

Thèse de doctorat
de l'Université Sorbonne Paris Cité
Préparée à l'Université Paris Diderot
Ecole doctorale BioSPC 562

Institut Pasteur : Unité de Biologie et Pathogénicité Fongiques

La candidose cutanéomuqueuse chronique: un modèle d'étude de l'adaptation génomique chez *Candida albicans*

Par Emilie Sitterlé

Thèse de doctorat en Microbiologie

Dirigée par Dr Marie-Elisabeth Bougnoux

Présentée et soutenue publiquement à Paris le 24 Septembre 2018

Président du jury : Silar, Philippe / Professeur / Université Paris-Diderot (LIDED)
Rapporteur : Fairhead, Cécile / Professeur / Université Paris Sud, Orsay (CNRS)
Rapporteur : Noël, Thierry / Professeur / Université de Bordeaux, (CNRS)
Examineur : Sanglard, Dominique / Professeur / Université de Lausanne,
Suisse (Institut de Microbiologie)
Directeur de thèse : Bougnoux, Marie-Elisabeth / MCU-PH / Université Paris Descartes
(Institut Pasteur, INRA)



Except where otherwise noted, this work is licensed under
<http://creativecommons.org/licenses/by-nc-nd/3.0/>

Résumé

Candida albicans est une levure commensale du tube digestif de l'Homme mais également un pathogène opportuniste responsable d'infections dont la gravité est fonction des défenses immunitaires de l'hôte. Au cours de l'interaction avec l'hôte, cette levure est confrontée à de nombreux stress environnementaux et immunitaires imposant des capacités d'adaptation rapides pour survivre. Le travail de thèse présenté dans ce manuscrit est consacré à l'étude de la diversité génomique de *C. albicans* chez l'individu sain mais également au cours de l'interaction pathologique de longue durée chez l'Homme. Dans une première partie nous montrons qu'il existe un niveau élevé d'hétérogénéité génomique essentiellement dus à des évènements de perte d'hétérozygoties, entre les isolats de *C. albicans* issus de prélèvements buccaux de porteurs sains et démontrons que celle-ci n'est pas le reflet de problèmes inhérents à la technique et à l'analyse des données de séquençage à haut débit. Puis, nous avons évalué dans quelle mesure le génome de *C. albicans* était capable d'évoluer, dans un contexte pathologique de longue durée grâce à l'étude de souches chronologiques provenant de patients atteints de candidose cutanéomuqueuse chronique. Nous montrons qu'il existe une importante dynamique dans l'apparition et l'élimination de variants phénotypiques et génotypique au cours de l'interaction pathologique chez l'hôte pouvant être la conséquence de l'adaptation des souches à l'interaction pathologique chronique. Un phénotype d'importance est la survenue de résistance de ces souches aux antifongiques. Une étude approfondie des gènes impliqués dans la résistance aux antifongiques a permis de décrire de nouvelles mutations dans les gènes *ERG11* et *TAC1* impliqués dans la résistance aux antifongiques azolés.

Abstract

Candida albicans is a common component of the human digestive tract and is considered the major opportunistic fungal pathogen. During interaction with the host, this yeast is confronted with numerous environmental and immune stresses imposing rapid adaptation capacities to survive. This manuscript aimed to study the genomic diversity of *C. albicans*, in healthy individuals and during long-term pathological interaction. In the first part of this work, we showed that there is a high level of genomic heterogeneity, especially linked to loss-of-heterozygosity, between isolates of *C. albicans* in oral samples from healthy carriers. Furthermore, we demonstrated that the heterogeneity observed is not reflective of technical problems nor associated to the high throughput sequencing data analysis. Then, we evaluated to which extent the *C. albicans* genome was able to evolve in a long-term pathological context, thanks to the study of chronological strains isolated from patients suffering from chronic mucocutaneous candidiasis. We have shown that there is an important dynamic in the appearance and the elimination of phenotypic and genotypic variants during the pathological interaction with the host. This may be the consequence of the adaptation of the strains to the chronic pathological interaction with its host. One of the interesting phenotype is the appearance of resistance to antifungal agents in these strains. An in-depth study of the genes involved in resistance to antifungal agents has enabled the detection and the description of new mutations in the *ERG11* and *TAC1* genes, involved in resistance to azole antifungals.

Remerciements

Cette thèse a pu être réalisée grâce à un financement DIM Malinf de la Région Ile de France.

J'aimerais tout d'abord remercier ma directrice de thèse, Madame Marie-Elisabeth Bougnoux pour m'avoir confié ce projet et permis de le mener à bien dans les meilleures conditions possibles. Merci pour son expertise, son enthousiasme ainsi que son soutien tout au long de ce travail de thèse, malgré ses nombreuses responsabilités.

Un grand merci à Christophe d'Enfert pour m'avoir accordé sa confiance et accueilli dans son Unité dans des conditions de travail privilégiées, mais également pour ses compétences humaines et ses précieux conseils.

J'exprime tous mes remerciements à l'ensemble des membres de mon jury : Madame Cécile Fairhead et Messieurs Thierry Noël et Dominique Sanglard pour avoir accepté de juger ce travail et Philippe Silar pour avoir également accepté de présider mon jury de thèse.

Je remercie les membres de l'unité que j'ai pu côtoyer tout au long de ces quatre années dans une ambiance de travail formidable.

Je remercie en particulier Adeline Feri pour son aide et plus particulièrement pour sa maîtrise de l'Anglais qui m'a permis de faire des progrès considérables. Merci à Sophie Bachellier pour ses encouragements et pour avoir pris le temps de lire cette thèse. Merci à Corinne Maufrais pour sa patience concernant mes multiples demandes saugrenues de nouveaux scripts. Je remercie également spécifiquement Natacha, Murielle, Mennat, Timéa, Mélanie, Jeanne, Guilhem, Sadri et Kévin qui m'ont entouré, conseillé ou aidé au quotidien dans le laboratoire ainsi que Christelle pour son aide et sa bonne humeur. Et bien sûr merci à tous les autres membres de l'équipe BPF...

Merci aux collaborateurs qui ont participé aux différents travaux : Matthieu Palayret, Alix Coste, Thomas Obadia, Anne Puel, Dominique Sanglard, Salomé Leibundgut-Landmann et son équipe.

Je remercie également Christiane Bouchier et Laurence Ma, de la plateforme génomique de l'Institut Pasteur pour leur aide technique dans les différents projets de séquençage (toujours urgents !) inclus dans mon travail de thèse.

Mes remerciements vont aussi à ma famille et mes amis qui, avec cette question récurrente, « quand est-ce que tu la soutiens cette thèse ? », bien qu'angoissante, m'ont permis de ne jamais dévier de mon objectif final. Merci à eux pour leur soutien dans les moments difficiles mais également dans les moments de détente.

Enfin un merci tout particulier à André et Gabriel pour leur soutien, compréhension et patience tout au long de ce parcours...

Sommaire

Abréviations	9
Préambule	11
INTRODUCTION	13
1. Le génome de <i>Candida albicans</i>	14
1.1 Organisation générale du génome de <i>C. albicans</i>	14
1.2 Variations du génome de <i>C. albicans</i>	17
1.2.1 Variations de structure du génome.....	17
1.2.1.1 Modification de la ploïdie.....	17
1.2.1.2 Modification de la taille des chromosomes, translocation et inversion	20
1.2.2 Variations de composition du génome	21
1.2.2.1 Hétérozygotie et mutations ponctuelles (SNPs)	21
1.2.2.2 Perte d'hétérozygotie (LOH).....	23
• Mécanismes dépendants de la recombinaison	23
• Mécanismes indépendants de la recombinaison	24
1.3 Méthodes d'étude de la variabilité génomique chez <i>C. albicans</i>	25
1.3.1 Techniques appliquées à l'analyse des variations génomiques	26
1.3.2 Techniques de génotypage moléculaire des souches de <i>C. albicans</i>	28
1.3.3 Séquençage du génome complet de <i>C. albicans</i>	29
1.3.3.1 Approches utilisant le clonage/séquençage.....	29
1.3.3.2 Approches de séquençage à haut débit	30
1.3.3.3 Analyse des données de séquençage à haut débit	32
2. Evolution et adaptation génomique de <i>C. albicans</i>	34
2.1 Evolution et diversité du génome de <i>C. albicans</i>	35
2.1.1 Evolution génétique de <i>C. albicans</i> au sein des <i>Saccharomycetales</i>	35
2.1.2 Diversité génétique de <i>C. albicans</i>	40
2.1.2.1 Structure de la population de <i>C. albicans</i>	40
2.1.2.2 Caractéristiques distinguant les clades	41
2.1.2.3 Variabilité génomique de <i>C. albicans</i>	45
2.1.2.3.1 Variations de la composition du génome de <i>C. albicans</i> et conséquences phénotypiques	45
• Hétérozygotie	45
• Perte d'hétérozygotie (LOH).....	46
• Conséquences moléculaires et phénotypiques de l'hétérozygotie/SNPs/LOH	46
2.1.2.3.2 Variabilité de la structure du génome de <i>C. albicans</i> et conséquences phénotypiques	47
• Variations dans le nombre de copies (CNVs) des ORFs et conséquences évolutives.....	47
• Réarrangements chromosomiques et conséquences phénotypiques	48
2.2 Adaptation de <i>C. albicans</i> à son environnement	50
2.2.1 Adaptation <i>in vitro</i> de <i>C. albicans</i> en réponse aux stress.....	51
2.2.1.1 Survenue de SNPs et de LOH en condition de stress et conséquences phénotypiques.....	51
• Cas particulier du stress induit par les antifongiques azolés.....	53
• Cas particulier du stress induit par les macrophages	54

2.2.1.2	Survenue d'aneuploïdies en condition de stress et conséquences phénotypiques.....	55
•	Cas particulier du stress induit par les antifongiques azolés.....	56
2.2.2	Adaptation <i>in vivo</i> de <i>C. albicans</i> au cours de l'interaction avec l'hôte.....	58
2.2.2.1	Evolution chez la souris.....	58
2.2.2.2	Evolution chez l'Homme.....	60
2.2.2.2.1	Diversité génétique de <i>C. albicans</i> au cours du commensalisme.....	60
2.2.2.2.2	Adaptation de <i>C. albicans</i> au cours de l'infection.....	62
•	Exemple du contexte pathologique retrouvé au cours de la mucoviscidose.....	62
•	Exemple de l'adaptation génomique de <i>C. albicans</i> au cours du traitement par fluconazole..	63
2.2.3	La Candidose cutané-Muqueuse Chronique (CCMC) un exemple d'interaction pathologique de très longue durée entre <i>C. albicans</i> et l'hôte.....	66
2.2.3.1	La CCMC : un modèle d'infection à long terme.....	66
2.2.3.2	La CCMC : un contexte d'interaction pathologique et inflammatoire particulier entre <i>C. albicans</i> et les cellules épithéliales.....	69
2.2.3.3	La CCMC : un contexte d'utilisation à long terme de molécules antifongiques.....	75
2.2.3.4	La CCMC : problématique de la résistance aux antifongiques.....	78
	Objectifs du travail de thèse.....	80
	RESULTATS-DISCUSSION.....	82
	1. Analyse des données de séquençage de génomes proches de <i>C. albicans</i>.....	83
1.1	Sélection des séquences de bonne qualité.....	83
1.2	Alignement des séquences.....	84
1.3	Analyse des SNPs/LOH.....	84
1.4	Analyse des variations du nombre de copie (CNVs).....	88
	2. Etude de la diversité génétique de <i>C. albicans</i> au cours du commensalisme.....	90
	Publication 1: Within-Host Genomic Diversity of <i>Candida albicans</i> in Healthy Carriers: Implication for Microevolution Studies. (Manuscrit soumis à Scientific Reports, SREP-18-31879) (Annexe 1) Emilie Sitterlé, Corinne Maufrais, Matthieu Palayret, Natacha Sertour, Christophe d'Enfert, Marie-Elisabeth Bognoux.....	92
	3. Diversité phénotypique et génomique de <i>C. albicans</i> au cours de l'interaction pathologique prolongée chez l'Homme.....	93
3.1	Introduction.....	93
3.2	Matériels et méthodes et Résultats.....	95
3.2.1	Constitution de la cohorte et caractérisation génotypique des souches de patients atteints de CCMC et de leur famille.....	95
3.2.1.1	Description des différentes collections de souches.....	95
3.2.1.2	Caractérisation génotypique des souches de patients CCMC et de leur famille.....	97
•	Caractérisation génotypique des isolats au sein d'un même patient.....	97
•	Caractérisation génotypique des isolats entre individus (sains et/ou malades d'une même famille).....	98
3.2.2	Caractérisation phénotypique des souches de patients atteints de CCMC et de leurs familles.....	100
3.2.2.1	Sensibilité des souches aux antifongiques.....	102
3.2.2.2	Cinétique de croissance des souches.....	104
3.2.2.3	Réponse aux stress environnementaux.....	105
3.2.2.4	Caractéristiques morphologiques et capacités d'invasion et de lyse cellulaire.....	110

•	Capacité de filamentation des souches	113
•	Capacité des souches à envahir la gélose	116
•	Capacité des souches à lyser les kératinocytes humains	117
3.2.3	Caractérisation génomique des souches de patients atteints de CCMC et de leur famille	119
3.2.3.1	Description des remaniements affectant le génome des souches isolées chez les patients CCMC et leur famille en comparaison avec les souches contrôles.	121
•	Remaniements génomiques de type aneuploïdie et LOH	121
•	Remaniements génomiques de type SNPs	127
3.2.3.2	Evolution phénotypique et génotypique des souches de patients CCMC et pistes futures ...	139
•	Evolution des souches chez le patient P1	139
•	Evolution des souches chez le patient P9	141
•	Evolution des souches entre individu malade et sains de la famille F8	143
3.3	Discussion	145
4.	<i>Analyse des mécanismes de résistance aux antifongiques</i>	151
	Publication 2 : A Useful Tool to Highlight New Resistance Mutations in Genes Involved in <i>Candida albicans</i> Resistance to Azoles and Echinocandines (Manuscrit en cours de préparation) (Annexe 2) Emilie Sitterlé, Alixe Coste, Thomas Obadia, Corinne Maufrais, Murielle Chauvel, Natacha Sertour, Anne Puel, Dominique Sanglard, Christophe d'Enfert, Marie-Elisabeth Bognoux.....	154
	<i>Conclusions-Perspectives</i>	155
	<i>Références bibliographiques</i>	158
	ANNEXES	172
	Publication 1	173
	Publication 2	218

Table des Illustrations

Figure 1 : Représentation schématique des 8 chromosomes de <i>C. albicans</i>	16
Figure 2 : Réarrangements génétiques survenant au cours du cycle parasexuel chez <i>C. albicans</i>	19
Figure 3 : Représentation schématique de la structure d'une MRS.	20
Figure 4 : Heatmap représentant la densité de SNPs hétérozygotes sur les 8 chromosomes de la souche de référence SC5314	23
Figure 5 : Mécanismes de survenue des LOH.....	25
Figure 6 : Schéma des étapes de préparation de l'ADN en vue d'un séquençage par technique Illumina	31
Figure 7 : Processus global d'analyse des génomes diploïdes.....	33
Figure 8 : Positionnement de l'espèce <i>C. albicans</i> au sein du règne fongique.	36
Figure 9 : Comparaison des caractéristiques des génomes de 4 sous-groupes majeurs du sous-embranchement des Saccharomycotina.....	37
Figure 10 : Dendrogramme réalisé à partir des DST (Diploid Sequence Typing) obtenus par typage MLST de 1410 souches de <i>C. albicans</i>	41
Figure 11 : Comparaison du fitness d'organismes euploïdes ou aneuploïdes en fonction des conditions environnementales.....	49
Figure 12 : Impact de différents stress sur le type et la fréquence de survenue d'événements de LOH.	53
Figure 13 : Détection et impact de l'acquisition de certaines aneuploïdies et LOH dans la résistance au fluconazole.....	57
Figure 14 : Défaut de l'immunité innée IL17 prédisposant aux CCMC.....	68
Figure 15 : Immunité innée anti <i>C. albicans</i> au niveau de la muqueuse buccale.....	72
Figure 16 : Photographies de patients présentant des mutations STAT1-GOF.	75
Figure 17 : Heatmap représentant la densité de SNPs hétérozygotes sur les 8 paires de chromosomes de la souche de référence SC5314.	86
Figure 18 : Quantification en kb des événements distaux de LOH par chromosome pour la souche SC5314.....	87
Figure 19 : Détermination de la taille des LOH.	88
Figure 20 : Détection des aneuploïdies par l'analyse des variations du nombre de copies (CNVs).....	89
Figure 21 : Composition de la cohorte de patients atteints de CCMC et des membres non malades de leur famille et cinétique d'isolement des 63 souches de <i>C. albicans</i> de la collection.	96
Figure 22 : Caractérisation phénotypique des 37 souches provenant des patients CCMC et de leur famille.. ...	104
Figure 23 : Impact de l'exposition à différents stress sur les souches les plus récentes des patients atteints de CCMC.....	108
Figure 24 : Impact de la caféine à 15 mM sur les 13 souches les plus récentes des patients CCMC et les 2 souches d'individus sains ainsi que la souche contrôle SC5314.....	109
Figure 25 : Morphologie des formes levure, hyphe et pseudohyphe de <i>C. albicans</i>	111
Figure 26 : Score appliqué à l'analyse de la morphologie des colonies (capacité de filamentation et aspect). .	113
Figure 27 : Caractérisation morphologique et capacité de lyse cellulaire des 37 souches provenant de patients CCMC et de leur famille.....	116
Figure 28 : Arbre phylogénétique réalisé à partir des données de séquençage du génome complet des 37 souches de patients CCMC et de leur famille.	120
Figure 29 : Quantification en kb de la taille des événements de LOH distaux	123
Figure 30 : Comparaison de la quantité moyenne (en pb) de LOH distaux détecté pour les 13 souches provenant de patients atteint de CCMC (en jaune) et les 144 souches contrôles (en bleu).....	123
Figure 31 : Comparaison de la présence de larges événements de LOH de type BIR chez les souches de patients CCMC.....	125
Figure 32 : Survenue d'événements de LOH au cours du temps sur les séries chronologiques de souches provenant des patients CCMC.....	126
Figure 33 : Algorithme employé pour détecter les SNPs spécifiques aux génomes des souches isolées chez les patients atteints de CCMC (ici n=9) en comparaison avec des génomes de souches contrôles (n=10).....	127

Figure 34 : Arbre phylogénétique réalisé à partir des données de séquençage du génome des souches de patients CCMC et de leur famille en comparaison avec les 144 génomes contrôles.....	132
Figure 35 : Représentation graphique des Odds ratio obtenus pour les gènes identifiés par l'approche de régression linéaire de type Lasso.	134
Figure 36 : Représentation graphique des score obtenus (coefficient de Gini normalisé) pour les gènes identifiés par l'approche de type classification (Random Forrest).....	135
Figure 37 : Evolution phénotypique et génomique de 4 souches provenant du patient P1.	140
Figure 38 : Evolution phénotypique et génomique des 4 souches provenant du patient P9.	142
Figure 39 : Comparaison de l'évolution des génomes de <i>C. albicans</i> chez un individu sain et malade.....	144
Tableau 1 : Résumé des principales caractéristiques associées aux clades MLST de <i>C. albicans</i>	44
Tableau 2 : Déficits immunitaires primitifs impliqués dans l'immunité innée IL-17 chez les patients présentant une CCMC « isolée » ou « syndromique »	67
Tableau 3 : Principales structures et principaux gènes de <i>C. albicans</i> codant pour des protéines impliquées dans l'interaction avec les cellules épithéliales de l'hôte.....	71
Tableau 4 : Prise en charge thérapeutique des patients ayant des mutations STAT1-GOF.....	77
Tableau 5 : Caractéristiques MLST des 63 souches de <i>C. albicans</i> isolées chez les patients CCMC et leur famille	98
Tableau 6 : Liste des 37 souches de <i>C. albicans</i> , provenant de patients atteints de CCMC et de leur famille, sélectionnées au sein de notre collection pour le séquençage du génome.	100
Tableau 7 : Caractérisation phénotypique des souches dans différentes conditions.....	101
Tableau 8 : Résumé des principales caractéristiques des données de séquençages obtenues, en comparaison avec la souche de référence SC5314, pour les 37 génomes de patients CCMC et leur famille.....	121
Tableau 9 : Lien ontologique entre les 50 gènes communs altérés dans le génome de la majorité des souches de patients CCMC.....	129
Tableau 10 : Caractéristiques statistique des gènes identifiés par approche de régression linéaire de type Lasso.	134
Tableau 11 : Liste des 19 gènes identifiés par les deux approches (Lasso et Random Forrest).	137
Tableau 12 : Caractéristiques des 22 souches de <i>C. albicans</i> issues des patients CCMC présentant une résistance acquise aux azolés et/ou aux échinocandines.....	153

Abréviations

ABhet	Ratio allélique (A/B) d'une position hétérozygote
AD	Autosomique dominant
ADNr	ADN ribosomique
AGPI	Acide gras polyinsaturé
AMP	Peptide antimicrobien
AND	Anidulafungine
AR	Autosomique récessif
BAM	Binary Alignment Map
BIR	Break Induce Replication
CASP	Caspofungine
CCMC	Candidose Cutané-Muqueuse Chronique
CEN	Centromère
CGD	Candida Genome Database
CGH	Comparative Genomic Hybridation Array
CHEF	Contour-clamped Homogeneous Electric Field
Chr	Chromosome
i(5L)	Isochromosome du bras gauche du chromosome 5
CLR	Lectines de type-C
CMI	Concentration Minimale Inhibitrice
CNVs	Copy Number Variations
COP	Candidose oropharyngée
CR	Rouge Congo
CW	Calcofluor White
DO	Densité optique
DST	Diploid Sequence Type
DT	Doubling Time
FACS	Fluorescence-assisted cell sorting
5-FC	5-fluorocytosine
Fluco	Fluconazole
GATK	Genome Analysis Tool Kit
GO	Gene Ontology
GOF	Gain of Fonction
GPI	Protéines glycosyl-phosphatidylinositol
Itra	Itraconazole
ITS	Internal Transcribed Spacer
LDH	Lactate déshydrogénase
LiCl	Chlorure de lithium
LOH	Loss of heterozygosity
LTR	Long Terminal Repeat
MLMT	Multilocus Microsatellite Typing
MLST	Multi Locus Sequence Typing
MRS	Major Repeat Sequence
MTL	Mating Type-Like

mTOR	mammalian Target Of Rapamycin
NETs	Neutrophil Extracellular Traps
NGS	Next Generation Sequencing
NS	Non synonyme
ORF	Open Reading Frame
PKA	Protein kinase-A
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
RPS	Repetitive Sequence
SAM	Sequence Alignment Map
SD	Milieu synthétique
SDS	Sodium dodecyl sulfate
SNP	Single Nucleotide Polymorphism
TE	Transposable Element
TLO	Telomere-associated genes
TLR	Toll-Like Receptor
Vori	Voriconazole
YPD	Yeast extract-Peptone-Dextrose

Préambule

Candida albicans est un redoutable pathogène opportuniste responsable d'infections sévères des muqueuses et de candidoses invasives chez les patients immunodéprimés. Les candidoses invasives sont associées à une mortalité extrêmement élevée de l'ordre de 40 à 50% (Pfaller and Diekema, 2007). Or, le nombre d'immunodéprimés ne cesse d'augmenter alors que les progrès diagnostiques et thérapeutiques pour identifier et traiter les infections fongiques, restent faibles (Brown et al., 2012). Plus de 50% des infections fongiques retrouvées chez les individus immunodéprimés sont dues à la seule levure *C. albicans*, qui est, par ailleurs, la principale levure commensale du tractus oro-digestif chez l'adulte sain (Ghannoum et al., 2010). La colonisation digestive et la traversée de la barrière intestinale par *C. albicans* figurent parmi les éléments clefs impliqués dans le développement des infections des muqueuses et des candidoses systémiques chez les patients immunodéprimés (Eggimann et al., 2003). La transition entre cet état de commensalisme et l'initiation d'une infection est ainsi influencée par la réponse immunitaire de l'hôte et la capacité de la levure à s'adapter et à échapper aux mécanismes de défense de l'hôte (Cheng et al., 2012; Richardson and Moyes, 2015). La capacité de *C. albicans* à coloniser les différents sites du corps humain implique le fait que même une interaction dite « saine », entre cette levure commensale et son hôte, est basée sur un équilibre très complexe (Hube, 2004) qui est le résultat d'une longue coévolution de ce pathogène opportuniste chez son hôte (Cottier and Pavelka, 2012).

C. albicans est le premier champignon pathogène dont le génome a été entièrement séquencé (Jones et al., 2004). La disponibilité de sa séquence a sans aucun doute permis de faire avancer nos connaissances sur la biologie de ce pathogène opportuniste, sa pathogénicité et ses mécanismes adaptatifs, ce qui fait de cette levure, un organisme modèle pour l'étude d'autres champignons pathogènes (Kabir et al., 2012). En effet, tous les organismes vivants ont la nécessité de s'adapter à leur environnement naturel mais également d'initier rapidement des stratégies d'adaptation pour survivre en cas de situation de stress (Forche, 2014).

C'est également le cas chez *C. albicans* dont la plasticité de son génome permet de générer de façon rapide de la diversité génomique en condition de stress (Selmecki et al., 2010). Des événements majeurs de remaniements génomiques comme des aneuploïdies ou des pertes d'hétérozygotie (LOH), ont été observés au sein d'isolats cliniques (Ford et al., 2015; Hirakawa et al., 2015; Ropars et al., 2018). Ces événements ont largement été étudiés dans le cadre de l'adaptation de *C. albicans* en réponse aux antifongiques mais également en réponse à diverses modifications environnementales *in vitro* et *in vivo* (Forche, 2014). En revanche, peu d'études se sont portées sur la diversité génomique de *C. albicans* chez l'Homme.

