1</i>	2497	CC	CT	1.3			R833* ^h
<i>MRR1</i>	2750	CC	CT	9.4	TT	1.3	T917M
<i>MRR1</i>	2756	CC	CT	2.0			A919V ^h
<i>MRR1</i>	2768	CC	CT	9.4	TT	1.3	T923I
<i>MRR1</i>	2811	CC	CG	0.65	GG	18.1	N938K
<i>MRR1</i>	2900	GG	GA	0.65			G967D ^h
<i>MRR1</i>	3058	GG	GC	30.7	CC	7.9	E1020Q
<i>MRR1</i>	3082	GG	GA	0.65			E1028K ^h
<i>MRR1</i>	3096	TT	TA	2.1	AA	18.8	F1032L
<i>MRR1</i>	3110	CC	CT	2.1	TT	17.2	S1037L
<i>MRR1</i>	3265	GG	GT	0.65			A1089S ^h

633 **Table 1:** (Continued)

Gene	ORF position	SC5314 reference strain genotype	Nucleotide substitution state/frequency (%)				Amino-acid substitution
			Heterozygous		Homozygous		
<i>FKS1</i>	71	CC	CT	7.1	TT	0.65	G24E
<i>FKS1</i>	109	CC			TT	0.65	G37S
<i>FKS1</i>	340	AA	AC	0.65			Y114D ^h
<i>FKS1</i>	400	CC			TT	0.65	G134S
<i>FKS1</i>	457	CC	CT	0.65			A153T ^h
<i>FKS1</i>	859	CC	CT	0.65			A287T ^h
<i>FKS1</i>	943	CC	CT	0.65			G315S ^h
<i>FKS1</i>	1679	TT	CT	0.65			Y560C ^h
<i>FKS1</i>	1891	AA	AG	0.65			F631L ^h
<i>FKS1</i>	2355	TT	TA	0.65			E785D ^h
<i>FKS1</i>	2356	TT	TC	0.65			I786V ^h
<i>FKS1</i>	2496	CC	CA	0.65			M832I ^h
<i>FKS1</i>	3274	CC	CA	0.65			A1092S ^h
<i>FKS1</i>	3430	CC	CT	0.65			D1144N ^h
<i>FKS1</i>	4834	AA			TT	1.3	L1612M
<i>FKS1</i>	4873	CC	CT	0.65			D1625N ^h
<i>FKS1</i>	5512	GC			CC/GG	45.5/9.1	P1838A
<i>FKS1</i>	5531	TT	TG	0.65			N1844T ^h
<i>FKS1</i>	5609	CC	CT	1.3			R1870K ^h
<i>FKS1</i>	5657	GC			CC/GG	39.4/24.5	T1886S