Dans le premier chapitre de l'Introduction de ce manuscrit, nous décrivons la structure du génome de *C. albicans* et sa variabilité puis dans un deuxième chapitre introductif, nous aborderons l'évolution du génome de *C. albicans* associé aux mécanismes mis en place par cette levure pour répondre *in vitro* et *in vivo* aux différents stress auxquels elle peut être soumise. Enfin, nous présenterons un modèle spécifique d'interaction à long terme de *C. albicans* avec son hôte que l'on observe chez des patients atteints de Candidose Cutanéomuqueuse Chronique (CCMC) (Kirkpatrick, 2001). Cette dernière constitue un modèle privilégié d'étude des mécanismes génomiques adaptatifs durant une interaction pathologique hôte-pathogène. L'environnement particulier dans lequel évolue *C. albicans* au cours de cette pathologie, du à la réponse immunitaire et inflammatoire inappropriée de l'hôte, mais également à l'utilisation prolongée d'antifongiques, est propice à l'observation de modifications génomiques.

INTRODUCTION

1 Le génome de *Candida albicans*

1.1 Organisation générale du génome de *C. albicans*

Candida albicans est dans la grande majorité des cas un organisme diploïde, bien que des souches haploïdes, tétraploïdes ou aneuploïdes ont été obtenues *in vitro* et observées *in vivo* (Hickman et al., 2013).

Le séquençage du génome de *C. albicans* a débuté dans les années 90 à l'Institut Sanger au Royaume Uni (Tait et al., 1997), avec la première description du chromosome 7 (Chibana et al., 1998). Plusieurs assemblages préliminaires ont ensuite été réalisés, mais ce n'est qu'en 2004, qu'une équipe du « Stanford Genome Technology Center » aux Etats-Unis, a publié la première description du génome complet d'une souche de *C. albicans* (Jones et al., 2004), la souche SC5314 isolée de l'hémoculture d'un patient présentant une infection disséminée (Gillum et al., 1984). Cette première version du génome, appelée « Assemblage 19 », est un assemblage diploïde. Un deuxième assemblage haploïde a ensuite été généré en éliminant de façon arbitraire un des deux homologues des 146 paires de contigs de l'assemblage 19. Des analyses complémentaires combinant cartographie des contigs sur les chromosomes et utilisation de données de séquence du génome de *C. dubliniensis*, espèce proche de *C. albicans*, ont abouti en 2007 à la publication de l'assemblage 21, assemblage haploïde constitué de 8 séquences chromosomiques (van het Hoog et al., 2007).

Finalement, c'est l'assemblage 22 qui est maintenant la séquence de référence de *C. albicans* disponible dans la base de données CGD (Candida Genome Database ; <http://www.candidagenome.org>). Cette version du génome, décrite en 2013 par Muzzey et al., a été obtenue en associant les données de séquençage Illumina de la souche SC5314 et de souches congéniques homozygotes pour les différents chromosomes (Muzzey et al., 2013). Cette approche a permis d'obtenir des séquences phasées de chacun des deux haplotypes -nommés A et B- pour les différents chromosomes. La correction de la séquence de référence du génome de SC5314 est un projet en évolution. Les axes d'améliorations portent actuellement sur les régions intergéniques, les séquences répétées ainsi que sur les duplications segmentales (Skrzypek et al., 2017).

Depuis le séquençage du génome complet de SC5314, d'autres génomes de souches de *C. albicans* ont été décrits. Butler et collaborateurs (2009) ont rapporté la séquence du génome complet de la souche WO-1 qui se caractérise par sa capacité à subir une conversion phénotypique dite « white-opaque » à haute fréquence (Butler et al., 2009). Plus récemment, plusieurs génomes de souches cliniques ont été séquencés et sont maintenant disponibles (Ford et al., 2015; Hirakawa et al., 2015; Ropars et al., 2018).

Le génome de *C. albicans* a une taille de 15,845 méga bases (Mb) sous forme haploïde. Il est composé de 8 paires de chromosomes. Sept de ces chromosomes sont numérotés de 1 à 7 selon la décroissance de leur taille, de 3,2 Mb pour le Chr1 à 0,95 Mb pour le Chr7. Le chromosome R, qui contient l'ADN ribosomique (ADNr), possède une taille proche de celle du Chr1 mais qui peut naturellement varier selon le nombre de répétitions d'ADNr (de 25 à 175 répétitions en tandem) (Jones et al. 2004; Van het Hoog et al. 2007). Chaque chromosome possède un petit centromère régional (3 à 5 Kb) et des télomères couvrant plusieurs centaines de paires de bases protégeant ses extrémités. Au niveau des extrémités télomériques se trouvent un ensemble de gènes appelés « TLO » (TeLOmere-associated genes) (*TLO1* à *TLO16*) qui ont un rôle dans la régulation transcriptionnelle de gènes impliqués dans l'adaptation chez l'hôte et la virulence (Flanagan et al., 2018; Haran et al., 2014; Sullivan et al., 2015; van het Hoog et al., 2007) (**Figure 1**).

En plus des répétitions d'ADNr présentes sur le ChrR, tous les chromosomes, à l'exception du Chr3, contiennent une à deux séquences MRS « Major Repeat Sequence » qui est composée d'unités répétitives allant de 50 à 130 kb (Lephart et al., 2005).

On retrouve également d'autres types de séquences répétées dispersées dans le génome, résultant de la présence de transposons et rétrotransposons. En effet *C. albicans* contient dans son génome des éléments transposables (TEs), qui sont des fragments d'ADN contenant plusieurs ORFs (open reading frames) capables de se dupliquer et de s'insérer dans le génome par un mécanisme de transposition. Ces éléments peuvent se déplacer dans le génome et être transmis verticalement. Les TEs présents dans le génome de *C. albicans* sont essentiellement de type rétrotransposons, c'est à dire dont le mécanisme de transposition passe par une étape ARN et qui sont flanqués aux extrémités 5' et 3' de longues répétitions terminales que l'on appelle LTR (Long terminal repeat). Il existe 34 familles distinctes de LTRs chez *C. albicans*.

Ils appartiennent respectivement aux groupes Ty1/copia et Ty3/gypsy qui ont fait l'objet d'études approfondies chez *Saccharomyces cerevisiae* (Zhang et al., 2014). A noter que dans la plupart des cas ces LTRs ne sont pas fonctionnels (Solo LTRs ou fragments de LTRs) (Zhang et al., 2014).

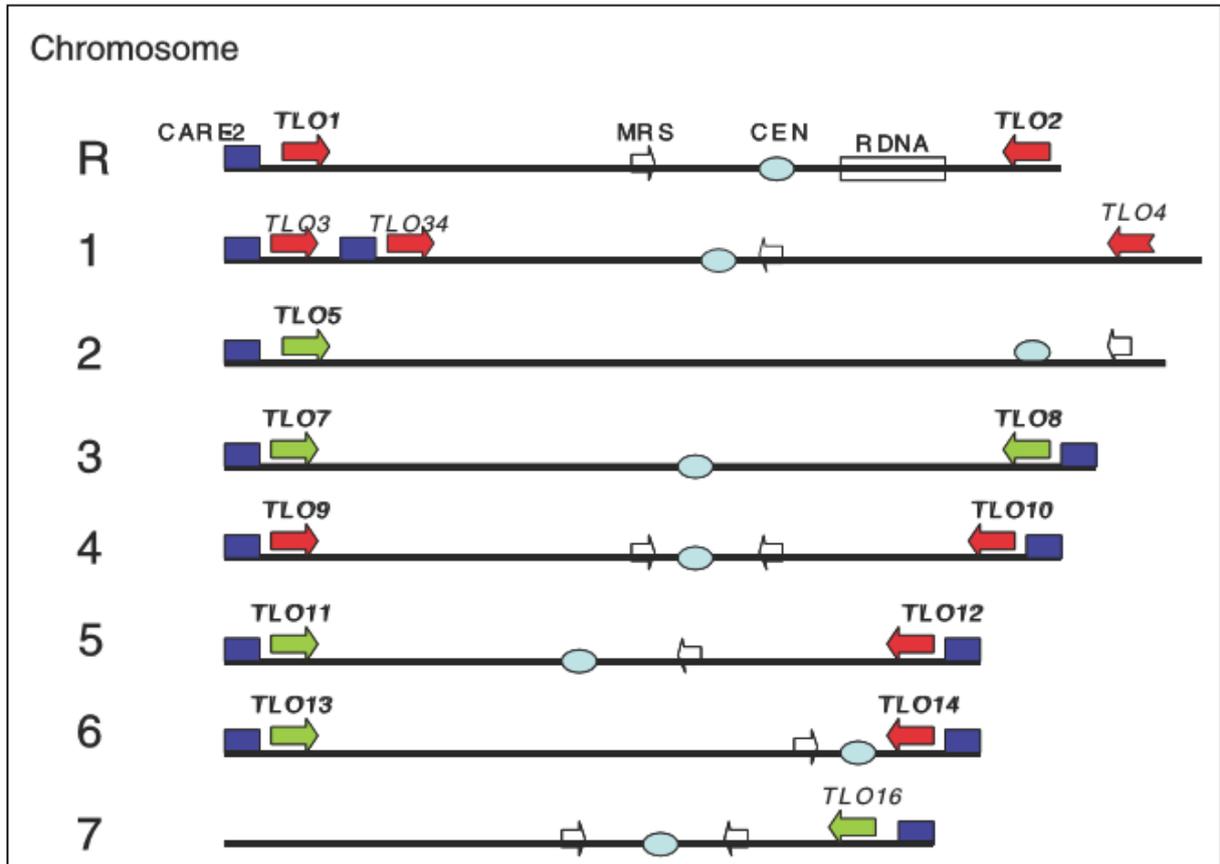


Figure 1 : Représentation schématique des 8 chromosomes de *C. albicans*, incluant une représentation des structures majeures comme les centromères (CEN), les séquences répétées MRS (Major Repeat Sequence) et CARE-2, l'ADN ribosomique et la famille de gènes associés aux télomères « TLO » (TeLOmere-associated genes) (van het Hoog et al., 2007).

L'annotation du génome de *C. albicans* réalisée en 2004 par Jones *et al.* avait permis d'identifier 6 419 ORFs. Une nomenclature officielle de ces gènes a été adoptée (similaire à celle de *S. cerevisiae*) et introduite dans la base de données « Candida Genome Database » (CGD). La dernière mise à jour de la base de données CGD décrit 6198 ORFs dont 72 % non caractérisés (CGD, 24 février 2018, (Arnaud et al., 2005; Binkley et al., 2014; Braun et al., 2005)).

A titre de comparaison, le génome de *S. cerevisiae* contient 6604 ORFs dont seulement 10 % non caractérisés (Saccharomyces Genome Database; <https://www.yeastgenome.org>; (Cherry et al., 2012)).

1.2 Variations du génome de *C. albicans*

L'analyse comparative des génomes de différentes souches de *C. albicans* a permis de révéler l'existence de nombreuses variations attestant d'une grande variabilité génomique au sein de l'espèce. Dans la suite de ce paragraphe, je ne décrirai que les modifications du génome de *C. albicans* associées à des variations de composition et des variations de structure. Ces variations jouent un rôle majeur chez les organismes asexués car elles permettent de générer une diversité génétique au sein de l'espèce (Hickman et al., 2015, 2013).

1.2.1 Variations de structure du génome

1.2.1.1 Modification de la ploïdie

C. albicans est un organisme diploïde ; son génome est composé de 8 paires de chromosomes. Un changement du nombre des chromosomes (n) est appelé une « aneuploïdie ». Cependant, le terme d'aneuploïdie est parfois utilisé de manière plus large pour désigner les altérations du génome résultant en une modification du nombre de copies (CNVs, Copy Number Variations) de régions chromosomiques incluant les événements de translocation, de duplication ou de délétion de fragments de chromosomes (**Figure 2**).

Les aneuploïdies peuvent survenir à la suite d'anomalie de ségrégation (inadéquate ou inégale) des chromosomes, au cours de la division cellulaire, à la suite de défauts dans la réparation des cassures double brins de l'ADN (Forche, 2014) et peuvent être générées au cours du cycle parasexuel de *C. albicans* (Forche et al., 2008) (**Figure 2**).

En effet, bien que cela soit rare, il a été observé en condition de stress, que *C. albicans* pouvait avoir recours à un cycle de reproduction dit parasexuel (**Figure 2**) (Dumitru et al., 2007; Hull et al., 2000; Magee and Magee, 2000; Zhang et al., 2015).

Durant ce cycle parasexuel, les cellules diploïdes ($2n$) homozygotes au locus *MTL* (Mating Type-Like) (i.e. *MTLa/a* ou *MTL α / α*) subissent une conversion phénotypique de la forme dite « white » à celle dite « opaque ». Les cellules opaques peuvent fusionner et former des cellules tétraploïdes (*a/a/ α / α*), souvent instables. L'état diploïde est restauré par la perte aléatoire et concertée des chromosomes surnuméraires (Forche et al., 2008; Noble and Johnson, 2007). Cependant, à l'issue de cette étape certaines cellules peuvent rester aneuploïdes ou présenter des événements de recombinaison génétique (**Figure 2**) (Hickman et al., 2015; Hirakawa et al., 2017; Selmecki et al., 2010).

L'analyse d'isolats de divers champignons pathogènes comme *C. glabrata*, *Cryptococcus neoformans* et *Cryptococcus gattii* a montré que ces événements d'aneuploïdie sont répandus chez ces organismes qui possèdent la capacité de supporter des changements drastiques au sein de leur génome. À titre d'exemple, la caractérisation génomique de 21 souches cliniques de *C. albicans* par Hirakawa et collaborateurs a montré que 6 d'entre elles (28%) présentaient une trisomie pour un ou plusieurs de ses chromosomes, 2 présentaient des duplications segmentales au niveau des Chr1 et ChrR, et 1 souche était haploïde sur une portion du bras droit du Chr3. La présence d'aneuploïdies dans cette population confirmait ainsi la tolérance de *C. albicans* à ces variations de structure (Hirakawa et al., 2015).

Cependant, dans cette étude, la fréquence élevée de souches aneuploïdes peut être non représentative de la population naturelle de *C. albicans*. En effet, les souches sélectionnées avaient une fréquence anormalement élevée (57 %, (12/21)) d'homozygotie au locus *MTL* en comparaison à la fréquence attendue, de l'ordre de 3 à 8% (Noble and Johnson, 2007). Cette proportion élevée pourrait s'expliquer par une exposition préalable de ces souches aux antifongiques azolés (d'Enfert et al., 2017), qui peut être associée à la survenue d'événements de perte d'hétérozygotie (LOH) et d'aneuploïdies au niveau du Chr5, sur lequel se trouve le locus *MTL* (Coste et al., 2006; Selmecki et al., 2006). De façon intéressante, dans les cellules aneuploïdes, le changement du niveau d'expression d'un grand nombre de gènes entraîne un déséquilibre au niveau de la synthèse protéique pouvant conférer un avantage ou un désavantage sélectif (Hirakawa et al., 2017).

Ces événements d'aneuploïdies, qui peuvent survenir en réponse à divers stress cellulaires causés par des modifications environnementales chez l'hôte, permettraient une adaptation rapide à des conditions défavorables. Nous discuterons cette hypothèse dans un chapitre spécifique.

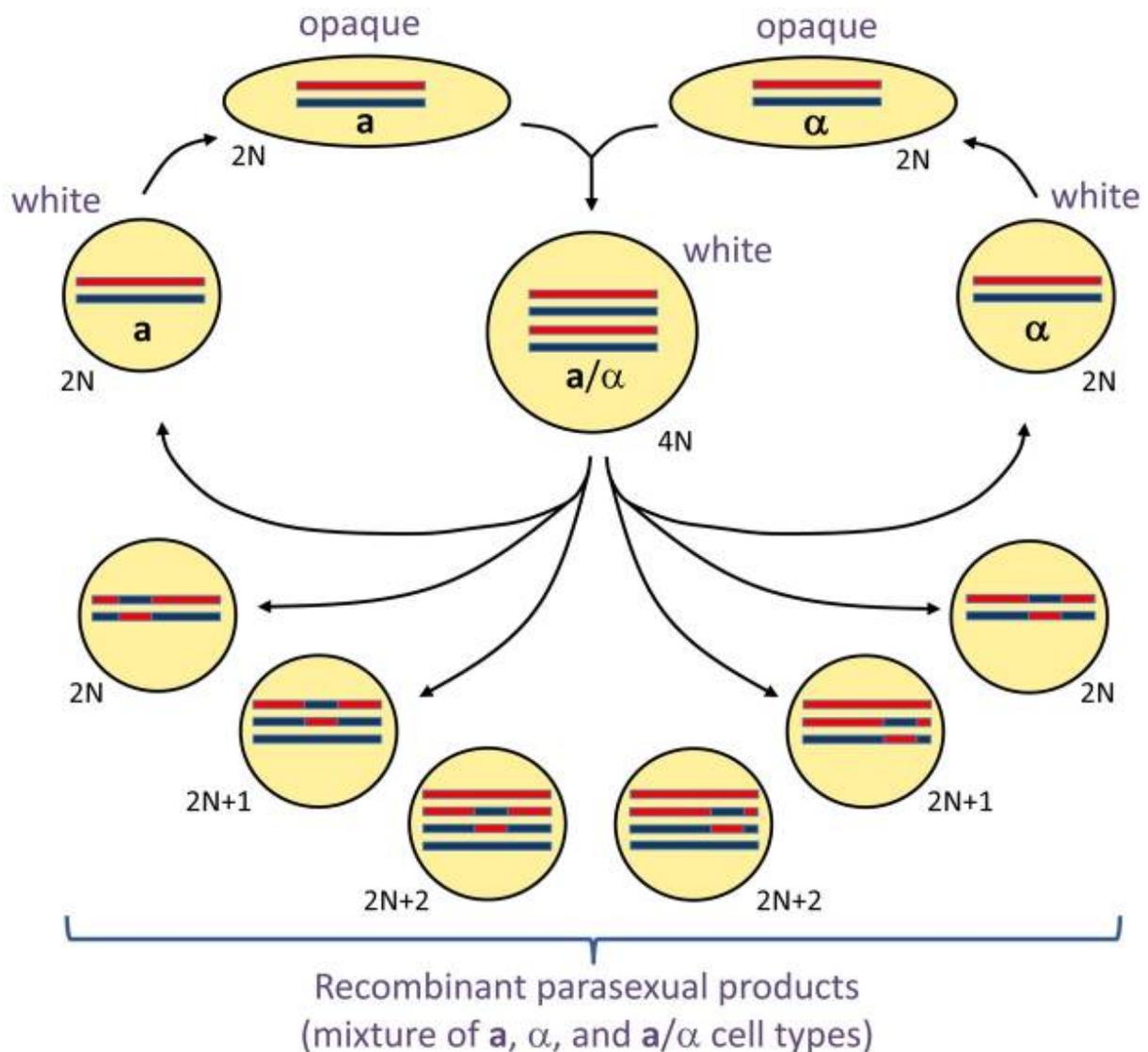


Figure 2 : Réarrangements génétiques survenant au cours du cycle parasexuel chez *C. albicans*. Dans ce schéma n'est représenté qu'un seul des 8 chromosomes. A la suite de la fusion de 2 cellules compétentes (opaque), les cellules tétraploïdes (4N) qui en résultent peuvent conserver leur tétraploïdie ou subir une perte concertée des chromosomes surnuméraires qui restaure la diploïdie des cellules. Enfin, un certain nombre d'entre elles peuvent présenter des événements de recombinaison génétique ou d'aneuploïdies (d'après RJ. Bennett *Curr Opin Microbiol.* 2015) (Bennett, 2015).

1.2.1.2 Modification de la taille des chromosomes, translocation et inversion

D'autres événements plus rares modifiant la structure du génome ont pu être identifiés chez *C. albicans* comme une modification de la taille des chromosomes, des translocations ou des inversions souvent associées aux régions répétées MRS (Selmecki et al., 2010). A ce jour, 9 MRS ont été identifiées dans le génome de *C. albicans* : 1 sur les chromosomes R, 1, 2, 5, et 6 et deux sur les chromosomes 4 et 7 (**Figure 1**). Les MRS sont composées d'une région RPS (Repetitive Sequence) de taille variable, flanquée des régions HOK et RB2 représentées sur la **Figure 3** (Chindamporn et al., 1998, 1995). La région RPS est composée d'un nombre variable de répétitions en tandem de 172 pb chacune, appelées "alts" (*C. albicans* tandem repeating unit), dont les séquences ne varient que légèrement à l'intérieur d'un chromosome et d'un chromosome à l'autre (Selmecki et al., 2010). La région RPS contient également 6 à 8 copies d'une séquence conservée de 29 pb (appelée COM29) contenant le site de restriction de 8 pb reconnu par l'endonucléase *SfiI* (**Figure 3**) (Lephart et al., 2005).

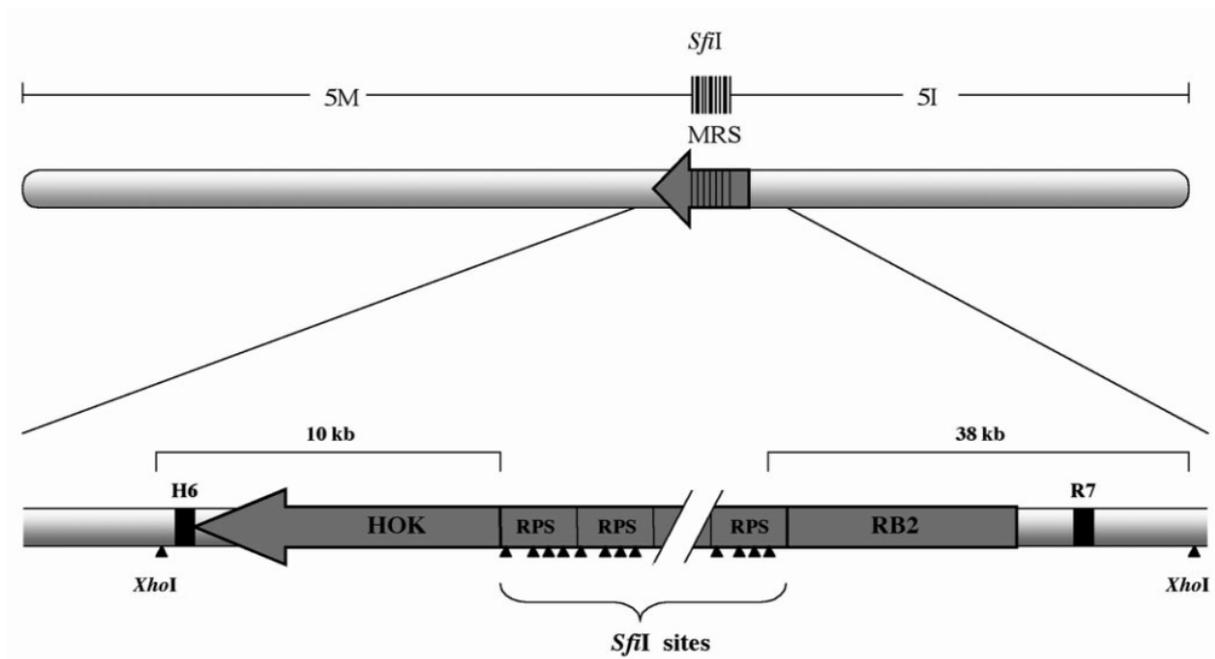


Figure 3 : Représentation schématique de la structure d'une MRS. Exemple du chromosome 5. La région RPS de taille variable est entourée de la région HOK et RB2 (Lephart et al., 2005).

Le nombre de répétitions incluses au sein des régions MRS peut, par conséquent, influencer sur la taille des chromosomes (Lephart et al., 2005). De plus, les MRS, du fait de leur structure répétée, sont également des régions préférentielles pour les recombinaisons et peuvent être associées à des événements de translocation réciproque lorsque la recombinaison se produit entre les MRS de deux chromosomes non-homologues (Lephart et al., 2005). Chu *et al.* ont décrit trois événements de translocation à proximité ou au sein de régions MRS des chromosomes 7 et 4, 5 et 6, et 5 et 1 dans la souche WO-1 (Chu et al., 1993). D'autres événements de translocation ont été observés dans des isolats cliniques de *C. albicans* et ont révélé un lien entre la survenue de ces événements et des modifications phénotypiques telles qu'une modification de l'assimilation des sucres ou l'absence de formation de chlamydospores (Chibana et al., 2000; Iwaguchi et al., 2001). Il est probable que ces événements de recombinaison au niveau des régions MRS sont induits par des conditions de stress (Lephart et al., 2005). Cependant, bien que décrits dans la littérature, ces événements sont extrêmement rares. A titre d'exemple, aucun événement de translocation n'a été identifié au sein des 21 isolats de *C. albicans* séquencés par Hirakawa *et al.* (Hirakawa et al., 2015). En revanche, tous les isolats présentaient des inversions (1 à 12 inversions de régions dont la taille variait de 705 pb à 144 kb) (Hirakawa et al., 2015).

Les événements modifiant la structure du génome de *C. albicans* sont relativement rares et ne constituent pas la source majeure de diversification génétique chez *C. albicans*. Celle-ci est essentiellement assurée par des variations de la composition du génome.

1.2.2 Variations de composition du génome

1.2.2.1 Hétérozygotie et mutations ponctuelles (SNPs)

Chez les organismes diploïdes les différences alléliques entre les deux chromosomes homologues sont désignées sous le terme « d'hétérozygotie ». La caractérisation du génome de *C. albicans* a permis de montrer que l'hétérozygotie apportait une source importante de diversité au sein d'un isolat mais également au sein de l'espèce (Muzzey et al., 2013; Ropars et al., 2018).

En effet, le séquençage de la souche de référence de *C. albicans* SC5314 a montré que son génome était fortement hétérozygote avec 65 787 SNPs entre les deux allèles soit 1 position hétérozygote toutes les 217 pb (Jones et al., 2004; Muzzey et al., 2013).

A titre de comparaison, cette valeur est plus de 6 fois supérieure à ce que l'on peut observer chez l'Homme chez qui environ 9×10^5 SNPs hétérozygotes ont été décrits, soit une moyenne de 1 SNP tous les 1 331 pb (Sachidanandam et al., 2001). Chez *S. cerevisiae* le nombre de positions hétérozygotes peut varier de façon importante en fonction des souches étudiées, allant de 1 SNP tous les 324 pb pour les génomes les plus hétérozygotes à seulement 1 SNP toutes les 36 000 pb pour ceux qui le sont moins (Magwene et al., 2011).

Les SNPs sont localisés aussi bien dans les régions intergéniques que dans les ORFs. Dans ce dernier cas, lorsque le polymorphisme entraîne une mutation non synonyme cela conduit à la traduction d'une protéine différente entre les deux allèles. En revanche, la densité des SNPs est hétérogène, certaines régions étant plus riches que d'autres. Cela peut s'expliquer par la survenue de grands événements de perte d'hétérozygotie (LOH) (Jones et al., 2004; Muzzey et al., 2013) que nous décrivons plus loin. Une représentation de la densité des SNPs hétérozygotes sur les 8 chromosomes de la souche de référence SC5314 permet d'illustrer la différence d'hétérozygotie entre chromosomes (**Figure 4**). Dans cette souche, les Chr 4, 5 et 6 présentent une densité de SNPs hétérozygote importante tandis que les Chr3 et 7 ainsi que les deux extrémités du ChrR sont largement homozygotes.

En plus des positions polymorphes, un grand nombre (6 103) d'insertions et délétions (Indels) de petite taille ont été détectées entre les deux haplotypes de la souche SC5314. Ces événements sont plus fréquents au niveau des régions répétées de certains promoteurs et des origines de réplication du génome (Muzzey et al., 2013).

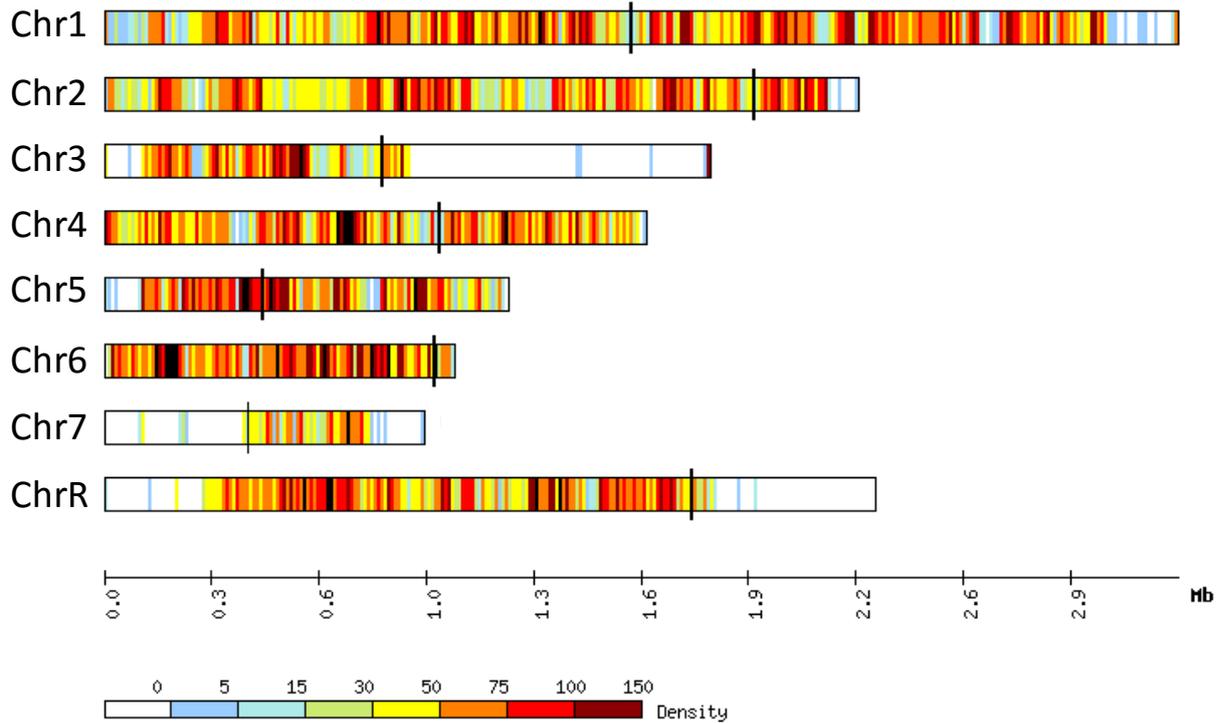


Figure 4 : Heatmap représentant la densité de SNPs hétérozygotes sur les 8 chromosomes de la souche de référence SC5314 (Fenêtre de 10 Kb). Les zones rouges représentent les régions les plus hétérozygotes et les zones blanches les régions homozygotes. (Données non publiées provenant du laboratoire)

1.2.2.2 Perte d'hétérozygotie (LOH)

Un autre moyen de diversification génétique observé chez les organismes diploïdes, comme *C. albicans*, est la perte d'hétérozygotie (LOH). Plusieurs mécanismes peuvent être à l'origine d'un LOH tel que la formation d'une cassure double brin ou un défaut de ségrégation des chromosomes lors de la mitose (**Figure 5**) (Carr and Gottschling, 2008). Dans le cas d'une cassure double brin, le mécanisme selon lequel celle-ci sera réparée peut donner lieu à différents types de LOH. Ces mécanismes peuvent être séparés en deux catégories selon qu'ils dépendent ou non d'événements de recombinaison.

- Mécanismes dépendants de la recombinaison

La réparation d'une cassure double-brin, peut se faire via un mécanisme de type NHEJ (Non-Homologous End-Joining) qui aboutit à une délétion locale après résection et ligation de la cassure double-brin.

La réparation de la cassure par conversion génique conduit au remplacement de la zone ayant subi une cassure double-brin par une copie de l'allèle homologue conduisant à la formation d'un LOH localisé de petite taille (**Figure 5, A**). La mise en place d'un mécanisme de réparation par BIR (Break-Induced Replication) conduit à la réplication de la zone allant du site de cassure jusqu'au télomère, conduisant à un LOH de grande taille (**Figure 5, A**). Ceci est également le cas lors d'une réparation par crossing-over, qui conduit en plus à l'apparition d'un LOH dans les deux cellules filles, ce qui n'était pas le cas pour l'ensemble des mécanismes précédents (Bennett et al., 2014; Carr and Gottschling, 2008).

- Mécanismes indépendants de la recombinaison

Une cassure double brin de l'ADN peut conduire à la perte du chromosome entier et ainsi former une aneuploïdie. Le chromosome restant peut être redupliqué, ce qui aboutit à la formation de deux chromosomes homozygotes (**Figure 5, B**).

Comme pour les aneuploïdies, les LOH peuvent survenir rapidement, permettant de survivre à un stress (Forche, 2014). Cependant, bien que l'aneuploïdie soit un processus réversible, le LOH ne l'est pas. Celui-ci entraîne une homozygotie de tous les gènes de la région concernée.

Une analyse détaillée de ces événements chez diverses souches a montré que tous les isolats de *C. albicans* présentaient au moins un LOH de grande taille sur un des chromosomes (Butler et al., 2009; Hirakawa et al., 2015; Jones et al., 2004). Par exemple, la souche de référence SC5314 présente plusieurs LOH de grande taille impactant les deux extrémités distales du ChrR, la majorité du bras droit du Chr3, ainsi qu'une grande partie distale des deux bras du Chr7 (**Figure 4**).

La localisation de ces LOH diffère d'une souche à l'autre suggérant qu'il s'agit d'événements non spécifiques, à l'exception des LOH du bras droit du ChrR qui débutent pour l'ensemble des isolats au niveau de la région contenant l'ADNr (Butler et al., 2009; Hirakawa et al., 2015; Jones et al., 2004; Ropars et al., 2018).

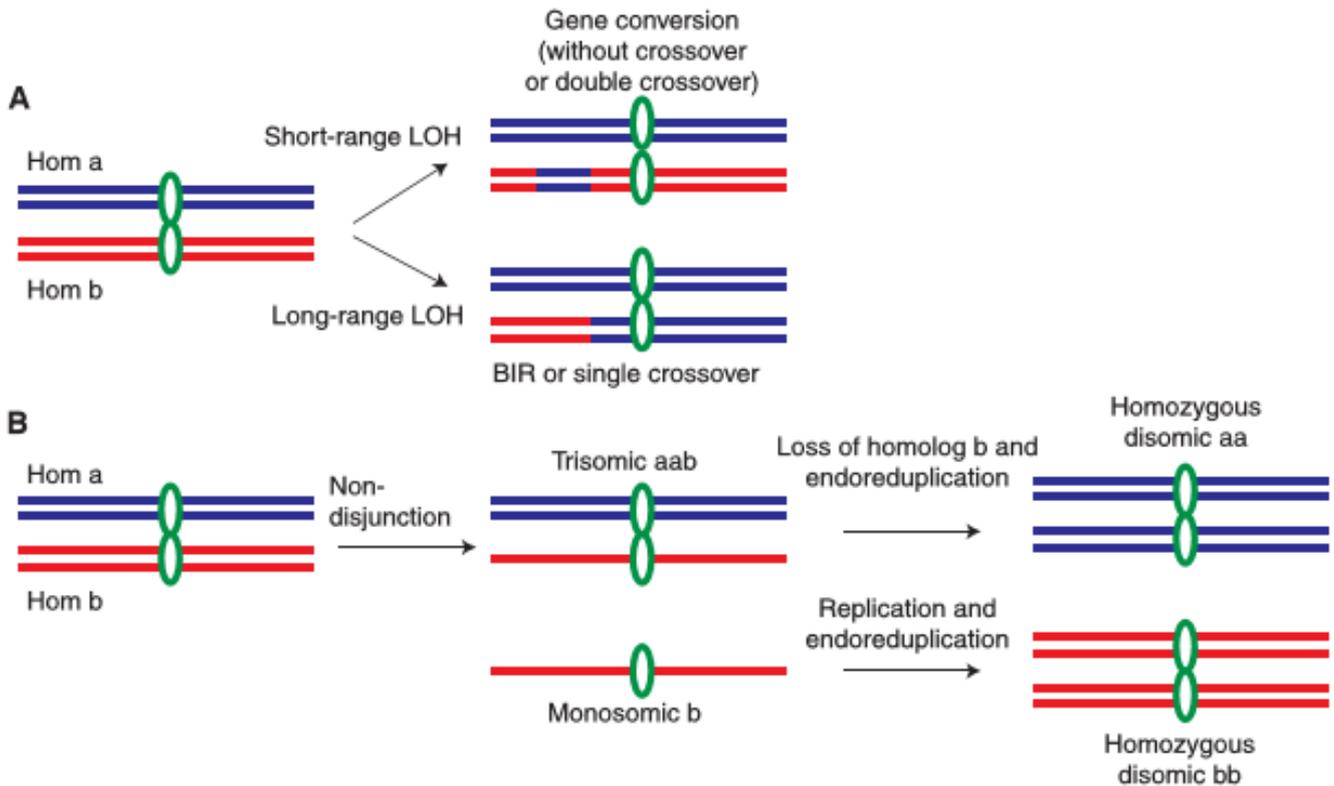


Figure 5 : Mécanismes de survenue des LOH. (A) Une recombinaison mitotique peut générer des LOH de petite taille par conversion génique ou double crossing-over de l'ADN, ou des LOH de grande taille par simple crossing-over ou réplication à la suite d'une cassure double brin de l'ADN (BIR, Break Induced Replication). (B) Une non-disjonction chromosomique peut générer des LOH d'un chromosome entier (Bennett et al., 2014).

1.3 Méthodes d'étude de la variabilité génomique chez *C. albicans*

Il existe de nombreuses méthodes permettant d'analyser les variations de structure du génome telles que les modifications de la taille ou du nombre de chromosomes mais également les variations de composition incluant les LOH et les SNPs.

Quatre techniques ont particulièrement été utilisées chez *C. albicans* et sont maintenant remplacées par les techniques de séquençage de nouvelles générations.

Il s'agit de la cytométrie en flux, de l'électrophorèse en champs pulsé -et notamment le CHEF (Contour-clamped Homogeneous Electric Field), des puces d'hybridation génomique comparative (CGH) (Array Comparative Genomic Hybridation) et des puces à SNP (SNP Array) (Selmecki et al., 2010). Nous ne détaillerons pas l'aspect technique de ces quatre méthodes mais résumerons les principaux avantages et limites de celles-ci. Par ailleurs, nous décrirons la principale technique actuellement utilisée pour génotyper les souches de *C. albicans* et finirons par les approches de séquençage de nouvelles générations qui permettent maintenant d'étudier de façon plus fine les variations du génome.

1.3.1 Techniques appliquées à l'analyse des variations génomiques

Les quatre techniques principalement utilisées chez *C. albicans* pour l'analyse des variations génomiques (cytométrie en flux, CHEF, CGH et SNP array) fournissent des informations importantes sur la structure du génome mais présentent également plusieurs limites.

La méthode qui a longtemps été la plus utilisée pour l'étude de l'organisation du génome consiste à évaluer la taille des chromosomes sur un caryotype obtenu en séparant les chromosomes par des techniques d'électrophorèse sur gel en champs pulsé (Scherer and Magee, 1990). La méthode CHEF (Contour-clamped Homogeneous Electric Field) permet d'identifier des modifications importantes dans la structure des chromosomes telles que des variations majeures de la taille des chromosomes ou l'observation de translocations réciproques grâce à une étape d'hybridation avec des sondes (« Southern blot ») (Selmecki et al., 2010). Cependant, certaines variations sont trop petites pour être détectées par ces méthodes, et nécessitent de réaliser des cartographies plus fines. Une digestion enzymatique préalable de l'ADN chromosomique peut être réalisée avec différentes enzymes de restriction permettant soit d'améliorer l'identification des fragments chromosomiques réarrangés soit de déterminer des variations plus précises de la taille des chromosomes. Néanmoins, cette méthode est limitée par le nombre de sondes utilisées et il est difficile de détecter des événements tels que des trisomies de chromosomes entiers du fait de la superposition de la migration des chromosomes.

Les puces d'hybridation génomique comparative aussi appelée « Array comparative genomic hybridization » (CGH) sont utilisées pour l'analyse des variations du nombre de copies (CNVs) de l'ADN. Elles permettent de détecter des aneuploïdies segmentales mais également des CNVs pour chacun des gènes du génome (Berman and Forche, 2017). Cependant, cette technique comme le CHEF, ne permet pas de détecter les modifications de la ploïdie du génome entier ni les translocations réciproques et les LOH (Selmecki et al., 2010).

Les puces à SNP, ou « SNP Array » ont été développées grâce à l'obtention de la séquence génomique de la souche SC5314. Ces puces permettent l'analyse à haut débit des génotypes de souches de *C. albicans* (Forche et al., 2005). Dans cette technique, de petits oligonucléotides (2 versions alléliques pour chaque SNP testé) sont fixés sur des lames. Les sondes d'une souche hétérozygote s'hybrident sur les deux allèles, tandis que les sondes d'une souche homozygote ne s'hybrident qu'à l'un des 2 allèles. Le génotype de chaque SNP est déterminé en calculant la fraction allélique pour chacun des trois génotypes possibles (Forche et al., 2009). De cette façon, la quantification des ratios alléliques sur l'ensemble des SNPs permet de détecter d'une part des CNVs tels que des trisomies de chromosomes entiers ou segmentaires mais également des événements de LOH (Selmecki et al., 2010). Cependant cette technique ne permet pas de détecter des translocations et les variations de ploïdie du génome entier. De plus, l'analyse des puces à SNP est limitée par le nombre de SNPs inclus dans la puce mais également par la difficulté d'analyser des souches ne dérivant pas de la souche de référence SC5314 (Forche et al., 2005).

Finalement, la cytométrie en flux permet de détecter des variations dans la ploïdie du génome entier par l'analyse du contenu génique des cellules. En revanche, elle ne permet pas de détecter une duplication ou la perte d'un seul chromosome. Un décalage de la ploïdie n'est observé que lorsque les souches contiennent plusieurs aneuploïdies (Selmecki et al., 2010). Ainsi, seule la combinaison des quatre méthodes permettrait de fournir une vue complète des variations du génome.

1.3.2 Techniques de génotypage moléculaire des souches de *C. albicans*

Les méthodes classiques de génotypage sont basées sur l'analyse du polymorphisme génétique d'individus d'une même espèce tels que le polymorphisme de longueur des microsatellites (MLMT, Multilocus microsatellite typing) (Dalle et al., 2000), le polymorphisme de la taille de fragments d'ADN (Ca3 fingerprint) (Pujol et al., 1997) ou le polymorphisme nucléotidique (MLST, Multilocus Sequence Typing) (Bougnoux et al., 2003). L'efficacité de la méthode de typage est définie par son pouvoir discriminant qui caractérise son aptitude à distinguer deux souches n'ayant aucune relation entre elles (Hunter, 1995). D'autres paramètres sont considérés pour le choix d'une méthode par rapport à une autre comme la reproductibilité et la standardisation de la méthode permettant de comparer les résultats de typage des souches entre les différents laboratoires. La méthode MLST est probablement la méthode la plus standardisée et la plus utilisée pour typer les souches de *C. albicans* (Bougnoux et al., 2004, 2002).

Cette méthode est basée sur l'analyse du polymorphisme nucléotidique (SNPs) de courtes séquences (450 à 600 bp) de plusieurs régions ou loci situées dans des gènes de ménage (housekeeping genes), codant pour des protéines essentielles (Odds et al., 2007). Ces régions sont le siège de polymorphismes stables dans le temps mais néanmoins suffisamment fréquents pour permettre de distinguer des génotypes de souches très proches génétiquement, faisant de cette méthode une des méthodes de typage la plus discriminante. Le typage MLST permet de caractériser les souches de *C. albicans* par détermination de leur DST pour « Diploid Sequence Type » (Bougnoux et al., 2004). En effet, *C. albicans* étant diploïde, l'existence de séquences différentes entre les deux chromosomes homologues induit l'apparition de positions polymorphes hétérozygotes. Le DST de la souche est obtenu par concaténation des séquences nucléotidiques de 7 loci distribués sur 5 des 8 chromosomes de *C. albicans* (Chr1, -2, -3, -4 et -6) (Bougnoux et al., 2003). Cependant, dans cette approche, seul un nombre limité de positions polymorphes sont analysés. Une base de données internationale accessible librement via internet (<https://pubmlst.org/calbicans/>) recense les données épidémiologiques des souches ainsi que leur DST. Elle représente un support très utile pour permettre de suivre la dispersion des souches à l'échelle internationale (Bougnoux et al., 2004).

De façon générale, les méthodes de typage actuelles permettent d'affirmer que des isolats de *C. albicans* sont génotypiquement différents. En revanche, lorsque deux isolats présentent le même profil, il n'est pas possible d'affirmer que ces isolats sont identiques (Odds and Jacobsen, 2008). Le séquençage du génome complet apporte un complément d'information permettant de comparer des isolats cliniques présentant le même profil MLST par exemple. Cependant, même le séquençage du génome peut poser des difficultés pour affirmer que deux souches sont identiques (Odds, 2010).

1.3.3 Séquençage du génome complet de *C. albicans*

L'évolution des techniques de séquençage a permis d'augmenter massivement le nombre de séquences de génome complet de *C. albicans* disponibles. Différentes stratégies ont été employées successivement pour décrire les génomes entiers de divers organismes. Celles-ci ont permis d'analyser ou de comparer les génomes de plusieurs souches au sein d'une même espèce ou d'étudier l'évolution génomique d'une même souche au cours du temps par l'analyse d'isolats séquentiels.

1.3.3.1 Approches utilisant le clonage/séquençage

Les approches de clonage/séquençage, initialement utilisées, consistaient en la fragmentation aléatoire du génome par diverses méthodes (ultrason, nébulisation ou enzymatique) générant des fragments d'environ 1kb. Ces fragments étaient clonés dans un vecteur plasmidique, séquencés par méthode de Sanger, puis assemblés de façon à obtenir des contigs les plus grands possibles. Une autre version de cette méthode consistait à cloner de grands fragments dans des plasmides particuliers appelés BAC pour « Bacterial Artificial Chromosome » qui ont la capacité d'accepter des fragments génomiques allant jusqu'à 250 kb (O'Connor et al., 1989). Les clones étaient ensuite cartographiés, soit par analyse des profils de restriction enzymatique, soit plus fréquemment par séquençage et hybridation des extrémités des BAC. Une fois cartographiés, les clones étaient fragmentés et séquencés individuellement puis assemblés par alignement bioinformatique (Jones et al., 2004). Ces stratégies étaient longues et fastidieuses et ont été largement remplacées par des approches de séquençage à haut débit.

1.3.3.2 Approches de séquençage à haut débit

Les techniques de séquençage à haut-débit ou « Next-Generation- Sequencing » (NGS), ont été développées pour augmenter de façon très importante la quantité de séquences obtenues en une seule expérimentation. Des quelques séquences de 800 à 1000 pb que l'on pouvait séquencer en quelques jours dans les années 80, on est arrivé à des techniques simplifiées qui peuvent séquencer des milliards de nucléotides par expérience. Par conséquent, de plus en plus de génomes sont séquencés ou en cours de séquençage. En Septembre 2016 le nombre de génomes séquencés était de plus de 9 000, et plus de 96 000 génomes sont en cours d'obtention ("Genomes OnLine Database" - GOLD, NCBI - Genome sequencing projects) (Goodwin et al., 2016).

Depuis les années 2000, plusieurs technologies ont été développées successivement. Quelle que soit la technique employée, le processus se déroule toujours en trois étapes. La première consiste à préparer des « librairies » d'ADN afin de « formater » l'ADN sous une forme lisible par les séquenceurs; la deuxième étape consiste à séparer chaque fragment d'ADN puis à l'enrichir; la troisième étape est celle du séquençage proprement dit, qui utilise des techniques de polymérisation ou de ligation et une détection du signal fondée sur la détection de pyrophosphate (pyroséquençage) pour la technologie Roche, d'ion H⁺ (conductimétrie) pour la technologie Ion Torrent ou de fluorescence pour la technologie Illumina. Le signal est ensuite converti en séquences qui doivent être analysées par des logiciels spécifiques afin d'être transformées en données biologiquement interprétables (Metzker, 2010).

Les séquenceurs dits à très haut débit utilisent essentiellement la technologie Illumina (Bennett, 2004). Celle-ci permet de générer des millions voire, des milliards de séquences de courte taille permettant d'obtenir une profondeur de séquençage (nombre de séquence moyenne couvrant le génome) très élevée, assurant une très bonne sensibilité pour la détection de variants.

Cette technologie se compose d'une étape de fragmentation aléatoire du génome par approche enzymatique ou mécanique, suivie d'une étape de réparation-ligation des adaptateurs aux extrémités des fragments. Les adaptateurs sont ensuite fixés, via des oligomères complémentaires aux adaptateurs, sur une lame en verre appelée « flow-cell ». Afin d'augmenter l'intensité du signal, chaque brin d'ADN est amplifié de manière clonale dans un périmètre géographique restreint par bridge-PCR (**Figure 6**) (Bennett, 2004). A l'issue de cette étape dite de « clustering », un séquençage par synthèse du brin complémentaire avec des dNTPs fluorescents permet la génération des séquences associées à chaque fragment d'ADN initial (**Figure 6**) (Bennett, 2004).

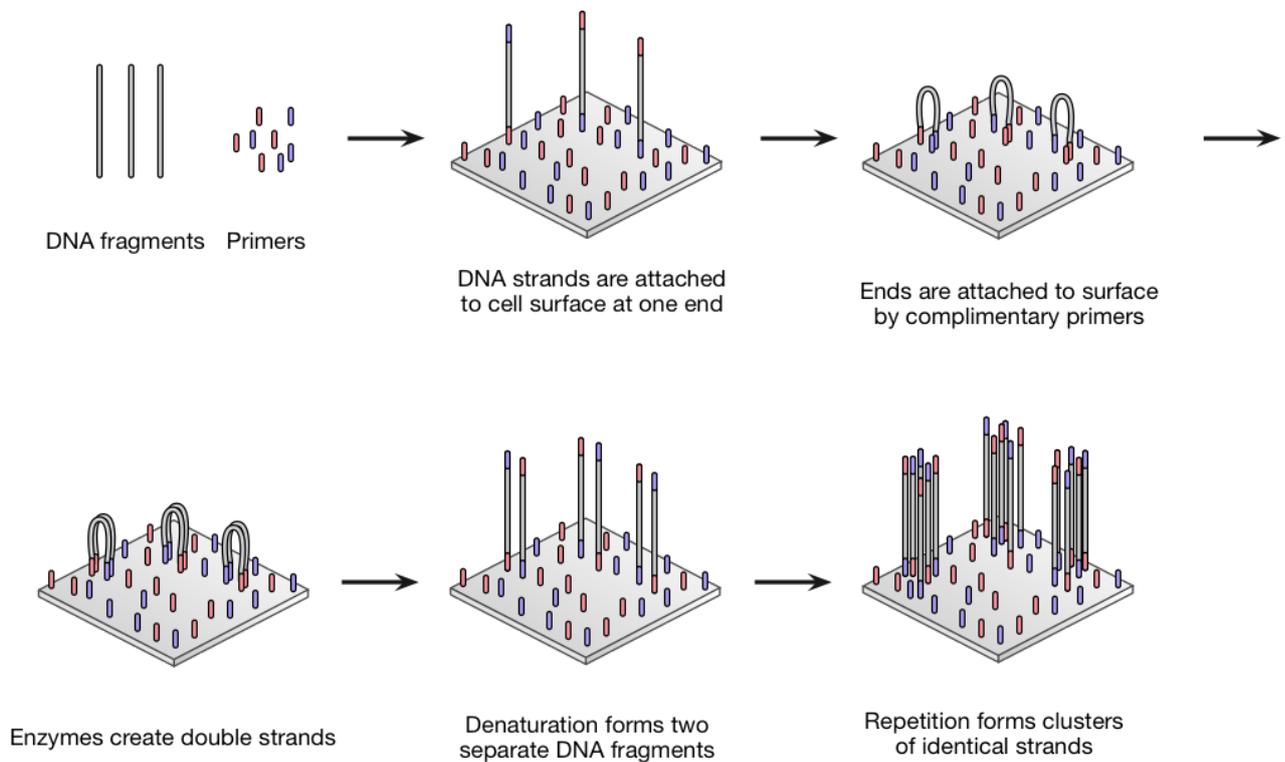


Figure 6 : Schéma des étapes de préparation de l'ADN en vue d'un séquençage par technique Illumina (Nucleic Acids Book Prof. Tom Brown).

Il est important de noter que cette technique a évolué au cours des dernières années. Tout d'abord grâce à l'amélioration des performances des polymérases qui a permis d'augmenter la longueur des séquences, initialement limitée à 36 pb, jusqu'à 300 pb (Goodwin et al., 2016).