634

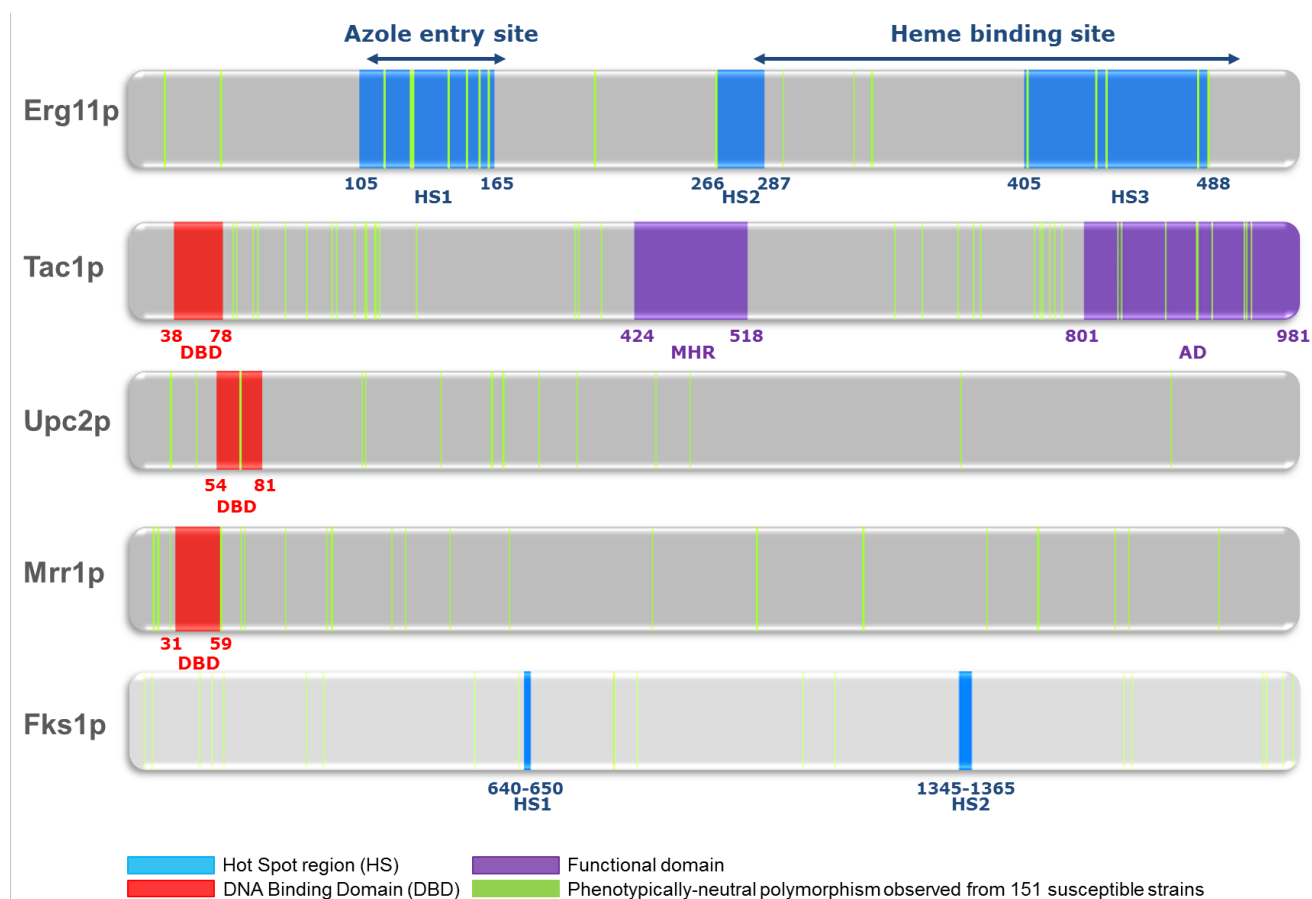
635

636 **Table 2:**

Strain	Origin	MIC (mg/L)				Amino-acid substitution			
		FLC	ITC	VOR	CAS	Erg11p	Tac1p	Upc2p	Fks1p
CEC718	P1	>256	>32	>32	0.047	Y132H, P236S , S405F	A736T	G657D	
CEC723	P2	16	1	0,094	0.064	P30S^h , S405F	A790V*		
CEC4504	P3	24	4	0,38	0.047	R157K ^h	H263Y* , H839Y* , C858Y		
CEC4503	P4	>256	>32	2	0.0125	D9V^h , R157K,	H263Y* , H839Y* , C858Y		
CEC4515	P5	>256	>32	>32	0.094	Y132H, G450E	H263Y* , P971S*	G648D ^h	
CEC4516	P5	>256	>32	>32	0.125	Y132H, G450E	H263Y* , P971S*		
CEC4521	P6	>256	12	4	0.064	Y132H, R467I	H263Y* , P971S*	D654N^h	
CEC4522	P6	>256	>32	12	0.19	Y132H, R467I	H263Y* , P971S*	D654N	
CEC4524	P6	>256	>32	3	0.5	Y64H , T123I, Y132H, R467I	H263Y* , P971S*	D654N^h	R647G
CEC4527	P7	8	3	0,125	0.047	G448E	A736T, N823K , C858Y		