Par ailleurs, pour pallier la difficulté liée à la génération de courtes séquences, il est également possible de réaliser un séquençage dit « Paired-end » correspondant au séquençage des 2 extrémités d'un fragment plus long permettant d'atteindre des séquences de 2X300 pb. Etant donné que la distance approximative entre chaque morceau séquencé est connue, les algorithmes d'alignement utilisent cette information pour positionner avec plus de précision les fragments dans les régions peu diversifiées telles que les séquences répétées. Les quantités d'ADN requises pour le séquençage ont également été fortement réduites. Cela a permis de s'affranchir des étapes de PCR initialement effectuées, évitant les nombreux biais comme les duplicats de PCR, gênant pour l'analyse des variations du nombre de copies ou diminuant la sensibilité de détection des mutations. Ces modifications expérimentales ont ainsi contribué non seulement à faciliter l'analyse des données, mais aussi à améliorer la qualité des résultats obtenus par les algorithmes d'analyse (Goodwin et al., 2016).

1.3.3.3 Analyse des données de séquençage à haut débit

Le séquençage de nouvelle génération a ainsi permis de générer de façon exponentielle des nouvelles données sur le génome de nombreux organismes. Le défi est maintenant le développement d'outils bioinformatiques capables d'analyser et représenter ces informations complexes d'une manière biologiquement pertinente et interprétable (Abbey et al., 2014).

La plupart des outils bioinformatiques actuellement disponibles pour l'étude des génomes eucaryotes ont été conçus pour l'analyse des données de séquences du génome humain, et notamment pour le dépistage de diverses maladies génétiques. Ils se présentent en général sous formes de modules qui, lorsqu'ils sont associés, permettent d'aller jusqu'à l'identification de variations même mineures dans le génome.

Ces associations, appelées également « pipeline » d'analyse, sont composées classiquement de trois grandes parties (**Figure 7**) (Abbey et al., 2014) :

- un module qui effectue l'alignement et le traitement des séquences brutes
- un module qui effectue les analyses de CNVs et SNP/LOH
- un module qui effectue des figures résumant toutes les analyses

Les différentes étapes qui constituent le pipeline d'analyse que nous utilisons pour comparer des génomes de *C. albicans* seront détaillées dans la partie résultat.

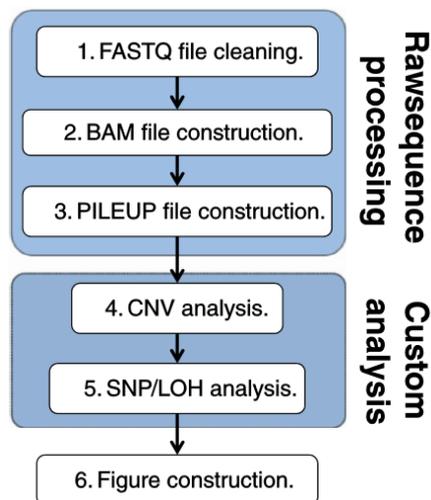


Figure 7 : Processus global (pipeline) d'analyse des génomes diploïdes (Abbey et al., 2014).

2 Evolution et adaptation génomique de *C. albicans*

Les organismes vivants ont la nécessité de s'adapter à leur environnement naturel mais doivent également être capables d'initier rapidement des stratégies d'adaptation pour survivre en cas de situation de stress (Forche et al., 2009). En biologie évolutive, l'adaptation se définit comme la modification d'un caractère anatomique, d'un processus physiologique ou d'un trait comportemental dans une population d'individus sous l'effet de la sélection naturelle. Le nouvel état de ce caractère améliore la survie et le succès reproductif des individus qui en sont porteurs (Grandcolas, 2009). Cette dernière définition est souvent critiquée, car elle ne précise pas ce qu'est l'adaptation, mais indique seulement ce qu'elle produit. L'adaptation est à la fois ce qui conditionne la sélection des individus dans leur milieu et le résultat de cette sélection. Par ailleurs, cette définition pose problème, car Charles Darwin a élaboré la théorie de la sélection naturelle précisément dans le but d'expliquer l'adaptation des êtres vivants à leurs conditions de vie. Dans « L'Origine des espèces (1859) », il n'avance aucune définition de l'adaptation, car il la considère comme une réalité évidente (Darwin, 1859). De fait, cette notion reste floue et mal définie, alors même qu'elle est centrale pour la biologie évolutive. L'adaptation d'un organisme à son environnement pourrait donc se définir par la présence de caractères (morphologiques, physiologiques, etc.) permettant d'augmenter les possibilités de survie de l'espèce en présence de conditions environnementales données ou modifiées à un instant t.

Les mécanismes mis en place au cours de l'évolution par les différents organismes sont extrêmement diversifiés voire même opposés. On peut par exemple distinguer la stratégie « d'extrême diversité » adoptée par les virus à ARN comme la grippe ou le VIH, qui présentent des cycles répliatifs très courts, un taux de réplication extrêmement élevé et un taux de viabilité très faible, et la stratégie « conservative » adoptée par certains eucaryotes comme l'Homme qui présente un temps de génération très long, un taux de reproduction très faible, mais un taux de viabilité élevé.

Par ailleurs, les mécanismes adaptatifs de tout être vivant peuvent également être considérés en fonction du temps nécessaire pour leur mise en place. On peut donc distinguer différentes échelles de temps de l'adaptation allant de la seconde à des millions d'années (Swynghedauw, 2012).

Il est ainsi impossible d'évoquer tous les mécanismes mis en jeu, ou tous les exemples d'adaptation tant ils sont nombreux, mais nous allons ici détailler les principaux mécanismes adaptatifs mis en place par *C. albicans*.

2.1 Evolution et diversité du génome de *C. albicans*

Dans l'évolution darwinienne la génération de diversité précède la sélection. La diversité peut être due à des mutations d'origines diverses survenant au hasard ; elle précède la sélection et la spéciation due soit à la pression sélective, soit au hasard. La constante de temps est ici le million d'années, correspondant au temps nécessaire pour qu'une mutation fonctionnelle soit sélectionnée (Swynghedauw, 2012). Dans cette partie, nous allons discuter de la façon dont le génome de *C. albicans* a évolué au cours du temps, pour générer de la diversité au sein de l'espèce, mais également au sein du règne fongique.

2.1.1 Evolution génétique de *C. albicans* au sein des *Saccharomycetales*

Sur le plan phylogénétique le genre *Candida* appartient au phylum des *Ascomycota*, et au sous-phylum des *Saccharomycotina* (**Figure 8**). Ce sous-phylum contient 18 classes dont la classe des *Saccharomycetes* contenant l'ordre des *Saccharomycetales* (Suh et al., 2006). Cet ordre se compose de 16 familles, dont deux ont divergé il y a environ 170 millions d'années (McManus and Coleman, 2014). La famille des *Saccharomycetales incertae sedis* contient le sous-groupe connu sous le nom de clade CTG. Chez ce dernier, le codon CTG est en effet traduit dans la majorité des cas en une sérine au lieu d'une leucine. Ce clade contient la majorité des espèces de *Candida* d'importance médicale (**Figure 9**). Une autre Famille dans l'ordre des *Saccharomycetales* se compose principalement du genre *Saccharomyces*, qui contient des espèces pour lesquelles les génomes ont subi une duplication complète appelée clade « WGD » (Whole Genome Duplication) dont l'espèce *S. cerevisiae* (McManus and Coleman, 2014).

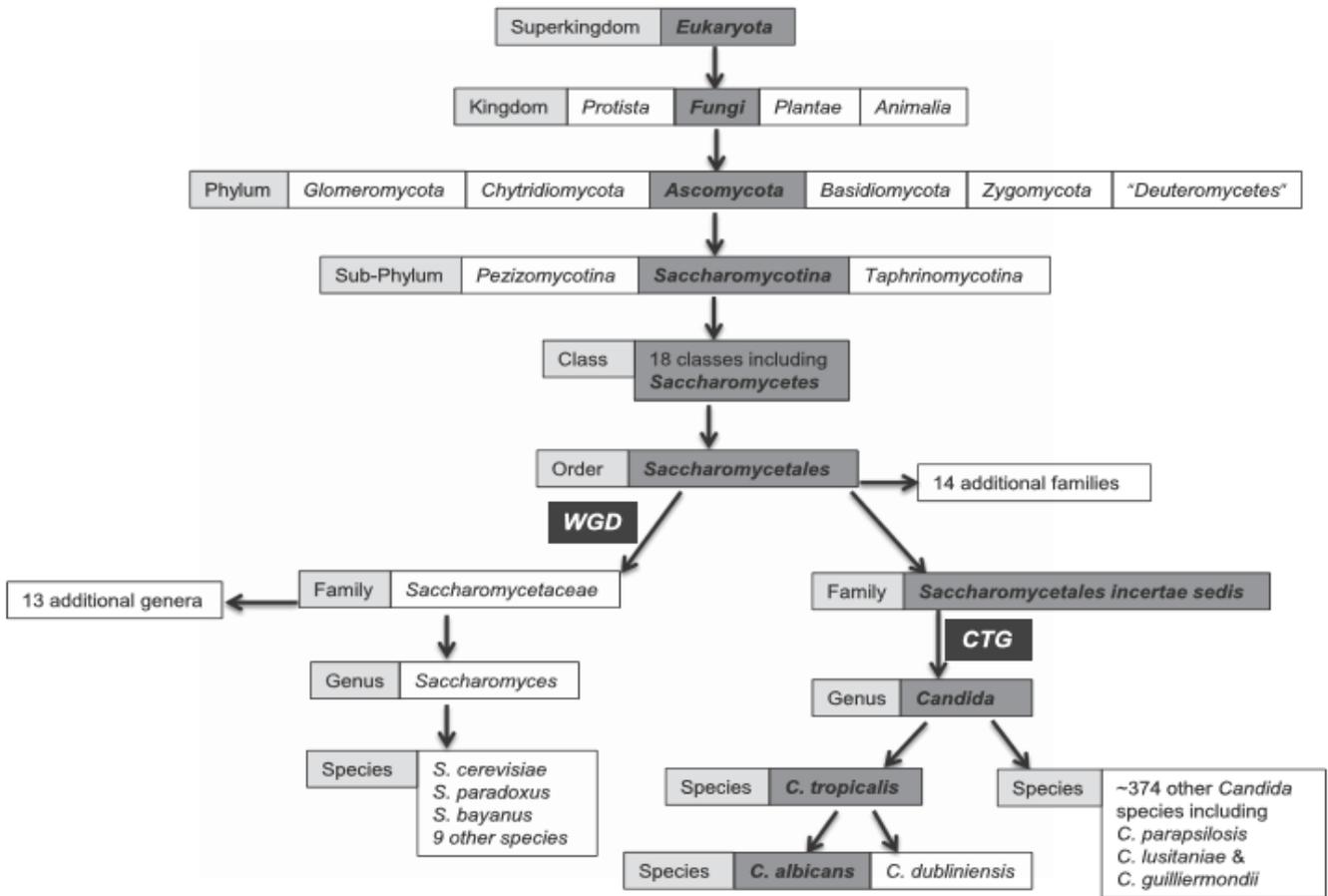


Figure 8 : Positionnement de l'espèce *C. albicans* au sein du règne fongique (McManus and Coleman, 2014).

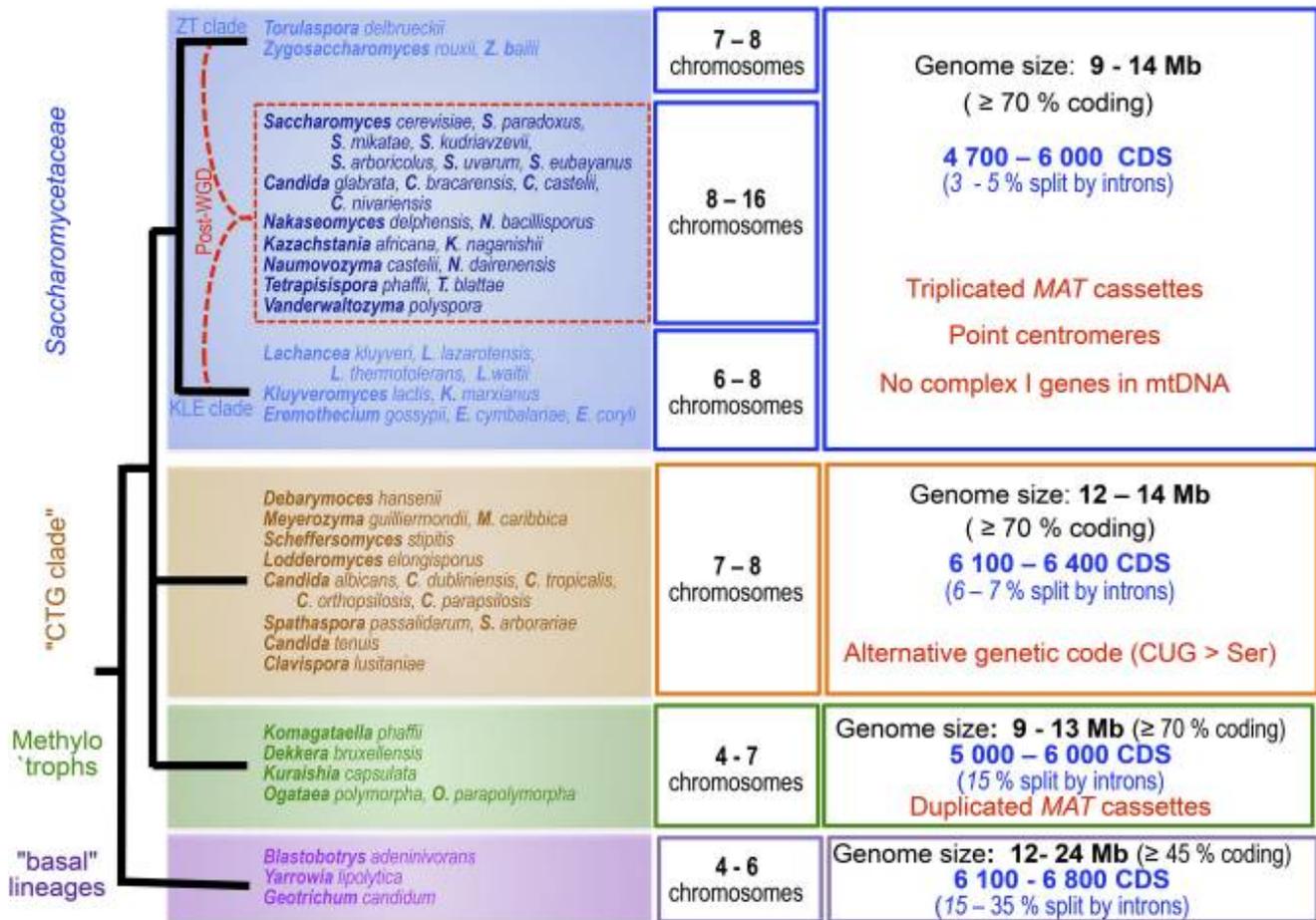


Figure 9 : Comparaison des caractéristiques des génomes de 4 sous-groupes majeurs du sous-embranchement des *Saccharomycotina* (B. A. Dujon et E. J. Louis ; Genetics 2017).

A l'échelle du genre, *C. albicans*, ainsi que plusieurs autres espèces de *Candida*, appartient au clade CTG qui se caractérise, comme nous l'avons vu précédemment, par la traduction des codons CUG, codant normalement pour une leucine, en résidu sérine. L'apparition du gène codant pour l'ARNt-CAG^{Ser} est survenue pour la première fois il y a environ 270 millions d'années (Gomes et al., 2007). Celui-ci était en compétition avec le type sauvage ARNT-CAG^{Leu} pour les 30 000 codons CUG de l'ancêtre. Le nouvel ARNT-CAG^{Ser} a été perdu au sein de la lignée *Saccharomyces*, qui a conservé le standard ARNT-CAG^{Leu} tandis que les *Candida* ont conservé l'ARNt-CAG^{Ser} (Gomes et al., 2007). L'utilisation du codon CUG chez *Candida* est rare ; il ne reste aujourd'hui que 1 à 2% des codons CUG originaux (Butler et al., 2009; Massey et al., 2003).

Les mécanismes évolutifs générant la diminution de ces codons est peu connue. Plusieurs études ont tenté de déterminer l'impact sur le protéome de l'altération du code génétique en réintroduisant le ARNt-CAG^{Leu} de *S. cerevisiae* dans une souche de *C. albicans* afin de recréer un niveau d'ambiguïté élevé dans la traduction en protéines (Gomes et al., 2007; Miranda et al., 2007). Les variations ainsi générées ont pour conséquence d'importantes variations morphologiques, la modification de l'expression de certains gènes, la réorganisation du caryotype, et l'augmentation de l'adhésion cellulaire et de la sécrétion d'enzymes hydrolytiques (Gomes et al., 2007; Miranda et al., 2007). Cela a permis de mettre en évidence le rôle majeur de l'altération du code génétique dans l'évolution des espèces du clade CTG, par l'intermédiaire de la génération d'une diversité phénotypique importante (Gomes et al., 2007; Miranda et al., 2007). Par exemple, l'augmentation de la diversité des protéines de surface rend *C. albicans* plus difficile à détecter par le système immunitaire (Miranda et al., 2013).

La génération de cette diversité phénotypique a donc représenté un fort potentiel sélectif pour les espèces de *Candida* soutenant l'hypothèse d'une accélération de l'évolution de nouveaux phénotypes.

L'une des premières études réalisées à l'échelle du génome, comparant les variations génomiques de 7 membres du clade CTG et 9 membres du clade WGD a été menée par Geraldine Butler et ses collaborateurs en 2009 (Butler et al., 2009). De façon intéressante, la comparaison de ces génomes a permis d'identifier plusieurs familles de gènes particulièrement représentées chez les espèces pathogènes de *Candida* par opposition aux levures non pathogènes. En effet, 21 familles de gènes sont surreprésentées dans les espèces pathogènes dont 3 familles associées à la paroi cellulaire, comprenant des gènes codant pour les adhésines Als (agglutinin-like sequence), les protéines Hyr1/Iff et les protéines glycosyl-phosphatidylinositol (GPI) de type Pga30-like (Butler et al., 2009).

L'expansion de ces familles de gènes est principalement survenue via la duplication de ces gènes. Par exemple chez *C. albicans*, le gène *ALS5* provient de la duplication du gène *ALS1* (Jackson et al., 2009).

De la même façon, ce type d'événement a été observé pour les familles de gènes *SAP*, *TLO* et la famille de gènes *IFA* riches en leucines codant pour des protéines transmembranaires putatives. Ce sont les *TLO* et *IFA* qui montrent la plus grande divergence entre les espèces de *Candida*. Une étude réalisée par Jackson *et al.* a montré que *C. albicans* possède 31 gènes de la famille *IFA* tandis que *C. tropicalis* n'en a qu'un, mais 19,4% d'entre eux (6/31) sont considérés comme non fonctionnels (Jackson *et al.*, 2009). La comparaison du génome de *C. albicans* à celui de *C. dubliniensis*, espèce moins pathogène, a montré que *C. dubliniensis* comptait seulement 21 membres de la famille des gènes *IFA*, dont la majorité 66,7% (14/21) n'est pas fonctionnelle (Jackson *et al.*, 2009). Le gène *TLO* ancestral, quant à lui, semble n'avoir subi qu'une seule duplication chez *C. dubliniensis*, alors que l'expansion de celui-ci est plus importante chez *C. albicans*. De plus, il n'existe pas de gène équivalent au gène *ALS3* codant pour l'invasine chez *C. dubliniensis* (ou chez toute autre espèce de *Candida*), ni de gène correspondant au gène *HYR1* de *C. albicans* (Jackson *et al.*, 2009). Les résultats de cette étude soulignent la divergence de l'histoire évolutive entre *C. albicans* et *C. dubliniensis*. Tandis que *C. albicans* a vu augmenter le nombre de ses gènes contribuant à la pathogénie, *C. dubliniensis* a perdu sa capacité génomique et ses fonctions pathogènes majeures, ce qui pourrait expliquer la différence de pouvoir pathogène entre ces 2 espèces (Jackson *et al.*, 2009).

Ces résultats suggèrent que certains traits de virulence peuvent survenir « rapidement » via de simples réarrangements génomiques. La duplication de certains gènes suivie d'une divergence des séquences est une source essentielle de variabilité génétique chez les espèces de *Candida* pathogènes. Ces événements permettent de sous-tendre de nouveaux caractères phénotypiques et sont généralement indicateurs d'une sélection directionnelle récente (Gabaldón *et al.*, 2016). Ces études ont permis de mettre en évidence l'importance des constituants de la paroi cellulaire, jouant probablement un rôle majeur dans les interactions hôte-pathogène (Gabaldón *et al.*, 2016).

2.1.2 Diversité génétique de *C. albicans*

Jusqu'à la fin des années 2000, les méthodes de typage moléculaire conventionnelles telles que les microsatellites (MLMT), le Ca3 fingerprint et le MLST, ont représenté les outils de base pour explorer la diversité génétique des souches de *C. albicans*. Grâce au génotypage de plusieurs centaines d'isolats cliniques, associée à des études de génétique des populations, la structure de la population de *C. albicans* a pu être déterminée (Odds et al., 2007).

2.1.2.1 Structure de la population de *C. albicans*

Quelle que soit la technique de typage moléculaire utilisée, celles-ci ont révélé l'existence de différents groupes génétiques ou « clades » regroupant des isolats génétiquement proches (Bougnoux et al., 2004; Chávez-Galarza et al., 2010; Dalle et al., 2000; Odds et al., 2007; Pujol et al., 1997). Les études de génomique des populations ont permis de montrer que la structure de la population de *C. albicans* est principalement clonale (Pujol et al., 1997).

Actuellement, on considère que la population de *C. albicans* est organisée en 18 clusters génétiquement distincts dont 5 clades majeurs (Clade 1, 2, 3, 4, 11) (**Figure 10**) (Odds et al., 2007; Shin et al., 2011).

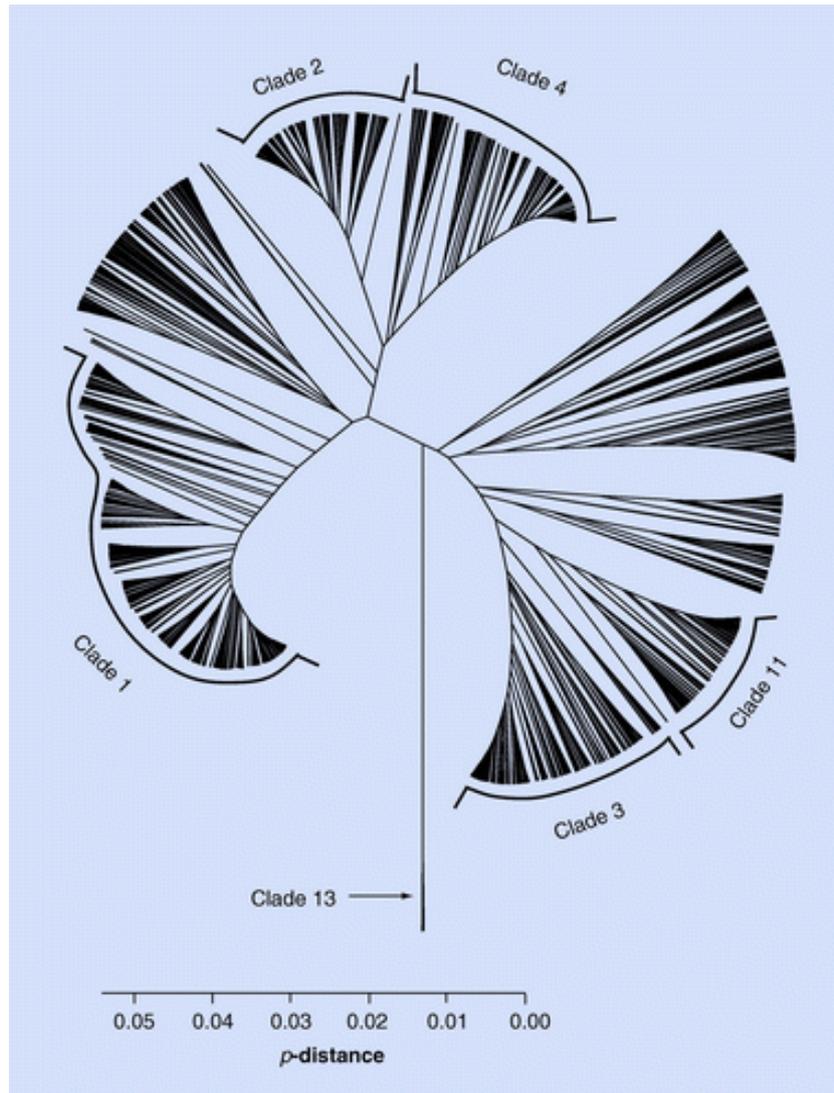


Figure 10 : Dendrogramme réalisé à partir des DST (Diploïdes Sequence Typing) obtenus par typage MLST de 1410 souches de *C. albicans*. Sont indiqués les 5 principaux clades (1, 2, 3, 4 et 11) ainsi que le clade 13 correspondant au clade le plus éloigné (Odds, 2010).

2.1.2.2 Caractéristiques distinguant les clades

L'étude réalisée par Odds *et al.* en 2007, sur un collectif très important de souches (1391 isolats d'origine géographique variée), a apporté une vision nouvelle sur la diversité au sein de la population de *C. albicans* (Odds *et al.*, 2007). Dans cette étude, il a pu être montré que certains clades étaient enrichis avec des souches d'origine géographique particulière (**Tableau 1**).

Par exemple le clade 8 était enrichi en isolats sud-américains, le clade 4 en isolats du Moyen-Orient et d'Afrique, le clade 2 en isolats du Royaume-Uni, le clade 11 en isolats d'Europe continentale, les clades 14 à 18 en isolats provenant d'Asie (**Tableau 1**) (McManus and Coleman, 2014; Odds et al., 2007; Shin et al., 2011).

Plusieurs équipes ont tenté de corréler l'observation de certains phénotypes des souches à l'appartenance à un clade spécifique. Par exemple, l'étude réalisée par Odds *et al.* en 2007 a montré l'absence de corrélation entre l'origine des souches (d'origine infectieuse ou commensale) et l'appartenance à un clade spécifique (Odds et al., 2007).

En revanche, de façon intéressante, le clade 1 (clade prédominant représentant 33,6 % des 1391 isolats), regroupe des isolats qui présentent des caractéristiques particulières comme :
i. une sensibilité diminuée à la 5-fluorocytosine (5-FC), en lien avec la diffusion clonale de la substitution C301T du gène *FUR1* (codant pour une uracile phosphoribosyl transférase), ii. une sensibilité diminuée à la terbinafine, iii. une tolérance à de fortes concentrations en NaCl, iv. des variations dans le nombre de répétitions en tandem dans des gènes impliqués dans la synthèse de protéines de surface cellulaire, comme le gène *ALS7*. Ces variations pourraient avoir un lien avec le caractère pathogène de certaines souches (Dodgson et al., 2004; MacCallum et al., 2009; Odds, 2010; Odds et al., 2007; Odds and Jacobsen, 2008; Tavanti et al., 2005; Zhang, 2003).

D'autres études ont également révélé une association significative entre la diminution de l'activité des phosphatases acides et les isolats appartenant au clade 2, ainsi qu'une diminution de la sensibilité à l'amphotéricine B chez les souches de clade 4 (Blignaut et al., 2005; MacCallum et al., 2009; Odds et al., 2007). Une diminution de la sensibilité des souches des clades 5 et 6 aux antifongiques azolés a été observée sans que toutefois cette différence soit significative (Tavanti et al., 2005).

Le clade 13, quant à lui, regroupe des isolats phénotypiquement atypiques, incapables, par exemple, de former des chlamydospores (Odds et al., 2007). Ces isolats sont nommés *Candida africana* parce que les 1ères souches décrites ont été isolées en Afrique. Ces souches ont en réalité une distribution mondiale et ont une extraordinaire capacité à coloniser la sphère génitale des femmes et à causer des vaginites (Chowdhary et al., 2017).

Leur distance génétique par rapport aux autres isolats de *C. albicans* ne permet cependant pas d'affirmer qu'elles forment une espèce indépendante.

Enfin, une étude réalisée par Calderon-Norena *et al.* sur 144 isolats, a montré que les souches appartenant aux clades 1 et 3 présentaient un SNP hétérozygote sur le gène codant pour le facteur de transcription Ace2, tandis que les isolats des autres clades présentaient ce SNP à l'état homozygote (Calderón-Noreña *et al.*, 2015). Cette modification induit un codon stop prématuré sur les deux allèles, ce qui entraîne la synthèse d'une version tronquée d'Ace2. Cette anomalie a pour conséquence une altération de la croissance des hyphes de *C. albicans* (Calderón-Noreña *et al.*, 2015) dont le rôle est essentiel dans l'initiation de l'infection. Cependant l'impact de ce phénotype au cours de l'interaction de *C. albicans* avec l'hôte reste à élucider.

Par ailleurs, les isolats de *C. albicans* peuvent également se distinguer selon leur génotype « A, B, ou C » déterminé en fonction de la présence (à l'état homozygote ou hétérozygote) ou de l'absence d'un transposon retrouvé dans la région ITS (Internal Transcribed Spacer) du gène de l'ARNr 28S (McCullough *et al.*, 1999). Certains des clades MLST sont enrichis avec des génotypes ABC particuliers (**Tableau 1**) (Odds *et al.*, 2007).

Clade ^a	Origin of isolates	ABC genotype ^b	Geographical enrichment ^c	Reduced susceptibility	Other clade-specific associations	References
1	Superficial infection Vaginal infection Oral carriage	A	Worldwide	FLC, 5FC, Terbafine	Growth in 2 M NaCl	Tavanti et al. (2005), Odds et al. (2007), McManus et al. (2012), Ge et al. (2012), MacCallum et al. (2009) and Abdulrahim et al. (2013)
2	Bloodstream	A	UK		Lower acid phosphatase activity	Tavanti et al. (2005), Takakura et al. (2008), Odds et al. (2007) and MacCallum et al. (2009)
3	Oropharyngeal	B	USA			Odds et al. (2007), Tavanti et al. (2005), Takakura et al. (2008)
4	Bloodstream	B/C	Middle east/Africa	Amphotericin B		Blignaut et al. (2005), Odds et al. (2007), Tavanti et al. (2005)
5	Oropharyngeal	B/C	Europe/UK	FLC		Odds et al. (2007), Tavanti et al. (2005)
6	Oropharyngeal	B	UK	FLC		Odds et al. (2007), Tavanti et al. (2005)
7		A				Odds et al. (2007), Tavanti et al. (2005)
8	Bloodstream Wildlife animals	A/B	South America			Odds et al. (2007)
9		A				Odds et al. (2007)
10		B	Continental Europe			Odds et al. (2007)
11		A/C	Europe/UK			Odds et al. (2007)
12		B/C				Odds et al. (2007)
13	Vaginal	A	Africa			Odds et al. (2007)
14		B	Asia			Gong et al. (2012), Odds et al. (2007)
15		B/C	Asia			Odds et al. (2007)
16		B	Asia			Odds et al. (2007)
17		A/B	Asia			Odds et al. (2007)
18	Dyspeptic patients	ND	Asia			Gong et al. (2012), Shin et al. (2011)

Abbreviations: UK, United Kingdom; USA, United States of America; FLC, fluconazole; 5FC, 5-fluorocytosine.

a. Clade numbers defined by MLST (Odds et al., 2007).

b. ABC genotypes are defined according to the absence, presence or heterozygous presence of an intron in the 25S rRNA gene (McCullough et al., 1999) and demonstrated statistically significant clade-enrichments (Abdulrahim et al., 2013; Odds et al., 2007).

c. Geographical locations demonstrated statistically significant clade-enrichments (Blignaut et al., 2005; Odds et al., 2007; Takakura et al., 2008;

Tableau 1 : Résumé des principales caractéristiques associées aux clades MLST de *C. albicans*. (d'après McManus and Coleman, 2014). (McManus and Coleman, 2014)

À ce jour, la base de données MLST disponible en ligne (<https://pubmlst.org/calbicans/>) contient 3490 profils MLST (DST) obtenus à partir de 4325 isolats de *C. albicans* isolés à travers le monde (date d'accès 12/05/18). La mise à disposition de ces données permet d'améliorer continuellement nos connaissances sur la dispersion des souches de *C. albicans*, ainsi que leur évolution au cours du temps. Si les études de génétique des populations, basées sur l'utilisation du typage moléculaire (que nous avons décrit précédemment), ont révélé très tôt que la structure de la population de *C. albicans* était principalement clonale, l'utilisation des données de séquençage du génome de *C. albicans* a permis de mieux comprendre l'évolution de cette espèce. L'étude génomique des populations a ainsi dévoilé le rôle de la parasexualité dans la diversification de *C. albicans* et l'apparition de nouveaux clades (Ropars et al., 2018).

2.1.2.3 Variabilité génomique de *C. albicans*

Dans l'étude d'Hirakawa *et al.* que nous avons citée précédemment, 21 isolats épidémiologiquement non liés de *C. albicans* ont été sélectionnés pour une analyse comparative phénotypique et génomique étendue. Ces isolats provenaient de différentes origines géographiques et étaient distribués au sein de 7 des 18 clades précédemment décrits (dont les 5 clades majeurs) (Hirakawa *et al.*, 2015). L'analyse de la diversité génétique entre ces isolats a révélé des variations dans la structure de leurs génomes, avec la mise en évidence d'événements d'aneuploïdie, mais également des variations significatives en termes de composition du génome (hétérozygotie/SNPs/LOH) pouvant avoir des conséquences phénotypiques.

2.1.2.3.1 Variations de la composition du génome de *C. albicans* et conséquences phénotypiques

- Hétérozygotie

Cette étude a permis de confirmer un niveau important d'hétérozygotie dans les génomes de *C. albicans* avec en moyenne un 1 SNP hétérozygote toutes les 267 pb. L'analyse comparative des génomes appartenant au même clade que la souche de référence SC5314 (clade 1) a montré, en moyenne, 1 variation homozygote toutes les 1404 pb, alors que celle-ci est de 1 SNP homozygote toutes les 235 pb pour les souches des autres clades. Grâce à ces données, la diversité nucléotidique moyenne calculée à l'échelle du génome entre deux isolats de *C. albicans* a été estimée à 0,37 % (Hirakawa *et al.*, 2015).

La répartition des SNPs n'était pas homogène sur le génome. L'analyse des gènes présentant des taux élevés de substitutions non-synonymes a permis d'identifier des protéines à évolution rapide, incluant des protéines de structure de la paroi cellulaire ou impliqué dans la voie de signalisation de l'intégrité de celle-ci. Cela souligne à nouveau le rôle central des protéines de la paroi cellulaire dans les interactions hôte-pathogène et dans l'adaptation aux modifications environnementales de *C. albicans* (Hirakawa *et al.*, 2015).

- Perte d'hétérozygotie (LOH)

L'analyse de la distribution des SNPs hétérozygotes à travers les différents chromosomes a permis de détecter les zones ayant subi des LOH. Toutes les souches dans cette étude présentaient au moins un LOH et tous les chromosomes étaient impactés, dont certains en totalité (Chr 3, 5, 6, 7 et R). En revanche, l'origine du LOH différait d'une souche à l'autre. Comme nous l'avons vu précédemment, seule l'extrémité du bras droit du ChrR contenant l'ADNr était homozygote dans toutes les souches (Hirakawa et al., 2015).

- Conséquences moléculaires et phénotypiques de l'hétérozygotie/SNPs/LOH

Le niveau élevé d'hétérozygotie, confirmé à l'échelle du génome sur tous les isolats de *C. albicans* séquencés, est l'occasion d'aborder les conséquences moléculaires et phénotypiques de ce type de variation.

Les conséquences d'une mutation sur le phénotype de la souche vont dépendre de la région du génome qui est affectée. Par exemple, les mutations survenant dans des régions intergéniques ou des introns n'ont souvent pas de conséquences directes, tandis qu'une mutation dans la séquence codante d'un gène peut modifier son activité. Une mutation dans une région codante peut induire une modification de l'acide aminé traduit, on parle alors de SNP « non-synonyme ». La mutation peut être « conservative » ou « non-conservative » si l'acide aminé modifié dispose ou non des mêmes propriétés physico-chimiques que l'acide aminé initial. Enfin, les mutations dites « non-sens » peuvent induire la survenue ou la suppression d'un codon stop conduisant à une protéine potentiellement défectueuse, comme dans le cas décrit précédemment (Ace2) (Calderón-Noreña et al., 2015).

Les mutations hétérozygotes (affectant l'un des deux allèles des chromosomes homologues) peuvent être responsables de la synthèse de deux protéines différentes, pouvant avoir un impact fonctionnel significatif sur la souche. De la même façon, il a été montré que l'expression de l'un des deux allèles pouvait être favorisée sur le plan transcriptionnel et traductionnel (Muzzey et al., 2014, 2013).

De façon intéressante dans l'étude réalisée par Hirakawa *et al.*, l'hétérozygotie moyenne retrouvée sur la totalité du génome (allant de 48 à 89%) était corrélée avec le taux de croissance des souches (YPD à 30°C) (Hirakawa *et al.*, 2015). Les souches les plus hétérozygotes présentant un meilleur fitness (Hirakawa *et al.*, 2015). Ces résultats sont concordants avec de précédentes études ayant montré l'avantage de l'hétérozygotie sur le fitness des souches (perte de fitness de la souche SC5314 rendue complètement homozygote par auto-diploïdisation du mutant haploïde) (Hickman *et al.*, 2013). En revanche il n'a pas été retrouvé de corrélation avec la virulence des souches dans différents modèles d'infection tant chez la souris que chez *Galleria sp.* (Hirakawa *et al.*, 2015).

2.1.2.3.2 Variabilité de la structure du génome de *C. albicans* et conséquences phénotypiques

- Variations dans le nombre de copies (CNVs) des ORFs et conséquences évolutives

En ce qui concerne le contenu génétique des 21 isolats de *C. albicans* étudiés par Hirakawa *et al.*, celui-ci était largement conservé entre les différents génomes analysés. En moyenne, 98,1% (6069/6189) des gènes décrits chez SC5314 ont également été identifiés dans les 21 isolats séquencés (Hirakawa *et al.*, 2015). Cependant, le nombre moyen de gènes identifiés par génome était de 6268, suggérant la présence de gènes supplémentaires absents dans la souche SC5314.

Indépendamment des aneuploïdies, des CNVs de certains ORFs ont également été observées dans le génome de quelques isolats, notamment aux niveaux de rétrotransposons actifs dans le gène *CYP5* et dans les membres de la famille des gènes *TLO* (Hirakawa *et al.*, 2015).

Le gène *CYP5* code pour une enzyme de la famille des cyclophilines. Ces protéines peuvent être impliquées dans divers processus comme la réponse au stress, et ont montré un rôle dans la pathogénicité chez les champignons (Viaud *et al.*, 2002), suggérant que la duplication de ce gène chez *C. albicans* pourrait également contribuer à l'amélioration du fitness ou à la virulence de ces souches.

En plus des CNVs, les gènes *TLO* ont présenté des variations de localisation dans le génome. De récentes études ont rapporté que les gènes *TLO*, qui codent pour la sous-unité Med2 du complexe formant le Médiateur, ont un rôle dans la régulation transcriptionnelle de gènes impliqués dans l'adaptation chez l'hôte et la virulence (Flanagan et al., 2018; Haran et al., 2014; Sullivan et al., 2015). Comme nous l'avons vu précédemment cette famille de gènes présente d'importantes variations, en termes de nombre de copies, entre *C. albicans* et les autres espèces de *Candida* moins pathogènes, indiquant que ces gènes ont probablement eu un rôle important en termes d'évolution génomique chez *C. albicans* (Butler et al., 2009).

- Réarrangements chromosomiques et conséquences phénotypiques

Des modifications structurales des chromosomes ont également été observées entre les 21 isolats étudiés par Hirakawa *et al.* En effet, de façon surprenante plus d'un tiers des isolats dans cette étude présentaient des événements d'aneuploïdies, dont la fréquence importante a déjà été discutée dans le paragraphe 1.2.1.1. De façon plus détaillée, parmi les 21 souches étudiées, 7 présentaient un ou plusieurs chromosomes trisomiques (Chr4 à Chr7), suggérant que ces événements sont bien tolérés. De plus, des duplications segmentales (tri- et tétrasomiques) ont été observées au niveau de deux chromosomes (ChrR et Chr1) et une souche était hémizygote (une seule copie d'un allèle présent) sur les 200 derniers kb du bras droit du Chr 3 (Hirakawa et al., 2015).

Dans cette étude, les souches aneuploïdes présentaient des défauts de croissance en comparaison avec les souches euploïdes. Ces résultats sont concordants avec d'autres études menées chez *C. albicans* ou *S. cerevisiae* montrant une altération du fitness de souches aneuploïdes cultivées sur des milieux complets (Forche et al., 2008; Pavelka et al., 2010; Siegel and Amon, 2012; Torres et al., 2007). Cependant, toutes les souches aneuploïdes ne présentent pas forcément un défaut de croissance, et peuvent donc être maintenues pendant un certain temps en l'absence de sélection. Dans certaines conditions de stress, comme l'utilisation d'antifongiques azolés, l'aneuploïdie peut même conférer un avantage sélectif comme l'illustre la **Figure 11** (Bennett et al., 2014; Janbon et al., 1998; Wertheimer et al., 2016).

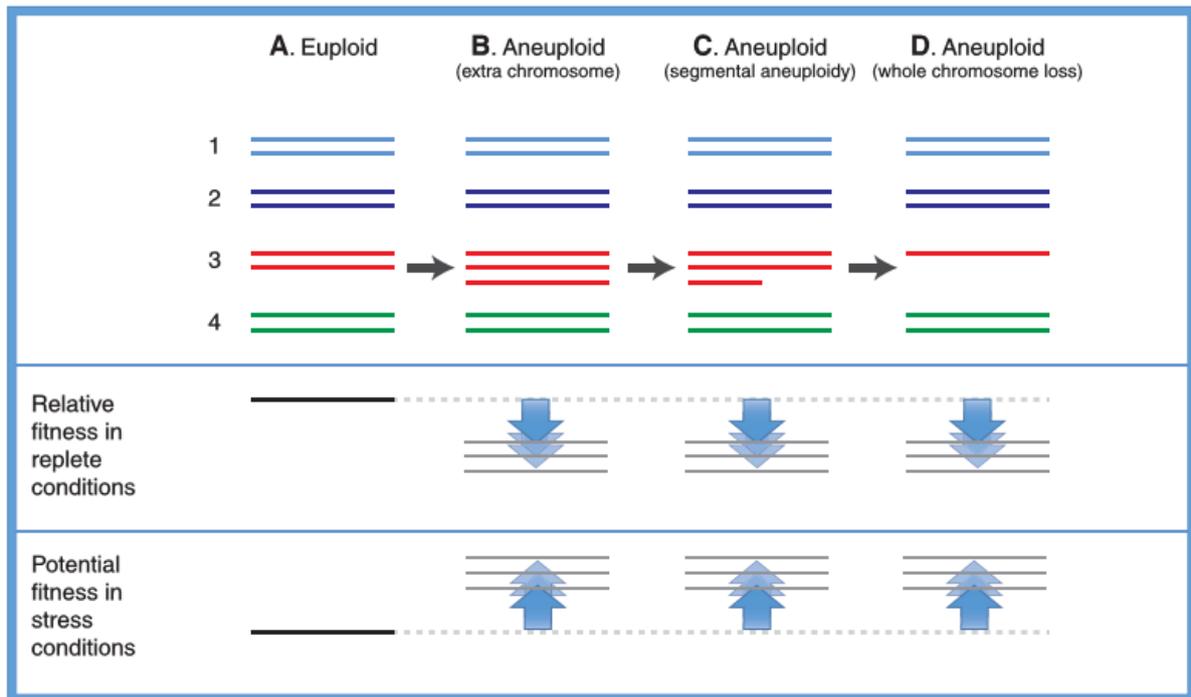


Figure 11 : Comparaison du fitness d'organismes euploïdes ou aneuploïdes en fonction des conditions environnementales. Panel supérieur : Représentation schématique des variations de la ploïdie (A) Exemple de 4 paires de chromosomes (euploïdies). (B) Trisomie du chromosome 3 (Chromosome surnuméraire) (C) Aneuploïdie segmentaire du bras gauche du Chr3. (D) Perte d'un des deux Chr3. Panel inférieur : Impact sur le fitness des événements d'aneuploïdies. En l'absence de stress environnemental, la plupart des souches aneuploïdes présentent un défaut de fitness (condition de culture en milieu riche) alors que confrontées à un stress particulier, certaines aneuploïdies spécifiques peuvent être favorable et conférer un avantage sélectif (Bennett et al., 2014).

Ainsi, l'analyse de la diversité génétique entre des isolats de *C. albicans* épidémiologiquement non liés a révélé de nombreuses variations en termes de composition du génome (hétérozygotie/LOH/SNPs) et un certain nombre de variations dans la structure de celui-ci, avec la mise en évidence d'aneuploïdies. Il est donc intéressant de noter que les variations phénotypiques observées entre les isolats de *C. albicans* sont essentiellement dues aux SNPs/LOH qui représentent la principale source de variabilité génétique chez cette levure. L'étude d'Hirakawa *et al.* représente une photographie de la diversité génomique au sein de l'espèce à travers l'analyse de 21 génomes (Hirakawa et al., 2015).

Nous allons décrire dans la suite de cette introduction les types de signaux environnementaux qui peuvent induire la survenue de ces événements (CNVs, LOH, SNPs), et générer à court terme de la variabilité génomique chez *C. albicans*.

2.2 Adaptation de *C. albicans* à son environnement

C. albicans est d'une part un commensal du tube digestif de l'homme, qui constitue son principal réservoir, et d'autre part un pathogène opportuniste responsable d'infections dont la sévérité est fonction de facteurs favorisant chez l'hôte. Lors de l'interaction avec l'hôte, que ce soit durant le commensalisme ou durant l'infection, *C. albicans* est confronté à un certain nombre de stress environnementaux et immunitaires. Cette levure dispose donc d'une capacité à s'adapter à des environnements changeants chez l'hôte, tels que l'hypoxie, le pH, le stress oxydatif par contact avec les macrophages, la fièvre ou les traitements antifongiques... (Forche, 2014). Cela peut se produire à court terme, par une modification réversible de la régulation de l'expression de certains gènes, de façon à adapter l'activité enzymatique en fonction des besoins cellulaires du moment (Kumamoto, 2008).

De nombreuses études réalisées chez les bactéries ou chez d'autres champignons ont démontré le rôle de la diversification génétique dans l'adaptation des organismes aux modifications environnementales (Edmond et al., 1999; Hickman et al., 2013; Wertheimer et al., 2016). Chez *C. albicans*, la génération de variants génomiques est grandement facilitée par la plasticité de son génome. En effet, comme nous l'avons vu précédemment, en plus de l'hétérozygotie importante, des événements majeurs de remaniement des chromosomes peuvent survenir induisant une modification du nombre de copies de gènes situés sur le chromosome affecté. La variation du nombre de copies de ces gènes peut ainsi conférer un avantage sélectif dans certaines conditions (Janbon et al., 1998; Selmecki et al., 2008). Cependant les aneuploïdies sont instables et donc souvent transitoires, l'état diploïde peut être retrouvé en l'absence de pression de sélection (Gerstein and Berman, 2015). De simples mutations ponctuelles peuvent également survenir dans certains gènes dont la fonction permet aux cellules, de façon durable, de mieux tolérer l'environnement défavorable.

L'acquisition de mutations avantageuses est fréquemment suivie d'une perte d'hétérozygotie (LOH) de l'allèle muté par des événements de recombinaison mitotique, ou par perte du chromosome contenant l'allèle de type sauvage (Coste et al., 2006; Dunkel et al., 2008). La survenue de ces événements a largement été étudiée dans le cadre de l'adaptation de *C. albicans* en réponse aux antifongiques (Hampe et al., 2017; Selmecki et al., 2010) mais également en réponse à diverses modifications environnementales *in vitro* et *in vivo*, que nous allons passer en revue dans cette partie (Forche, 2014).

2.2.1 Adaptation *in vitro* de *C. albicans* en réponse aux stress

2.2.1.1 Survenue de SNPs et de LOH en condition de stress et conséquences phénotypiques

Chez les organismes ayant un mode de reproduction clonale, l'hétérozygotie est sensée rester « fixe », ce qui se traduit par une diversité génétique élevée au niveau de l'isolat, mais par un nombre réduit de génotypes différents au sein de la population (Forche, 2014). Cependant, les événements de LOH sont beaucoup plus fréquents (10^{-6} par génération) que les mutations ponctuelles (10^{-9} par génération) (Lang and Murray, 2008). Malgré cela, *C. albicans* a maintenu un niveau élevé d'hétérozygotie ainsi qu'une grande diversité génotypique au sein de la population. Les échanges génétiques entre les clades et au sein des clades sont rares, cela suggère que l'hétérozygotie pourrait être un avantage (contre sélection des LOH), ou que d'autres mécanismes permettent de maintenir la diversité génétique de cette espèce (Bougnoux et al., 2008; Forche, 2014; Ropars et al., 2018). Le recours au cycle parasexuel, en conditions de stress, pourrait permettre l'apparition de nouveaux variants génomiques au cours de la perte concertée de chromosomes (Bennett and Johnson, 2003; Forche et al., 2008).

Plusieurs études ont évalué la fréquence, mais également l'impact sur le fitness, de la survenue des LOH dans différentes conditions de stress ainsi que la contribution de ces LOH à la diversité génétique au sein de la population (Bougnoux et al., 2008; Forche et al., 2011; Hickman et al., 2013). Bien que ces événements tendent à diminuer la diversité allélique au sein de la cellule, ils facilitent par ailleurs l'adaptation de la souche, en révélant des allèles récessifs qui peuvent fournir un avantage dans des conditions défavorables (Abbey et al., 2011).

In vitro, l'augmentation de la fréquence des LOH chez la souche SC5314 a été observée en réponse à divers stress environnementaux, tels que le stress oxydatif, l'augmentation de la température (choc thermique), l'exposition au fluconazole (antifongique azolé) ainsi que l'augmentation du glucose ou sorbose dans le milieu de culture (conditions utilisées pour induire un cycle parasexuel) (Berman and Hadany, 2012; Forche et al., 2011).

De façon intéressante, la nature des LOH varie également en fonction du type de stress. Un stress oxydatif augmente la proportion de petits LOH de type Conversion génique, tandis que les autres stress augmentent la proportion de LOH impactant la totalité d'un chromosome (**Figure 12**). Biologiquement cela peut s'expliquer par le fait que le stress oxydatif induit des cassures de l'ADN, alors que le choc thermique et le fluconazole induisent vraisemblablement des anomalies dans la ségrégation des chromosomes (Berman and Hadany, 2012; Forche et al., 2011).