637

638 **Figure 1:**



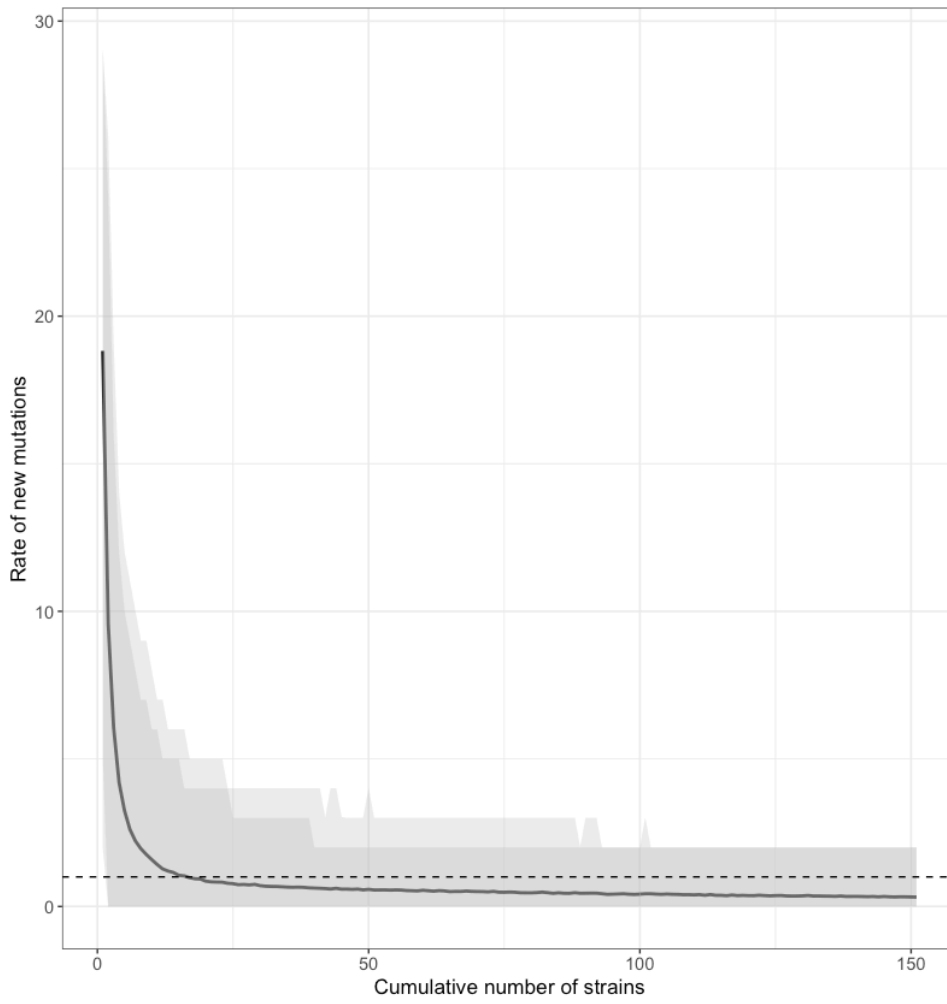
639

640 **Figure 2:**

641

642

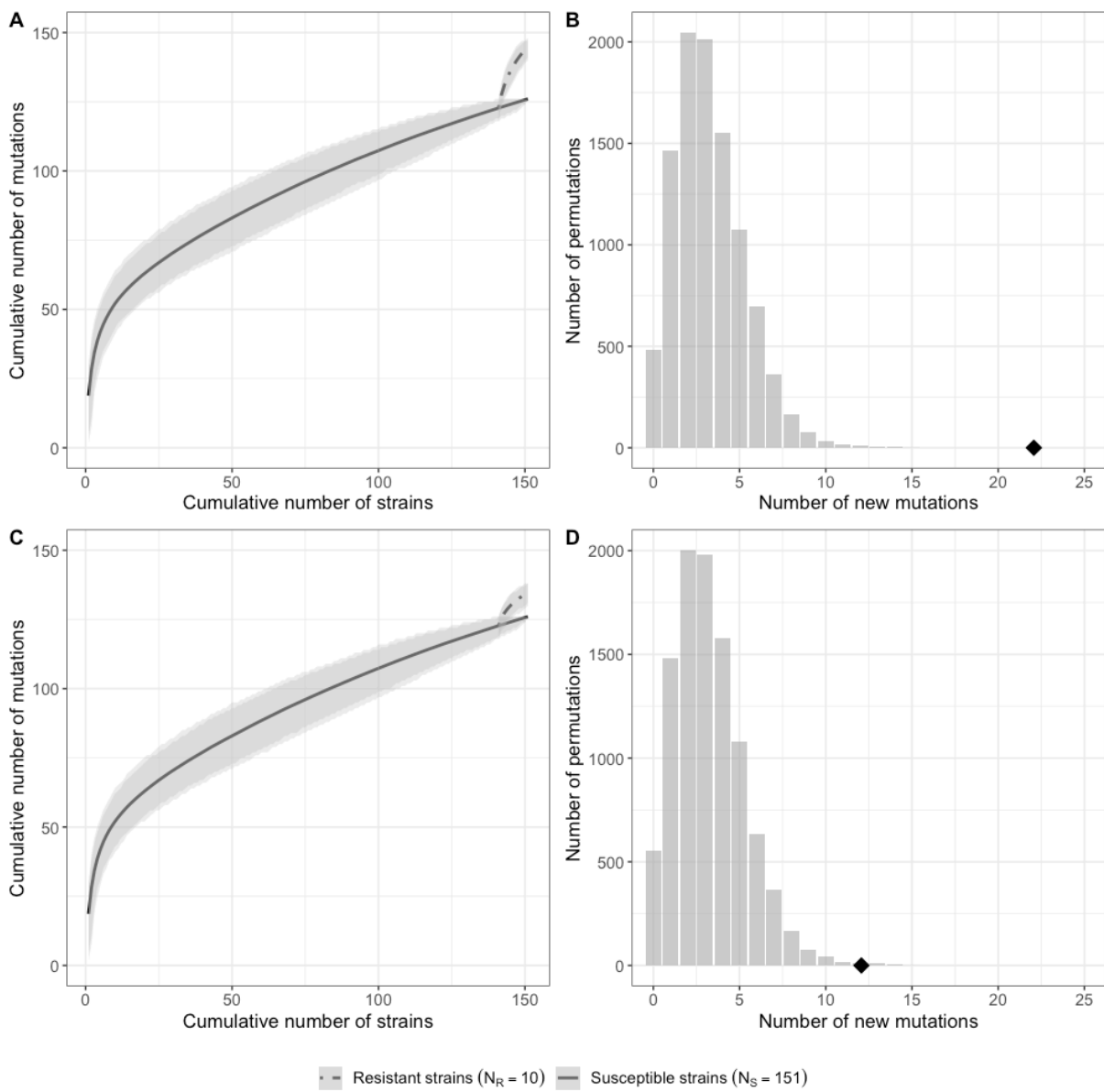
643 **Figure 3:**



644

645

646 **Figure 4:**



647

648