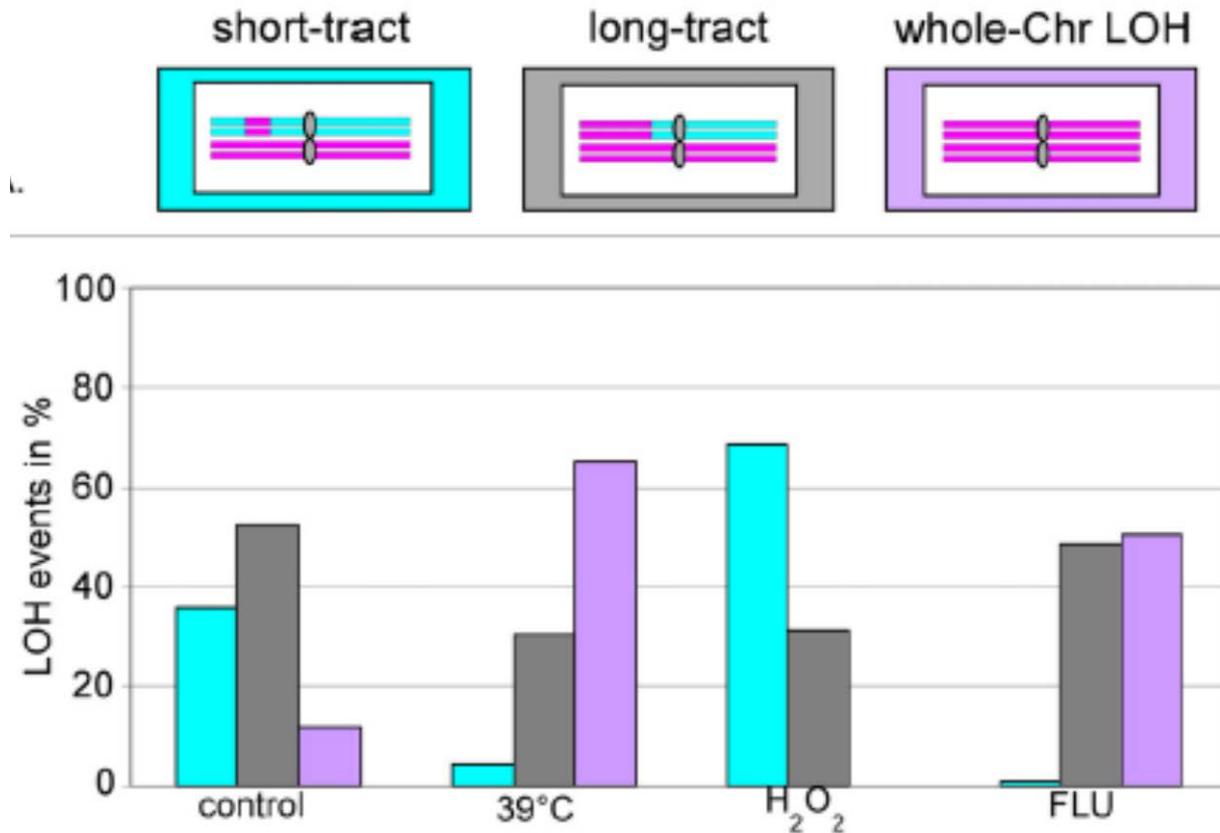


Figure 12 : Impact de différents stress sur le type et la fréquence de survenue d'événements de LOH. La taille des LOH : courts (bleu), longs (gris) et chromosome entiers (rose) sont illustrés ci-dessus et encadrés dans la couleur correspondant à chacun des types. La fréquence de survenue de chaque type de LOH est représentée sur les histogrammes pour chaque stress. Le stress oxydatif (H₂O₂) augmente la proportion de petits LOH, tandis que les autres stress augmentent la proportion de LOH impactant un grand fragment, ou la totalité d'un chromosome (d'après Forche et al., 2011).

- Cas particulier du stress induit par les antifongiques azolés

Ces événements peuvent avoir des conséquences phénotypiques cliniquement importantes, comme par exemple au cours de l'acquisition de la résistance aux traitements antifongiques. Chez *C. albicans*, les mutations décrites comme étant associées à des niveaux élevés de résistance aux antifongiques sont généralement observées à l'état homozygote. La transition, via un mécanisme de LOH, d'une mutation initialement hétérozygote à l'état homozygote contribue à augmenter le niveau de résistance de la souche (Forche, 2014). En effet, une étude réalisée par Coste *et al.* a montré le rôle des LOH impactant une région du Chr5 sur le niveau de résistance au fluconazole (Coste et al., 2006).

Le gène *TAC1* (qui code pour un facteur de transcription qui régule l'expression de pompes à efflux *CDR1* et *CDR2*) est localisé au niveau du bras gauche du Chr5, à proximité du locus MTL (**Figure 13, b**). Les mutations gain de fonction (GOF) de *TAC1* acquises lors de l'exposition au fluconazole peuvent devenir homozygotes à la suite d'un LOH, conséquence d'une réplication induite par une cassure de l'ADN (BIR, Breack Induced Replication) ou d'un événement de recombinaison (single crossover) (Coste et al., 2006). Des mutations ponctuelles du gène *ERG11* (localisé sur le bras gauche du Chr5 et codant pour la cible du fluconazole, la lanostérol 14- α -déméthylase) mais également du gène *MRR1* (localisé sur le Chr3 et codant pour un facteur de transcription qui régule l'expression du transporteur *Mdr1*) peuvent également être rendues homozygotes par des événements de LOH similaires à ceux de *TAC1* (Coste et al., 2006; Dunkel et al., 2008).

- Cas particulier du stress induit par les macrophages

Un autre exemple de phénomène d'adaptation rapide observé *in vitro* chez *C. albicans* a été décrit par Wartenberg *et al.* en 2014, par une approche d'évolution expérimentale (Wartenberg et al., 2014).

Dans cette étude originale, les auteurs ont soumis un double mutant de *C. albicans* (*cph1 $\Delta\Delta$ /efg1 $\Delta\Delta$*), incapable de filamenter, à une pression de sélection continue par co-culture avec des macrophages, constituant un environnement hostile lié au phagosome (Wartenberg et al., 2014). De façon intéressante, le mutant a développé la capacité d'échapper aux macrophages en acquérant à nouveau la capacité de filamenter. De plus, le mutant évolué a montré une virulence accrue dans un modèle d'infection chez la souris et une modification de la composition de la paroi cellulaire, en comparaison avec la souche initiale. L'analyse du génome complet des deux souches (initiale et évoluée) a permis de détecter une mutation au niveau du gène *SSN3* qui code pour la sous-unité Cdk8 du médiateur. Cette mutation a permis à la souche de s'adapter à son environnement en contournant la voie de signalisation de la filamentation dépendante d'Efg1/Cph1. Ces résultats montrent qu'une modification phénotypique morphologique impliquant un réseau de transcription majeur peut être modifiée de façon rapide dans un contexte de pression de sélection spécifique *in vitro* (Wartenberg et al., 2014).

2.2.1.2 Survenue d'aneuploïdies en condition de stress et conséquences phénotypiques

L'exposition à des conditions de stress peut également induire l'apparition d'aneuploïdies, qui dans certains cas et pour certains chromosomes spécifiques, peuvent être stables, et également entraîner des modifications phénotypiques (Rustchenko et al., 1994). Le développement de méthodes d'analyse du génome, comme la CGH couplée à la cytométrie en flux, a grandement facilité la détection des réarrangements chromosomiques, permettant ainsi de décrire plus précisément la fréquence de survenue de ces événements dans différentes conditions (Abbey et al., 2011).

Plusieurs études ont, par exemple, montré que le choc thermique pouvait affecter la stabilité du génome de *C. albicans*. La première, réalisée en 1985 par Hilton *et al.*, avait déjà rapporté que l'exposition des souches à une température de 50°C était responsable de la perte de chromosomes (Hilton et al., 1985). Une étude plus récente, utilisant le CHEF, a cette fois-ci permis de détecter des gains ou pertes de chromosomes entiers après une exposition très courte (90 à 120 secondes) à 50°C (Bouchonville et al., 2009).

Le choc thermique entraîne une diminution de la viabilité des cellules, et donne lieu à des variations dans la taille des colonies. Les petites colonies montrent deux fois plus de modifications du caryotype que les grosses, suggérant que l'exposition, même limitée, des souches à des températures élevées entraîne une augmentation du nombre d'événements de non-disjonction des chromosomes homologues.

D'autres études ont montré que la survenue de ces événements pouvait être favorable dans certaines conditions de cultures. Par exemple, des réarrangements majeurs des Chr2 et 6 permettent à *C. albicans* de mieux se développer en présence d'arabinose-D (Rustchenko et al., 1994), tandis qu'une monosomie du Chr5 lui permet de se développer sur milieu contenant du sorbose comme seule source de carbone (Janbon et al., 1998).

L'impact de ces événements sur le fitness a été démontré par Abbey *et al.* par l'analyse du temps de génération (temps de doublement) de la souche SC5314 et de dérivés génétiquement modifiés (Abbey et al., 2011). Clairement, le temps de doublement des souches aneuploïdes (aneuploïdie totale ou partielle d'un chromosome) est significativement supérieur à celui du témoin euploïde.

En revanche pour les LOH, seuls les grand événements (>100 kb) entraînent une augmentation des temps de doublement des souches (Abbey et al., 2011).

- Cas particulier du stress induit par les antifongiques azolés

L'analyse par CGH d'isolats cliniques de *C. albicans* a permis de révéler un lien direct entre la résistance au fluconazole et la présence d'un isochromosome du bras gauche du Chr5 « i(5L) » (aneuploïdie segmentaire) (**Figure 13, a**) (Selmecki et al., 2008, 2006). Cet isochromosome correspond à un chromosome surnuméraire qui se compose de deux bras gauches du Chr5, reliés par un centromère, contenant les gènes *ERG11* et *TAC1*, mentionnés précédemment (**Figure 13, b**). Ainsi, la formation de i(5L) contribue à augmenter la résistance au fluconazole en augmentant le nombre de copies de ces gènes (**Figure 13, c**) (Selmecki et al., 2006). La fréquence à laquelle cet isochromosome apparaît ainsi que le coût en termes de fitness, ont été testés expérimentalement en présence et en absence de fluconazole (Selmecki et al., 2009). Cette étude a permis de montrer qu'un i(5L) pouvait se former rapidement -en quelques générations seulement après exposition des isolats au fluconazole- et pouvait être perdu lorsque les isolats n'étaient plus cultivés avec l'antifongique (Selmecki et al., 2006, 2009). Cette étude a par ailleurs montré que des isolats porteurs de l'i(5L) avaient également acquis d'autres aneuploïdies sur des chromosomes portant des gènes impliqués dans la résistance aux antifongiques.

L'accumulation d'aneuploïdies était corrélée avec l'augmentation des CMI (Concentrations Minimales Inhibitrices) (**Figure 13, c**). De façon surprenante, dans ce contexte, la présence de chromosomes supplémentaires n'a pas altéré la croissance des isolats, que ce soit avec ou sans fluconazole (Selmecki et al., 2009) ce qui confirme l'intérêt de ces événements dans des contextes spécifiques.

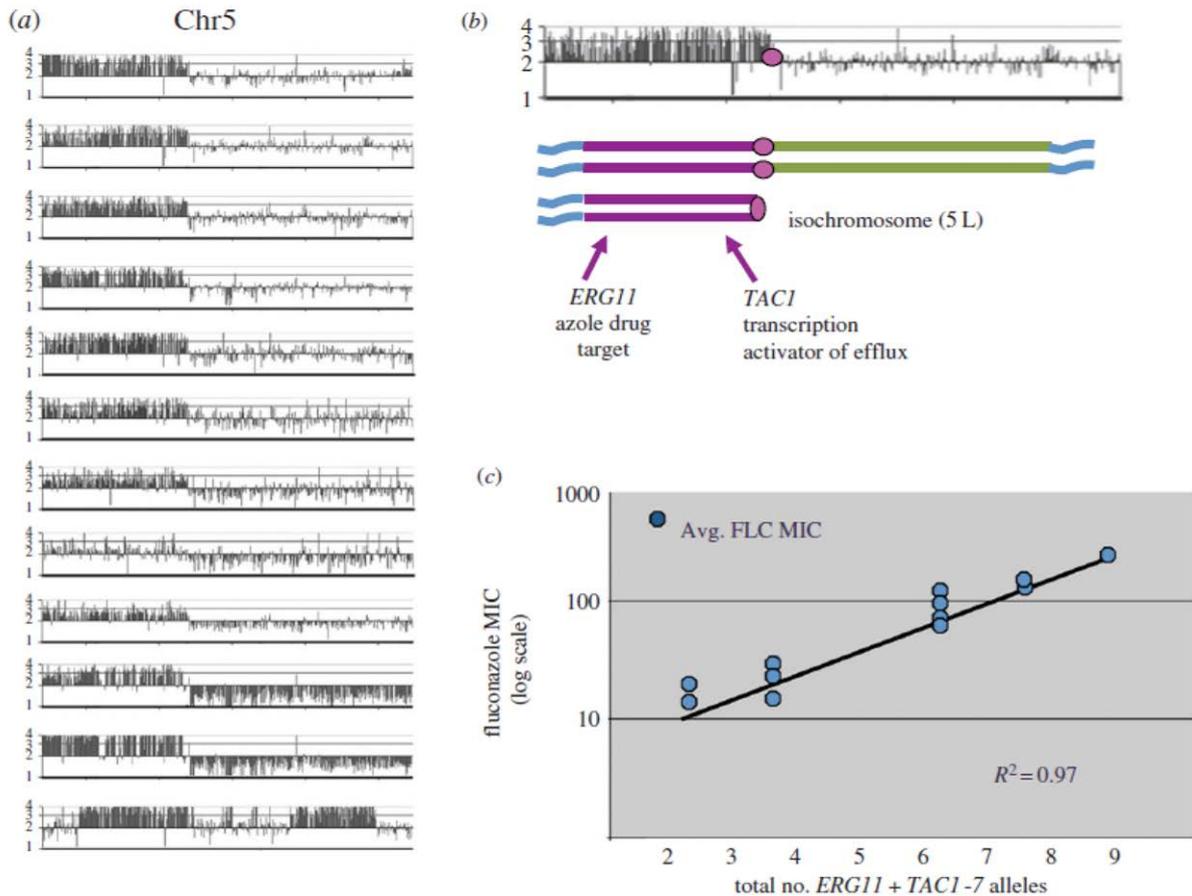


Figure 13 : Détection et impact de l'acquisition de certaines aneuploïdies et LOH dans la résistance au fluconazole. (a) Représentation du niveau de ploïdie (1, 2, 3 ou 4 copies) sur le Chr5 par calcul des CNVs en méthode CGH. Un isochromosome i(5L) a été détecté chez 12 isolats résistants au fluconazole à partir d'une étude portant sur 90 isolats cliniques ou de laboratoires (Selmecki et al., 2006). (b) Structure de l'i(5L). Le centromère (CEN5, cercle rose) correspond au point de démarrage de la détection de l'aneuploïdie du bras gauche du Chr5 (i(5L)) qui inclut le gène *ERG11*, codant pour la lanostérol 14- α -déméthylase -la cible des antifongiques azolés- et *TAC1*, qui code pour un facteur de transcription qui régule l'expression des gènes de pompe à efflux (*CDR1* et *CDR2*). (c) Augmentation du nombre de copies des gènes *ERG11* et *TAC1*-7. *TAC1*-7 est une version allélique hyperactive de *TAC1* (Coste et al., 2006), corrélée avec le niveau de résistance (CMI) des souches (Selmecki et al., 2008). (FLC = fluconazole). (d'après Wertheimer et al. *Phil. Trans. R. Soc. B*, 2016).

L'ensemble de ces études met en évidence la capacité de *C. albicans* à acquérir des variations génomiques comme des SNPs, des LOH et des aneuploïdies, lui permettant de survivre à des conditions défavorables.

Cependant, bien que ces études aient permis de mieux caractériser les conditions dans lesquelles ces événements pouvaient survenir, elles résultent d'observation *in vitro* ne reflétant que partiellement l'environnement complexe dans lequel ce pathogène évolue. En effet, plusieurs travaux suggèrent que *C. albicans* dispose d'un arsenal de mécanismes d'adaptation, mis en place au cours de l'interaction avec l'hôte, différents de ceux observés en laboratoire (Forche et al., 2018; Pande et al., 2013). C'est pourquoi seules des études réalisées dans des conditions de croissance *in vivo* peuvent permettre d'identifier les événements spécifiques (SNPs, LOH et aneuploïdies) favorables à ce champignon, pour survivre et persister dans les niches différentes de l'hôte et/ou provoquer une infection.

2.2.2 Adaptation *in vivo* de *C. albicans* au cours de l'interaction avec l'hôte

2.2.2.1 Evolution chez la souris

Plusieurs études ont abordé ces questions en utilisant un modèle de colonisation digestive ou d'infections invasives ou oropharyngées chez la souris.

Une première étude réalisée par Forche *et al.*, a permis de montrer que des variants génétiques et phénotypiques pouvaient apparaître rapidement au cours d'un passage unique chez l'hôte (Forche et al., 2009). En effet, après un passage *in vivo* d'isolats de *C. albicans* dans un modèle d'infection invasive chez la souris, des taux élevés de LOH (détectés par microarray) et des variations chromosomiques (détectées par méthode CGH) ont été observés. Par exemple, la fréquence de survenue des LOH durant l'infection était 28 fois supérieure à celle observée chez les isolats témoins cultivés *in vitro* en milieu liquide. Enfin, les isolats collectés *in vivo* présentaient également divers événements de remaniements chromosomiques induisant des altérations phénotypiques (modifications du temps de croissance et de l'aspect morphologiques des colonies) (Forche et al., 2009).

Afin de mieux comprendre dans quelle mesure s'effectue l'acquisition de variations génomiques au cours de l'infection, cette même équipe a effectué une étude expérimentale d'évolution dans un modèle de candidose oropharyngée chez la souris (Forche et al., 2018).

Durant cette étude, réalisée avec plusieurs centaines d'isolats, les auteurs ont analysé la diversité génotypique et phénotypique des populations cellulaires de *C. albicans*, après un passage de plusieurs jours dans la cavité buccale des souris (n=17). La ploïdie, ainsi que les événements de LOH, ont été déterminés par cytométrie en flux et séquençage de fragments d'ADN de type ddRADseq (Double digest restriction-site associated DNA sequencing). A nouveau, la fréquence d'apparition de ces événements était bien plus élevée (100x) dans la population de *C. albicans* collectée après passage *in vivo* que dans la population témoin cultivée *in vitro*. Des aneuploïdies et des LOH ont été observés sur tous les chromosomes, mais les aneuploïdies touchaient plus fréquemment les petits Chrs, tandis que les LOH impactant la totalité du Chr étaient plus fréquents chez les grand Chrs (Forche et al., 2018). La plupart des isolats présentaient plusieurs types de variations génomiques, suggérant que l'environnement inflammatoire retrouvé au cours de la candidose buccale chez la souris génère de façon rapide, (en moins d'un jour d'infection) de la diversité *de novo*, permettant à *C. albicans* de survivre et potentiellement de s'adapter à son nouvel environnement.

De façon très intéressante, certains événements de remaniement génomiques, tels que l'homozygotie du Chr1 ou encore une trisomie du Chr6 et du Chr5, ont été détectés dans les isolats infectants plusieurs souris, confirmant l'idée d'une possible adaptation spécifique du génome aux conditions rencontrées dans la cavité buccale au cours des premiers stades de l'infection (Forche et al., 2018). En effet, le Chr6 contient plusieurs gènes importants pour : i. la virulence (aspartyl-protéinases sécrétées, lipases et adhésines), ii. l'utilisation du carbone (gènes *NAG*, N-acetylglucosamine) et iii. la réparation des cassures double-brins de l'ADN (*RAD52*) (Forche et al., 2018). Il est intéressant de noter que la surexpression du gène *RAD52* augmente l'instabilité du génome. Par conséquent, une copie supplémentaire de ce gène pourrait entraîner une instabilité accrue du génome et l'amplification d'allèles spécifiques pouvant conférer un avantage au cours de l'infection oropharyngée chez la souris.

Ces études ont considérablement amélioré nos connaissances de la variabilité génomique de *C. albicans in vivo*. Cependant, le modèle murin de commensalisme ou d'infection ne correspond pas à l'environnement naturel dans lequel évolue *C. albicans* à long terme chez l'Homme.

2.2.2.2 Evolution chez l'Homme

Bien que l'altération des fonctions immunitaires et la dysbiose chez l'Homme demeurent les déterminants principaux de la survenue des infections opportunistes, l'espèce *C. albicans* dispose de capacités particulières permettant de passer d'un état commensal à un état pathogène. Cependant, toutes les souches de *C. albicans* ne sont pas équivalentes. La variabilité génotypique et phénotypique au sein de l'espèce peut également contribuer à des variations dans le développement et la sévérité de l'infection (Braunsdorf and LeibundGut-Landmann, 2018; Cavalieri et al., 2017). Comme nous l'avons déjà mentionné, au cours de l'infection, *C. albicans* rencontre divers environnements auxquels il doit s'adapter pour pouvoir croître et survivre. L'hétérogénéité génomique à l'échelle de la population cellulaire pourrait faciliter l'adaptation de cette levure à ces changements environnementaux en générant des variants capables de survivre dans un contexte ou un environnement particulier chez l'hôte.

Le tube digestif de l'Homme constitue le principal réservoir de cette espèce mais également le point de départ le plus fréquent des infections. En effet, il a été montré que la souche responsable de l'infection était souvent la même que la souche endogène retrouvée à l'état commensal (Kullberg and Arendrup, 2015; Lott et al., 2005). Par conséquent, le tube digestif représente une source intéressante de souches de *C. albicans* pour l'étude de la diversité génétique de cette espèce au cours du commensalisme, mais également des mécanismes génomiques adaptatifs survenant de façon naturelle chez l'hôte. Le défi est d'identifier des variants génétiques qui se traduisent par des modifications phénotypiques, impactant l'équilibre entre le commensalisme et la pathogénicité de *C. albicans* au cours de l'interaction avec l'hôte (Braunsdorf and LeibundGut-Landmann, 2018).

2.2.2.2.1 Diversité génétique de *C. albicans* au cours du commensalisme

La colonisation initiale par *C. albicans* peut se produire dès la naissance ou secondairement par transmission interhumaine (Bougnoux et al., 2006; Qi et al., 2005).

Différentes études - essentiellement des approches de typage moléculaire - ont montré qu'au cours du commensalisme, un individu pouvait être colonisé par une ou plusieurs souches de *C. albicans*, et qu'un nombre limité de souches génotypiquement apparentées pouvait coloniser un ou plusieurs sites d'un même individu (Bougnoux et al., 2006).

Ces études ont également montré qu'une même souche pouvait persister durant plusieurs années chez l'hôte, ou évoluer par le biais de variations génétiques mineures appelées « microvariations » ou « microévolution » (Angebault et al., 2013; Bougnoux et al., 2006; Jacobsen et al., 2008; Odds et al., 2006). Ces souches dites « microévoluées » diffèrent le plus souvent par un ou deux LOH au niveau des loci étudiés en MLST, résultant de mutations ponctuelles ou d'événements de recombinaison homologue (Bougnoux et al., 2006; Diogo et al., 2009; Jacobsen et al., 2008; Odds et al., 2006; Takagi et al., 2013; Xu et al., 1999). Une caractérisation plus fine de ces événements a montré qu'il s'agissait essentiellement de grands événements de LOH qui peuvent être concomitant à d'autres remaniements du génome tels que des altérations de la taille des chromosomes ou de la ploïdie (Diogo et al., 2009). Ces données suggèrent que le génome de *C. albicans* présente également une grande plasticité au cours du commensalisme.

Par ailleurs, la diversité génétique au sein de la population cellulaire de *C. albicans* chez l'hôte a été analysée par typage moléculaire (MLST), à travers des échantillons provenant d'individus sains ou porteurs d'infection vaginale (Jacobsen et al., 2008). Une diversité génétique plus importante a été observée entre les colonies collectées au sein des échantillons commensaux par rapport aux échantillons prélevés au cours des infections vaginales. Ce qui suggère l'existence d'un phénomène de sélection d'un ou plusieurs variants lors du passage à la pathogénicité, et/ou au cours du traitement par les antifongiques (Jacobsen et al., 2008). Nous allons détailler maintenant un exemple de chacune de ces deux hypothèses.

2.2.2.2.2 Adaptation de *C. albicans* au cours de l'infection

Différentes études ont examiné les capacités d'adaptation de *C. albicans* en dehors du contexte de commensalisme. En effet, l'importante hétérogénéité génétique observée au sein même de la population de *C. albicans* au cours du commensalisme pourrait faciliter l'adaptation de cette levure aux modifications environnementales. En d'autres termes, au cours de l'infection chez l'hôte, des variants génétiques mieux adaptés pourraient émerger en raison d'un phénotype avantageux.

- Exemple du contexte pathologique retrouvé au cours de la mucoviscidose

Un premier exemple provient d'une étude réalisée chez des patients atteints de mucoviscidose (Kim et al., 2015). Chez ces patients, certaines conditions locales telles que la viscosité anormale des sécrétions respiratoires, les lésions épithéliales des bronches d'origine infectieuse, dues notamment à *Pseudomonas aeruginosa*, vont favoriser la colonisation de l'arbre respiratoire par divers champignons, dont *C. albicans*. De façon intéressante, certaines molécules sécrétées par les pathogènes bactériens communément associés à la mucoviscidose peuvent moduler, chez *C. albicans*, le passage de la forme levure à la forme filamenteuse. Par exemple, *P. aeruginosa* a la capacité d'inhiber la filamentation de *C. albicans* en sécrétant des molécules telles que la pyocyanine et la 3-oxo-C12 homosérine lactone, et sélectivement détruire les filaments de *C. albicans* en formant des biofilms sur leurs surfaces (Kim et al., 2015). *Burkholderia cenocepacia* est également capable d'inhiber la filamentation de *C. albicans* en sécrétant de l'acide cis-2-dodécénoïque tandis que *Staphylococcus aureus* a la possibilité de former un biofilm mixte avec *C. albicans* ayant un impact défavorable sur l'évolution de la maladie dans un modèle de co-infection chez la souris (Kim et al., 2015). Ainsi, les voies respiratoires des patients atteints de mucoviscidose constituent un environnement intéressant mais complexe pour l'étude de l'adaptation de *C. albicans*.

Les auteurs ont évalué la dynamique de la communauté fongique (séquençage à haut débit de la région ITS1 de l'ARNr) sur 89 échantillons d'expectoration provenant de 28 malades, sur une période allant de 3 à 56 mois. Ils y ont associé une analyse phénotypique et génotypique de 1 603 isolats fongiques obtenus en culture.

Dans cette étude, *C. albicans* était l'espèce la plus fréquemment retrouvée, avec un total de 1 056 isolats provenant de 23 patients. L'analyse des caractéristiques phénotypiques de ces isolats a montré de façon surprenante que 25 d'entre eux, provenant de 6 patients différents, présentaient au cours du temps une croissance filamenteuse constitutive, indépendante de stimuli environnementaux.

Afin d'identifier le génotype responsable de ce phénotype, les auteurs ont sélectionné 3 souches d'un même patient : une souche ayant le phénotype de filamentation constitutive (F1) et 2 souches se développant sous forme levure (Y1 et Y2). La comparaison des 3 génomes a permis d'identifier une mutation homozygote non-synonyme dans le gène *NRG1* de la souche F1. Ce gène code pour un facteur de transcription connu pour réprimer la filamentation. L'implication de la mutation retrouvée (D271N) sur la survenue du phénotype a été validée par mutagenèse. Le gène *NRG1* des 24 autres isolats présentant le même phénotype a été séquencé, et pour 23/24 d'entre eux des mutations homozygotes pouvant expliquer le phénotype ont été identifiées (Kim et al., 2015). L'acquisition de la résistance aux effets de répression de la filamentation, induite par les pathogènes bactériens présents dans le microbiote respiratoire de ces patients, montre ainsi un exemple de mécanisme commun d'adaptation des souches de *C. albicans* émergeant dans le cadre de l'interaction chez l'hôte.

- Exemple de l'adaptation génomique de *C. albicans* au cours du traitement par fluconazole

Un deuxième exemple d'évolution génomique de *C. albicans* au cours de l'infection chez l'Homme peut être illustré par l'adaptation de cette levure au cours de l'exposition *in vivo* à un traitement par antifongique. Pour ce faire, Ford *et al.* ont étudié des isolats cliniques de *C. albicans* provenant de patients infectés par le VIH, traités par fluconazole dans le cadre d'une candidose oropharyngée (Ford et al., 2015).

Durant cette étude, 43 isolats, prélevés séquentiellement chez 11 individus sur une période de plusieurs mois, ont été analysés génotypiquement (séquençage à haut débit) et phénotypiquement. Les premiers isolats étaient généralement plus sensibles aux antifongiques que les isolats recueillis au cours du traitement.

Comme l'isolat initial de chaque patient a été prélevé avant le début du traitement, celui-ci a donc constitué la base génétique pour l'analyse de l'apparition de variants au cours du temps (Ford et al., 2015).

Dans la plupart des séries, l'analyse comparative des génomes a suggéré que les infections étaient dues à une souche unique qui restait génétiquement reliée à celle isolée dans les derniers échantillons. Cependant, de nombreuses variations génomiques ont été détectées entre les isolats séquentiels d'un même patient, comprenant des mutations nucléotidiques (SNPs), mais également des réarrangements chromosomiques, indiquant que l'accroissement de la résistance aux antifongiques est lié à une pression de sélection (Ford et al., 2015).

Parmi les réarrangements chromosomiques majeurs, les LOH étaient significativement associés à une résistance plus élevée au fluconazole, alors que les aneuploïdies étaient retrouvées de façon transitoire et n'avaient pas d'effet sur le niveau de résistance. Les événements de LOH étaient le plus souvent persistants ou récurrents et ont été détectés dans toutes les séries de patients. Des événements de LOH se sont produits de façon fréquente sur le bras droit du Chr3 et sur le bras gauche du Chr5. Ces régions comprennent les gènes clés impliqués dans la résistance aux antifongiques (*CDR1*, *CDR2* et *MRR1* sur le Chr3 et *ERG11* et *TAC1* sur le Chr5). La mesure du fitness *in vitro* de ces souches a permis de démontrer le bénéfice de ces variants dans l'adaptation au fluconazole. En effet, la souche la plus évoluée avait un meilleur fitness en présence de fluconazole qu'en l'absence de celui-ci (Ford et al., 2015).

L'analyse précise de la survenue de SNPs au cours du temps a également permis d'identifier des gènes soumis à pression de sélection au cours du traitement par le fluconazole (SNPs absents dans la souche initiale et fixés dans les suivantes). Pour ce faire, seuls les SNPs non-synonymes retrouvés dans des régions codantes, chez au moins trois patients différents, ont été considérés.

De telles mutations ont été observées dans un total de 240 gènes, codant pour un grand nombre d'entre eux pour des protéines impliquées dans la synthèse ou la constitution de la paroi fongique (telles que des protéines Hyr/Iff (HYR1 et 3, IFF8 et 6), des adhésines Als (Als1-4, 7, 9) et les protéines de type Pga30-like (7 gènes). Les mutations de ces gènes pourraient permettre à *C. albicans* de mieux échapper au système immunitaire de l'hôte.

D'autres gènes fréquemment mutés chez plusieurs patients ont été identifiés, tels que les gènes *CHO1*, *MNN2* et 7 membres de la famille de gènes *FGR*, impliqués dans la croissance du filament, ainsi que le gène *AXL1* qui code pour une endoprotéase putative et des gènes impliqués dans la résistance aux antifongiques comme *ERG11* (Ford et al., 2015).

Cette étude a permis de fournir une description globale des événements de remaniement génomique et donc des mécanismes adaptatifs mis en place par *C. albicans in vivo* au cours d'un traitement par le fluconazole. Elle a permis également d'identifier de nouveaux gènes potentiellement importants pour la résistance aux antifongiques et l'échappement au système immunitaire de l'hôte. Cependant, une limite importante de cette étude consiste en l'analyse d'une seule colonie pour chaque prélèvement. En effet, la comparaison des génomes des isolats issus d'un même patient a montré plusieurs milliers de SNPs de différence. Il n'est donc pas possible dans cette étude de faire la distinction entre l'apparition de nouvelles mutations et la sélection, par pression antifongique, de variants minoritaires préexistants.

La diversité génétique basale au sein d'un échantillon humain à l'échelle du génome est une information qui n'est pas connue. L'analyse de plusieurs colonies recueillies dans un même prélèvement chez un même individu aurait le potentiel de révéler la structure, la dynamique et la diversité de la population de *C. albicans*.

2.2.3 La Candidose cutané-Muqueuse Chronique (CCMC) un exemple d'interaction pathologique de très longue durée entre *C. albicans* et l'hôte

2.2.3.1 La CCMC : un modèle d'infection à long terme

La Candidose Cutané-Muqueuse Chronique (CCMC) est une maladie invalidante qui se caractérise par des infections sévères, persistantes ou récurrentes, de la peau, des ongles et des muqueuses par des levures du genre *Candida*, principalement *C. albicans* (Kirkpatrick, 2001).

La CCMC est fréquente chez des individus qui présentent un défaut acquis ou inné du nombre et/ou de la fonction des lymphocytes T (patients infectés par le VIH, sous traitement par des immunosuppresseurs, patients présentant des immunodéficiences combinées sévères (SCID), lymphopénie CD4 idiopathique...). Cependant, chez ces individus, la CCMC est fréquemment associée à d'autres manifestations cliniques, infectieuses ou auto-immunes potentiellement plus graves (Puel et al., 2012).

Il existe, en revanche, des patients présentant une CCMC comme principale manifestation clinique, qui débute en général très tôt dans l'enfance (Li et al., 2017; Puel et al., 2012). Une prédisposition génétique de ces formes de CCMC avait été proposée dans les années 1970 (Wells et al., 1972), mais ce n'est qu'en 2011 qu'ont été identifiées les premières étiologies (Puel et al., 2011). Ces déficits immunitaires primitifs incluent de nombreuses causes génétiques comme par exemple des mutations hétérozygotes dans le gène *STAT3* responsables du syndrome hyper-IgE (AD-HIES), des mutations homozygotes dans le gène *AIRE*, induisant une polyendocrinopathie auto-immune de type 1 (APS-1), des déficits autosomiques dominants (AD) en *CARD9*, *IL-12Rb1*, *IL-12p40* et *RORγ/γT*. Dans cette catégorie, on retrouve également les mutations hétérozygotes gain-de-fonction (GOF) de *STAT1* (Liu et al., 2011; van de Veerdonk et al., 2011).

On parle enfin de CCMC isolée (CCMCi), lorsque celle-ci constitue l'unique manifestation clinique résultant des déficits autosomiques récessifs (AR) en *IL-17RA*, *IL-17RC* et *ACT1* et le déficit AD en *IL-17F* (Boisson et al., 2013; Ling et al., 2015; Puel et al., 2011). Les principaux déficits et leurs caractéristiques retrouvés dans les CCMC dites isolées ou syndromiques sont résumés dans le **Tableau 2**.

Disease	Frequency of CMC	Other infections	Associated symptoms	Immunological phenotype	Gene	Inheritance
Syndromic CMC						
HIES	85%	<i>Staphylococcus</i> , <i>Aspergillus</i>	Eczema, scoliosis, pneumatocele, hyperextensibility, dysmorphic facial features, retention of primary teeth	Increased serum IgE, eosinophilia, decreased IL-17-producing T cells	<i>STAT3</i>	AD
APECED	70–98%		Ectodermal dysplasia, autoimmune dysfunction of parathyroid and adrenal glands, alopecia	Neutralizing antibodies against IL-17A, IL-17F and/or IL-22	<i>AIRE</i>	AR
CARD9 deficiency	35–86%		Dermatophytes, <i>Candida</i> , brain abscess	Decreased IL-17-producing T cells, impairment of <i>C. albicans</i> -killing by neutrophils	<i>CARD9</i>	AR
IL-12Rβ1 and IL-12p40 deficiency	6–25%	<i>Mycobacterium</i> , <i>Salmonella</i>		Decreased IL-17-producing T cells, impaired IL-12 signaling	<i>IL12RB1</i> <i>IL12B</i>	AR
STAT1 gain-of-function	98%	Bacteria, viruses, fungi, mycobacteria	Aneurysm, autoimmune diseases, endocrine diseases	Decreased IL-17-producing T cells, decreased switched memory B cells	<i>STAT1</i>	AD
RORγT deficiency	6/7 (86%)	<i>Mycobacterium</i>	Lack of peripheral lymph node, thymic hypoplasia	Defect of MAIT, type 1 NKT, IL-17-producing T cells, impaired antigen-specific IFN-γ production	<i>RORC</i>	AR
CMCD						
IL-17RA deficiency	3/3 (100%)	<i>Staphylococcus</i>		No response to IL-17A, IL-17E and IL-17F	<i>IL17RA</i>	AR
IL-17RC deficiency	3/3 (100%)			No response to IL-17A and IL-17F	<i>IL-17RC</i>	AR
IL-17F deficiency	5/7 (70%)			Impaired IL-17F and IL-17A/F function	<i>IL17F</i>	AD
ACT1 deficiency	2/2 (100%)	<i>Staphylococcus</i>		No response to IL-17A, IL-17E, and IL-17F	<i>TRAF3IP</i>	AR

Abbreviations: AD, autosomal-dominant; APECED, autoimmune polyendocrinopathy–candidiasis–ectodermal dystrophy; AR, autosomal-recessive; CARD9, caspase recruitment domain-containing protein 9; CMC, chronic mucocutaneous candidiasis; CMCD, CMC disease; HIES, hyper IgE syndrome; IFN-γ, interferon gamma; IL, interleukin; RORγT, retinoic acid-related orphan receptor γT

Tableau 2 : Déficiences immunitaires primitives impliquées dans l’immunité innée IL-17 chez les patients présentant une CCMC « isolée » ou « syndromique » (d’après Li et al., 2017).

Ces déficiences ont mis en évidence le rôle majeur chez l’homme de l’IL-17 dans l’immunité cutanéomuqueuse vis-à-vis de *C. albicans* (**Figure 14**) (Li et al., 2017; Puel et al., 2011). En effet, la reconnaissance de *Candida* par les cellules épithéliales induit la production de cytokines pro-inflammatoires. Les cytokines, IL-6 et IL-23 vont activer les lymphocytes T via STAT3, qui induit à la fois la production du facteur de transcription ROR et la différenciation des lymphocytes T producteurs de l’IL-17. Les lymphocytes T différenciés (localisés au niveau de la peau et des muqueuses) vont ainsi sécréter de l’IL-21, IL-22 et de l’IL-17A et IL-17F (**Figure 14**).

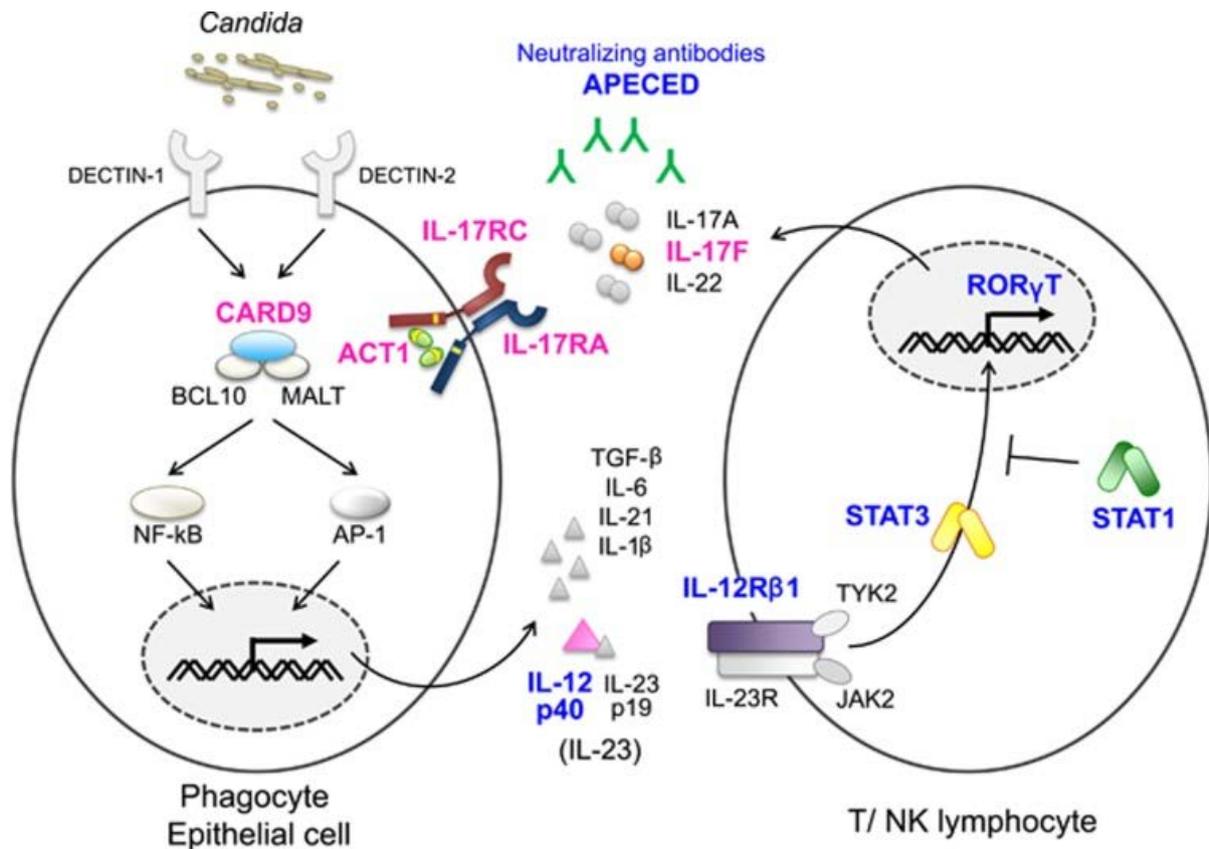


Figure 14 : Défaut de l'immunité innée IL17 prédisposant aux CCMC. Représentation schématique de l'immunité anti *Candida* chez l'homme. La reconnaissance de *Candida* par les phagocytes induit la production de cytokines pro-inflammatoires. Les cytokines, IL-6 et IL-23 activent les lymphocytes T via STAT3 qui, à la fois, induit la production du facteur de transcription ROR γ T et la différenciation des lymphocytes T producteurs de l'IL-17, ce qui conduit à la production d'IL-17A, IL-17F et IL-22. Les déficits dans la voie de signalisation STAT3 (IL-12R β , IL-12p40) ou AR-ROR γ T et AD-STAT1 (GOF) sont responsables de CCMC syndromique (bleu). Des déficits dans les quatre gènes (codant l'IL-17F, l'IL-17RA, l'IL-17RC et l'ACT1) directement impliqués dans la signalisation de l'IL-17 sont responsables de CCMC isolée (rose) (Okada et al., 2016).

Dans la CCMC, les malades sont porteurs très tôt pendant l'enfance (âge médian 1 an chez les patients STAT1) d'une souche généralement unique de *C. albicans* (Toubiana et al., 2016). Différentes études par typage MLST réalisées chez un faible nombre de patients ont montré qu'une même souche, ou des souches ayant subi des micro-variations, pouvaient persister chez le malade durant plusieurs années (jusqu'à 9 ans). Une même souche peut être responsable d'infections récurrentes malgré un traitement antifongique (McManus et al., 2011; Moorhouse et al., 2016; Siikala et al., 2010).

En effet le déficit immunitaire dont souffrent ces patients induit, en plus de la susceptibilité particulière à *Candida*, un défaut d'élimination du pathogène au niveau des muqueuses (Trautwein-Weidner et al., 2015). Ceci a été montré via un modèle murin de candidose oropharyngée (COP), utilisant des souris présentant différents déficits dans la voie de l'IL-17, qui sont dans l'incapacité d'éliminer la levure au niveau de l'épithélium. Les déficits dans la voie de l'IL-17 sont responsables d'un défaut d'induction des peptides antimicrobiens (AMP) au niveau de l'épithélium, entraînant une colonisation fongique persistante (Trautwein-Weidner et al., 2015). Ainsi, chaque souche peut évoluer chez un même patient durant des années, voire toute sa vie.

La CCMC constitue donc un modèle particulier d'interaction pathologique de longue durée entre *C. albicans* et l'hôte. Dans cette pathologie *C. albicans* évolue durant plusieurs années dans un environnement de l'hôte modifié par une réponse immunitaire et inflammatoire pathologique.

2.2.3.2 La CCMC : un contexte d'interaction pathologique et inflammatoire particulier entre *C. albicans* et les cellules épithéliales.

Pour mieux comprendre l'environnement spécifique dans lequel évolue *C. albicans* au cours de la CCMC, il est intéressant d'aborder en détail les étapes qui constituent l'interaction entre *C. albicans* et les cellules épithéliales de l'hôte au niveau de la bouche et de la peau, mais également la réponse immunitaire et inflammatoire qui en résultent.

Les cellules épithéliales situées à la surface des muqueuses et de la peau constituent la première ligne de défense contre *C. albicans*. L'adhésion de *C. albicans* au niveau de ces cellules résulte de l'interaction entre les constituants de la paroi fongique et des récepteurs et protéines de surface de l'hôte (**Tableau 3**) (Naglik et al., 2017). La reconnaissance de *C. albicans* par les cellules épithéliales orales, notamment la lignée cellulaire TR146, induit en 15 minutes trois voies majeures de signalisation : la voie du facteur nucléaire-kappaB (NF- κ B), la phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) et la voie Mitogen Activated Protein (MAP) kinase (p38, JNK et ERK1/2) (**Figure 15**) (Naglik et al., 2017).

C. albicans peut initialement être reconnu via ses mannanes et β -glucanes (paroi fongique) mais il a été montré que cela n'active pas complètement les cellules épithéliales, soulignant le rôle limité de ces polysaccharides dans l'induction de la réponse immunitaire épithéliale/kératinocytes (Moyes et al., 2010). En effet contrairement aux cellules myéloïdes, l'activation des cellules épithéliales n'est pas médiée par des récepteurs lectines de type-C (CLR) ou des récepteurs de type « Toll-Like Receptors » (TLR) qui reconnaissent les β -glucanes et mannanes (Moyes et al., 2010). De nombreuses autres protéines sécrétées ou retrouvées à la surface cellulaire de *C. albicans* (par ex. Sap1-3/9/10, Als1/3/4/9, Mp65, Phr1, Iff4, Sun41, Pra1, Eap1, Utr2 et Ecm33...) favorisent l'adhésion épithéliale ; il est donc possible que l'interaction fasse également intervenir l'une de ces protéines (Naglik et al., 2017).

Bien que l'adhésion et l'invasion de *C. albicans* induisent l'activation des différentes voies de signalisation, c'est la formation de l'hyphe (filament) de *C. albicans* qui provoque à la fois des lésions de l'épithélium et l'activation de l'immunité innée. Cela se fait, entre autres, via la sécrétion d'une toxine appelée candidalysine (codée par le gène *ECE1*), qui au niveau des cellules épithéliales orales, induit un afflux d'ions calcium et une libération de lactate déshydrogénase (LDH) (**Figure 15**) (Moyes et al., 2016). Il a en effet été montré que des mutants de *C. albicans* pour le gène *ECE1* conservent leur capacité d'invasion *in vitro* mais sont incapables d'induire une lyse tissulaire ou la libération de cytokines. Ces mutants ont donc un pouvoir pathogène diminué dans un modèle murin de candidose oropharyngée (Moyes et al., 2016).

Fungal component/gene	Epithelial function or target receptors
Structural polysaccharides	
β -Glucan	Induces epithelial signaling.
Mannans	Induces epithelial signaling. Receptors not identified.
Chitin	Induces epithelial signaling. Receptors not identified.
Adhesins	
<i>HWP1</i>	Adhesion to epithelial cells via transglutaminase activity. Specific host receptors unknown.
<i>ALS1-9</i>	Adhesin family. Structural studies indicate this family has multiple epithelial targets.
<i>INT1</i>	Interaction with epithelial integrins.
Toxins	
<i>ECE1</i>	Parent protein of Candidalysin. Induces c-Fos and MKP1 signaling. Receptor activation indicated but not identified.
Endocytosis	
<i>ALS3</i>	Activation of or interaction with E-cadherin, EGFR/Her2, AhR, NEDD9 and PDGF BB
<i>SSA1</i>	HSP70 family member. Activation of or interaction with EGFR/Her2
Active penetration/hydrolysis	
<i>SAP1-8</i>	Secreted aspartic proteases — digestion of epithelial tissues. Sap5 degrades E-cadherin
<i>PLB1</i>	Phospholipase B1 — digestion of epithelial tissues
<i>LIP1-10</i>	Lipase family — digestion of epithelial tissues

Tableau 3 : Principales structures et principaux gènes de *C. albicans* codant pour des protéines impliquées dans l'interaction avec les cellules épithéliales de l'hôte (d'après Naglik et al., 2017).

C'est donc la candidalysine qui, au niveau épithélial, active le système immunitaire, principalement par la voie de signalisation MAPK, en particulier la voie p38 (activation de c-Fos) et la voie ERK1/2 (activation de MKP1) qui régule la réponse immunitaire épithéliale. L'activation de ces deux voies a pour conséquence la production de cytokines pro-inflammatoires, chimiokines et peptides antimicrobiens (IL-1 α/β , IL-6, IL-8, GM-CSF et G-CSF, β -defensine 3, CCL20 et S100A8/9) nécessaires au recrutement des cellules immunitaires innées dont les neutrophiles et les macrophages (**Figure 15**) (Moyes et al., 2016). Ces cellules reconnaissent les mannanes de la paroi fongique ainsi que l'ADN via les TLR2/4 et 9, et le β -glucane via les CLRs (Dectin-1/-2, DC-SIGN ou Mincle) (Naglik et al., 2017, 2014). L'activation de ces récepteurs induit les voies de signalisation NF- κ B, MAPK et Syk et la production à nouveau de cytokines pro-inflammatoires (**Figure 15**).

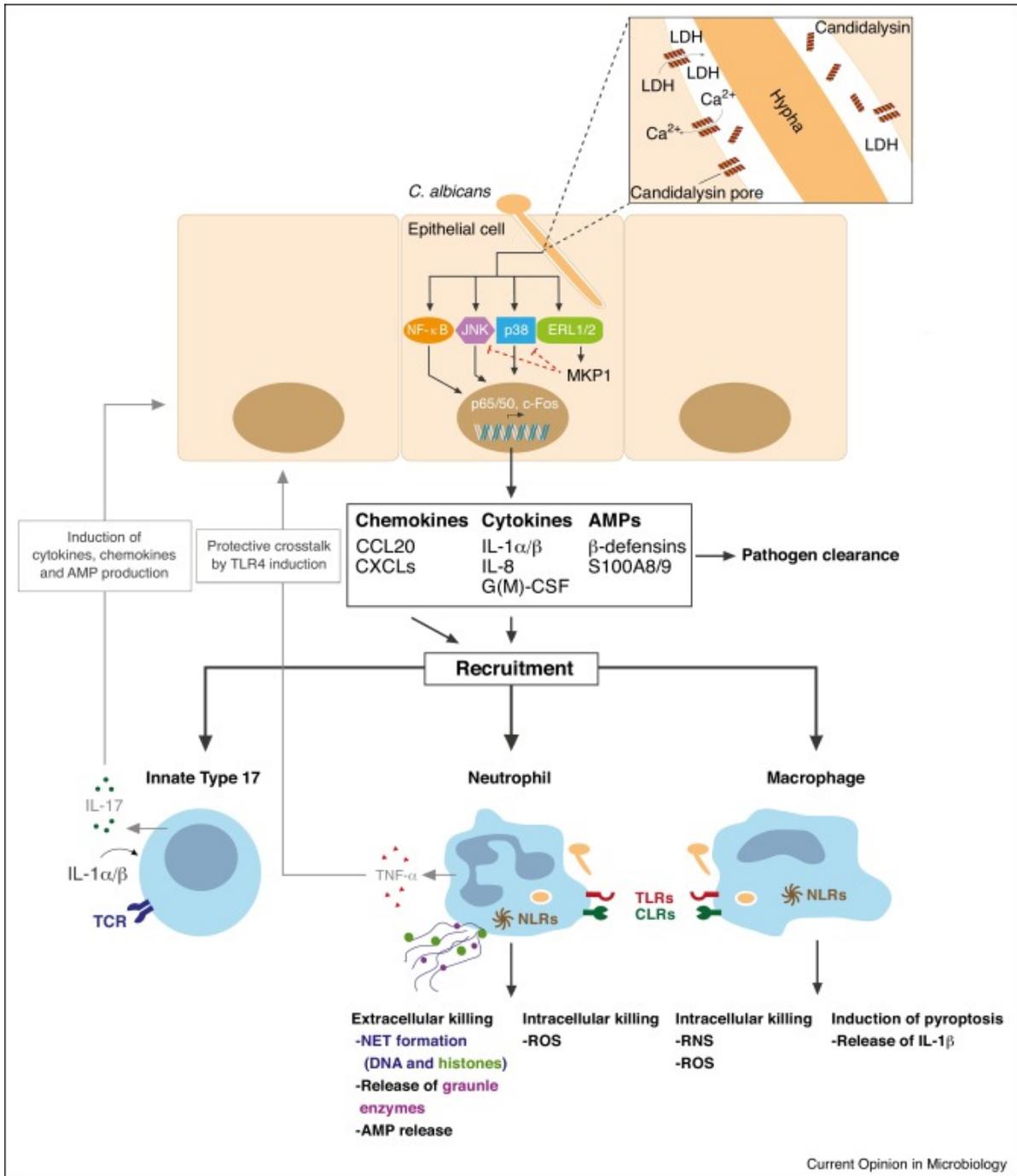


Figure 15 : Immunité innée anti *C. albicans* au niveau de la muqueuse buccale (d'après Naglik et al., 2017).

Les neutrophiles, rapidement recrutés sur le site d'infection, peuvent également être activés et libérer du TNF α (action protectrice sur les cellules épithéliales), sans contact direct avec la levure via les chimiokines et facteurs de croissance sécrétés par l'épithélium. Les neutrophiles phagocytent (par l'intermédiaire des CLR) et tuent les cellules et hyphes intracellulaires de *C. albicans* essentiellement par le biais de mécanismes oxydatifs.

Les hyphes qui ne peuvent être phagocytés, de par leur taille trop importante, sont soit inhibés, soit lysés de façon extracellulaire, via la formation de NETs (Neutrophil Extracellular Traps) par dégranulation d'enzymes et sécrétion de peptides antimicrobiens tels que la calprotectine (**Figure 15**) (Kenny et al., 2017).

Les macrophages, quant à eux, jouent un rôle plus mineur *in vivo* que les neutrophiles. Ils sont recrutés sur le site d'infection et phagocytent les levures non opsonisées via la reconnaissance par les TLRs et les CLR. Cependant, la lyse des levures dans le phagolysosome par des mécanismes oxydatifs et nitrosatifs (ROS/RNS) est moins efficace qu'avec les neutrophiles. Ainsi, *C. albicans* peut facilement survivre et s'échapper des macrophages *in vitro* (Naglik et al., 2017). Les macrophages peuvent également détecter *C. albicans* lorsque celui-ci a été internalisé par l'intermédiaire de récepteurs NLRs. Ces derniers peuvent activer l'inflammasome NLRP3, responsable de la production de cytokines pro-inflammatoires (IL-1 β et IL-18...) et entraîner la mort des cellules hôtes (**Figure 15**). En effet, la formation intracellulaire d'hyphes, induite par l'alcalinisation du phagosome, provoque la mort des macrophages soit par pyroptose, soit par perforation de la membrane du macrophage (Naglik et al., 2017).

C'est à ce stade qu'interviennent les cellules de l'immunité innée de type 17 dont nous avons détaillé les étapes dans la partie précédente.

De nombreuses cellules d'origine lymphoïde sont capables de produire de l'IL-17 mais dans le contexte de la candidose oropharyngée, ce sont les lymphocytes $\gamma\delta$ -T et les TCR β + « naturels » Th17 (nTH17) qui sont principalement concernés. Ce sont donc les cytokines (IL-1 α/β , IL-6 ; IL-23...) libérées par les différentes cellules (épithéliales, macrophages...) qui induisent la différenciation des lymphocytes Th17 qui, à leur tour, activent de nouveaux lymphocytes Th17. La sécrétion de l'IL-17 induit à nouveau la libération de chimiokines, de cytokines et de peptides antimicrobiens (AMP) à partir des cellules épithéliales, afin d'amplifier l'efficacité d'élimination de la levure au niveau buccal (**Figure 15**) (Naglik et al., 2017).

En effet, les fonctions antifongiques de l'IL-17 sont multiples. Premièrement, l'IL-17 est un puissant activateur de la réponse neutrophilique via l'induction de la sécrétion, par les cellules épithéliales, de cytokines et chimiokines tel que le G-CSF, CXCL1/2 et 5. Deuxièmement, l'IL-17 est également un puissant inducteur de la production de peptides antimicrobiens (AMPs), en particulier des β -défensines-1 et 3, mais également des histatines via une action sur les cellules des glandes salivaires (Naglik et al., 2017).

Comme nous l'avons mentionné précédemment, les diverses anomalies récemment identifiées dans la voie de signalisation de l'IL-17 sont responsables d'une susceptibilité accrue aux candidoses oropharyngées responsables de l'entité clinique nommée la Candidose Cutané-Muqueuse Chronique. Les patients présentant un déficit immunitaire primitif dans cette voie conservent leurs capacités de réponse inflammatoire via les neutrophiles et les macrophages et ne développent généralement pas d'infections systémiques ou disséminées dues à *C. albicans*. En effet, seuls 10% des patients présentant des mutations gain de fonction dans *STAT1* développeront des formes invasives (Toubiana et al., 2016). Chez les autres patients, les infections sont superficielles et siègent au niveau de la peau, des ongles et des muqueuses orales (**Figure 16**). Les infections sont persistantes et récurrentes, car les patients sont dans l'incapacité de fournir une réponse immunitaire complète pour éliminer efficacement la levure. Ces infections deviennent extrêmement invalidantes et nécessitent une prise en charge particulière essentiellement par l'utilisation d'antifongiques que nous allons voir maintenant.



Figure 16 : Photographies de patients présentant des mutations STAT1-GOF. (A) Muguet buccal (B-C) Onychomycose (D-I) Candidose cutanée (d'après Toubiana et al., 2016)

2.2.3.3 La CCMC : un contexte d'utilisation à long terme de molécules antifongiques

La plupart des patients atteints de CCMC sont traités par des antifongiques en topique et/ou systémique (Boisson et al., 2013; Ling et al., 2015; Liu et al., 2011; Puel et al., 2011; Toubiana et al., 2016; van de Veerdonk et al., 2011). Malheureusement, peu de molécules antifongiques sont actuellement commercialisées.

Les antifongiques appartiennent à différentes classes comprenant : les polyènes, les azolés, les échinocandines, les analogues nucléosidiques et les allylamines.

Ces molécules ont une efficacité variable selon le type et le site de l'infection mais également en fonction de la sensibilité de l'isolat (Pappas et al., 2016; Pfaller et al., 2013, 2010; Pfaller and Diekema, 2012). La proportion d'utilisation des différentes molécules est résumée dans le **Tableau 4** provenant d'une étude récente réalisée par Toubiana *et al.* sur une cohorte de 274 patients atteints de STAT1-GOF (Toubiana et al., 2016).

Chez les patients atteints de CCMC, ce sont les antifongiques azolés (triazolés), et notamment le fluconazole *per os* qui est la molécule utilisée en première intention, suivi de l'itraconazole et plus récemment du posaconazole ou du voriconazole (molécules azolées de deuxième génération). Une dernière molécule triazolée est maintenant disponible : c'est l'isavuconazole. Ces molécules ont succédé aux imidazolés qui étaient les premiers antifongiques azolés à être utilisés. Cependant, du fait de leur mauvaise absorption digestive et de leur hépato-toxicité, ces molécules sont maintenant restreintes à la voie locale pour le traitement de formes superficielles mineures. Les triazolés sont quant à eux largement utilisés en raison de leur bonne tolérance, et de leur facilité d'administration (bonne biodisponibilité par voie orale). Cependant, elles présentent des interactions médicamenteuses importantes.

Les azolés inhibent de façon sélective la 14- α -déméthylase, enzyme codée par le gène *ERG11*. La 14- α -déméthylase est impliquée dans la biosynthèse de l'ergostérol, un constituant essentiel de la membrane fongique. Cependant, l'activité fongistatique des azolés est un inconvénient majeur pour traiter ces patients et joue un rôle dans l'émergence de la résistance. Ainsi, seul un tiers des patients CCMC présentant des mutations GOF-STAT1 sont traités avec succès par des azolés, tandis qu'une réponse partielle est observée pour les autres (Toubiana et al., 2016). Par ailleurs, les traitements curatifs sont de longue durée et suivis d'une prophylaxie afin de prévenir les récurrences (Boisson et al., 2013; Ling et al., 2015; Liu et al., 2011, 2011; van de Veerdonk et al., 2011). Ces traitements prolongés par des antifongiques azolés sont la principale cause d'émergence de résistance (McManus et al., 2011; Moorhouse et al., 2016; Siikala et al., 2010).

Chez les patients intolérants aux antifongiques azolés ou présentant des souches résistantes à ces antifongiques, les échinocandines (caspofongine et micafungine) peuvent être utilisées (**Tableau 4**). Leur action cible l'enzyme β -1,3-glucane synthase, codée par le gène *FKS1* impliqué dans la synthèse du β -D-glucane, composant majeur de la paroi fongique.

Ces molécules utilisables uniquement par voie IV (Intraveineuse), présentent une activité fongicide et sont plutôt indiquées en 1^{ère} ligne dans le traitement des candidoses invasives. Bien que ces molécules aient été introduites relativement récemment (il y a environ 15 ans), une augmentation de la prévalence des isolats cliniques de différentes espèces de *Candida* résistants aux échinocandines a été observée dans diverses études (Chandraskar and Sobel, 2006; Deresinski and Stevens, 2003; Pfaller and Diekema, 2004).

Treatment	Patients (%)
	n = 274
No antifungal treatment	19 (7)
Intermittent antifungal treatment	53 (19)
Current long-term antifungal treatment	202 (74)
Local treatment only	8 (3)
Oral fluconazole	150 (55)
Oral posaconazole/itraconazole	53 (19)
Oral voriconazole	19 (7)
IV echinocandins	6 (2)
Oral terbinafine	3 (1)
IV amphotericin B	7 (3)
Antibiotic prophylaxis	66 (24)
Co-trimoxazole	41 (15)
Macrolides	20 (7)
Others*	12 (4)
Antiviral prophylaxis	4 (1)
Polyvalent immunoglobulins	37 (14)
Immunotherapy†	8 (3)
Immunosuppressive therapies‡	6 (2)
Hematopoietic stem cell transplantation (HSCT)	5 (2)

*Nebulized colimycin, topical fucidic acid, fluoroquinolone, tetracycline, amoxicillin.

†Granulocyte–colony stimulating factor (G-CSF)/granulocyte macrophage–colony stimulating factor (GM-CSF), interferon (IFN) α/γ , rituximab.

‡Cyclosporine, azathioprine, corticoids, or mycophenolate mofetil.

Tableau 4 : Prise en charge thérapeutique des patients ayant des mutations *STAT1*-GOF (d'après Toubiana et al., 2016)

Malheureusement, tous ces traitements ne sont que symptomatiques. Seule la transplantation de cellules souches hématopoïétiques pourrait être considérée comme une réelle option thérapeutique pour ces patients, en particulier pour ceux présentant les formes cliniques les plus graves (infections invasives, anévrismes cérébraux et développement de cancers) que l'on retrouve chez les patients GOF-STAT1 (Toubiana et al., 2016).

En attendant, l'utilisation d'antifongiques reste, pour ces patients, le traitement de choix. Mais leur utilisation devient de plus en plus complexe en raison du nombre restreint de molécules disponibles et de la survenue fréquente de résistance (McManus et al., 2011; Moorhouse et al., 2016; Siikala et al., 2010; Toubiana et al., 2016).

2.2.3.4 La CCMC : problématique de la résistance aux antifongiques

Chez les patients atteints de CCMC, l'utilisation prolongée des molécules antifongiques de type azolés est fréquente et pose de nombreux problèmes en termes d'apparition de résistance (McManus et al., 2011; Moorhouse et al., 2016; Siikala et al., 2010; Toubiana et al., 2016). Dans l'étude réalisée par Toubiana *et al.* en 2016 sur 274 patients STAT1-GOF, 255 (93 %) ont nécessité un traitement antifongique parmi lesquels 202 (79%) ont eu besoin d'un traitement à long terme (Toubiana et al., 2016). Une résistance clinique à au moins un antifongique a été observée chez 39 % (78/202) des patients soumis à un traitement à long terme, et chez 15% (8/53) des patients traités par intermittence. Parmi les patients traités, 10% (25/255) ont présenté une résistance clinique à plusieurs molécules antifongiques. Les résultats relatifs à la sensibilité des souches étaient disponibles pour 40 % des patients en échec thérapeutique et a révélé un pourcentage élevé de 98 % de résistance aux azolés (Toubiana et al., 2016). La plupart de ces patients a eu besoin de traitements de deuxième ou de troisième ligne, tels que du voriconazole, des échinocandines, de la terbinafine ou de l'amphotéricine B liposomale (Toubiana et al., 2016).

De façon générale, l'incidence de la résistance aux antifongiques chez *C. albicans* est faible mais dépend du type d'infection. L'incidence de la résistance aux azolés chez les isolats de *C. albicans* provenant de patients candidémiques est beaucoup plus faible (0 à 5 %) que celle observée pour les patients présentant une CCMC (Cleveland et al., 2015; Pfaller et al., 2015; Pfaller and Diekema, 2012).

Les isolats de *C. albicans* provenant des patients CCMC constituent donc une source majeure d'information pour l'étude approfondie des mécanismes génomiques de résistance aux antifongiques. Comme nous l'avons vu précédemment, un phénotype de résistance peut survenir à la suite de diverses altérations génomiques (SNPs/LOH/aneuploïdies) en réponse au stress induit par les antifongiques. L'étude de la survenue de ces événements chez les patients atteints de CCMC est un autre volet intéressant de l'étude des mécanismes adaptatifs mis en place par *C. albicans* pour persister chez l'Homme dans un environnement défavorable.

Objectifs du travail de thèse

Au cours de l'interaction avec l'hôte, *C. albicans* est confronté à un grand nombre de stress environnementaux et immunitaires imposant des capacités d'adaptations rapides pour survivre à ces environnements changeants. Comme nous l'avons vu dans la partie introductive, de nombreux événements de remaniements génomiques, incluant des aneuploïdies ou des pertes d'hétérozygotie (LOH), ont été observés au sein de diverses souches de *C. albicans*. La survenue de ces événements a largement été étudiée *in vitro* et *in vivo*, essentiellement chez la souris, en réponse à diverses modifications environnementales. En revanche, hormis les études réalisées par des approches de typage moléculaire, très peu de travaux se sont portés sur la diversité génétique de *C. albicans* et les modifications du génome survenant de façon naturelle au cours de l'interaction chez l'Homme, que ce soit durant le commensalisme ou au cours de l'infection.

Le but de ce travail était d'approfondir nos connaissances concernant l'étendue de la diversité génétique de *C. albicans* chez l'hôte naturel grâce à l'utilisation du séquençage de génomes entiers.

Après avoir travaillé sur la mise en place de méthodes d'analyses comparatives de génomes proches de *C. albicans*, nous nous sommes interrogés sur l'étendue de la diversité génétique qui pouvait exister entre des souches de *C. albicans* isolées au sein d'un même prélèvement chez des sujets « porteur sain », c'est-à-dire dans le cadre d'un simple portage commensal chez l'Homme en l'absence de pression de sélection qui pourrait être due à l'exposition à des antifongiques ou à un terrain particulier chez l'hôte.

Puis nous avons exploré si *C. albicans* était capable d'adaptation génomique spécifique dans un contexte d'interaction pathologique chez l'hôte, dans un « modèle » de patients atteints de CCMC. Dans ce modèle, *C. albicans* est confronté à un environnement modifié par une réponse immunitaire et inflammatoire inappropriée mais également par une exposition prolongée à des antifongiques.

Notre travail de thèse sera présenté sous la forme de quatre parties répondant aux objectifs suivants :

1. Développer un « pipeline » d'analyse des données de séquençage à haut débit pour la comparaison de génomes proches de *C. albicans* afin de détecter de façon exhaustive des variations génétiques fines entre divers isolats génotypiquement reliés.

2. Déterminer et quantifier le niveau basal d'hétérogénéité génétique au sein de la population de *C. albicans* au cours du commensalisme.

3. Etudier les mécanismes adaptatifs de *C. albicans* dans un contexte d'interaction pathologique chez l'Homme par l'analyse génomique et phénotype de séries chronologiques de souches de *C. albicans* isolées chez les patients atteints de CCMC :

- Caractériser génotypiquement les souches de *C. albicans* par typage moléculaire afin de déterminer leurs liens au cours du temps et au sein des différentes familles

- Caractériser phénotypiquement les souches issues de patients atteints de CCMC et de leur famille dans différentes conditions expérimentales *in vitro*, afin d'identifier les modifications phénotypiques des souches survenant au cours de l'interaction prolongée et pathologique avec l'Homme.

- Décrire de façon exhaustive les remaniements affectant le génome de *C. albicans* lors de l'interaction pathologique à long terme.

- Identifier des gènes dont l'altération surviendrait spécifiquement au cours de l'interaction prolongée et pathologique avec l'Homme.

4. Etudier les événements génétiques impliqués dans la survenue de la résistance *in vitro* aux antifongiques chez les souches de *C. albicans* isolées chez les patients atteints de CCMC.

RESULTATS-DISCUSSION

1 Analyse des données de séquençage de génomes proches de *C. albicans*

La première partie de ce travail a consisté à adapter à la problématique de comparaison de génomes apparentés la méthode d'analyse de génomes de *C. albicans* de souches épidémiologiquement non liées, utilisée dans le laboratoire. En effet, comme nous l'avons vu précédemment, la séquence génomique de *C. albicans* est constituée de 2 x 8 séquences linéaires (0,9 à 3,2 Mb) correspondant aux 8 paires de chromosomes avec une importante variabilité entre les séquences des chromosomes homologues liée à l'hétérozygotie (environ 70 000 SNPs par génome). Les modifications, même mineures, du génome peuvent avoir des conséquences phénotypiques importantes comme la résistance aux antifongiques. La détection précise et fiable ainsi que la possibilité de comparer la survenue de ces événements entre différents génomes proches a constitué une étape importante de ce projet.

Durant ce travail, j'ai ainsi pu participer au développement de l'outil bio-informatique mis en place dans l'unité afin d'avoir la possibilité de comparer des génomes proches de *C. albicans*. Le « pipeline » d'analyse inclut la détection et l'interprétation des polymorphismes nucléotidiques de haute qualité (SNPs), les variations du nombre de copies des gènes (CNVs) ainsi que les pertes d'hétérozygotie (LOH) majeurs (plusieurs kb) mais également de petites tailles (quelques pb).

1.1 Sélection des séquences de bonne qualité

La première étape de l'analyse consiste à détecter puis filtrer les séquences brutes « reads » de mauvaise qualité issues du séquenceur (format fastq). Les reads ou portions de reads sont supprimés ou coupés s'ils ne satisfont pas les critères qualité définis (détection FastQC v0.11.5 puis filtration par l'outil Cutadapt v1.8.3.).

1.2 Alignement des séquences

Les reads sont ensuite alignés sur une séquence de référence (ici Assemblage 22 du génome de référence SC5314, version A22-s06-m01-r01) par l'utilisation d'un logiciel spécifique de « mapping » BWA_mem générant un fichier de type SAM (Sequence Alignment/ Mapping) (Li et al., 2009).

Les logiciels SAMtools (SAMtools v1.2) et PicardTools (<http://broadinstitute.github.io/picard/>) sont utilisés pour compresser le fichier en un fichier binaire de type BAM (Binary sequence Alignment/Mapping) et organiser les séquences pour les étapes ultérieures (Li et al., 2009; Wysokar et al., n.d.).

L'outil « Genome Analysis Tool Kit » (GATK v3.1, (McKenna et al., 2010)) est ensuite utilisé pour la suite du prétraitement des données ; il s'articule en deux étapes. Tout d'abord un réalignement des séquences proches des Indels est effectué afin de diminuer la détection de SNPs erronés (diminution des faux positifs) au sein de ces zones grâce aux modules « RealignerTargetCreator » et « IndelRealigner ». Puis les alignements sont convertis en fichier BAM, puis en fichier texte « pileup » plus simple et contenant le minimum de données nécessaires à la suite de l'analyse. Les données brutes concernant la profondeur de séquençage à chaque position génomique sont conservées pour l'analyse des CNVs.

1.3 Analyse des SNPs/LOH

Les SNPs correspondent aux positions du génome qui ont deux allèles différents sur des chromosomes homologues ou une variation par rapport à la référence. La détection de ces positions (Variant calling) peut se faire avec l'outil GATK via le module « HaplotypeCaller». Les SNPs de mauvaise qualité sont filtrés à l'aide du module « VariantFiltration » de GATK en utilisant les recommandations de bonnes pratiques de filtration, comprenant différents paramètres (QD < 2.0, MQ < 40.0, FS > 60.0, HaplotypeScore > 13.0, MQRankSum < -12.5, ReadPosRankSum < -8.0).

Nous avons ensuite développé des scripts dédiés permettant de ne sélectionner que les SNPs de bonne qualité ayant une profondeur de séquençage supérieure à 18x pour la position donnée.

Un ratio allélique est également calculé par GATK pour chaque position, en divisant le nombre de séquences mutées par le nombre de séquences totales couvrant la position.

Toutes les positions avec un ratio allélique autour de 0,5 (0,15 à 0,85) sont considérées comme hétérozygotes (ratio ABHet). Des ratios alléliques $\geq 0,98$ sont considérés comme homozygotes. Entre 0,85 et 0,98 les résultats sont incertains et donc non considérés pour la suite de l'analyse des SNPs de haute qualité. Les fichiers de sortie rassemblant les SNPs de bonne qualité sont utilisés pour comparer les génomes des différentes souches.

La densité de SNPs hétérozygotes est calculée en découpant le génome en fenêtres de 1 ou 10 kb selon l'analyse. La densité de SNPs est généralement représentée selon une intensité de couleur « heatmap » sur l'ensemble des 8 chromosomes. Il en résulte des zones blanches ou claires correspondant aux régions homozygotes et des régions de plus en plus foncées correspondant aux régions où le nombre de SNPs est plus élevé (**Figure 17**).

Cette analyse nous permet ainsi de visualiser de façon rapide des régions du génome ayant subi de grands événements de LOH de type BIR (**Figure 17**, par exemple du bras droit du Chr3 et le bras gauche du Chr 7).

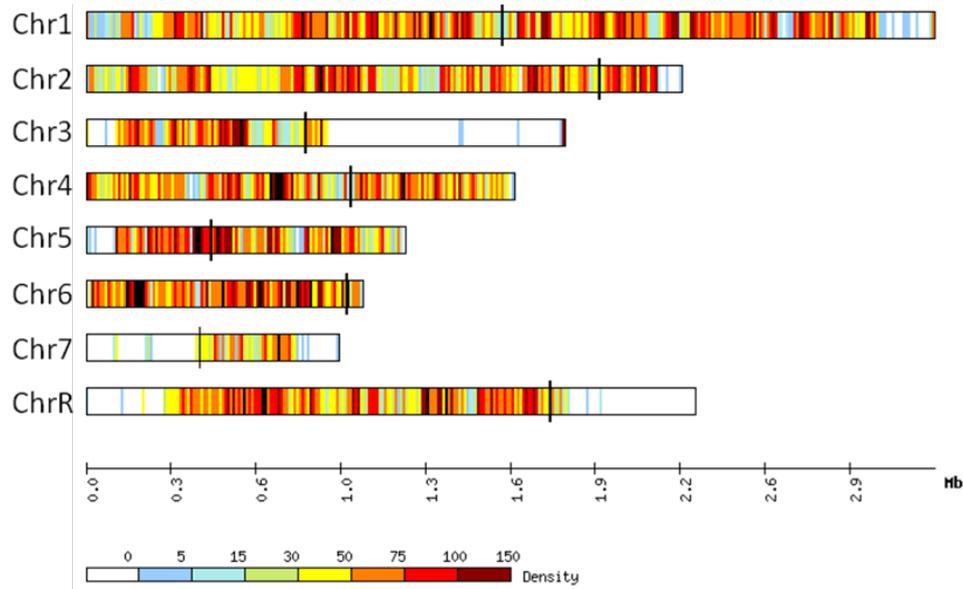


Figure 17 : Heatmap représentant la densité de SNPs hétérozygotes sur les 8 paires de chromosomes de la souche de référence SC5314 (Fenêtre de 10 kb). Les zones blanches à foncées représentent respectivement des régions très homozygotes à très hétérozygotes. Dans cette figure on observe plusieurs événements majeurs de LOH (Bras droit du Chr3 par exemple). Le chromosome 6 présente un niveau élevé d'hétérozygotie (données provenant du laboratoire).

Les données de densité de SNPs hétérozygotes sont également utilisées pour quantifier les événements majeurs de LOH au niveau des extrémités des chromosomes (LOH distaux). La quantification de ces LOH par chromosome peut être représentée sous forme de diagramme (**Figure 18**).

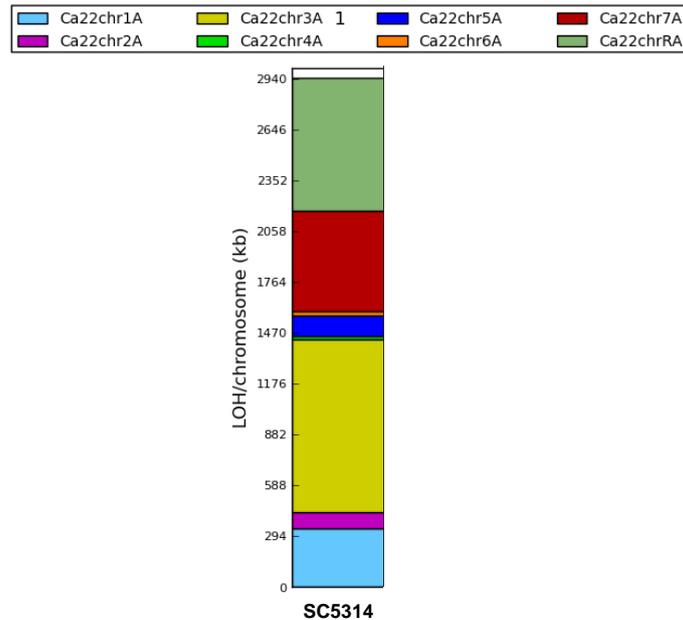


Figure 18 : Quantification en kb des événements distaux de LOH par chromosome pour la souche SC5314 (données provenant du laboratoire).

Cependant, cette approche ne nous permet pas de visualiser de petits événements de LOH, ni de les localiser précisément sur les chromosomes. C'est pourquoi, nous avons développé un script original nous permettant de comparer la survenue de LOH entre deux génomes de souches appariées.

Pour cela, nous avons considéré un événement de LOH lorsque nous observions au moins 2 positions successives présentant une perte d'hétérozygotie (transition d'un génotype hétérozygote à un génotype homozygote) (**Figure 19**).

La taille minimale du LOH a été définie par le nombre de pb située entre la première et la dernière position présentant une perte d'hétérozygotie. La taille maximale du LOH a été définie par le nombre de pb entre la dernière et la première position hétérozygote bordant l'événement de LOH (**Figure 19**).

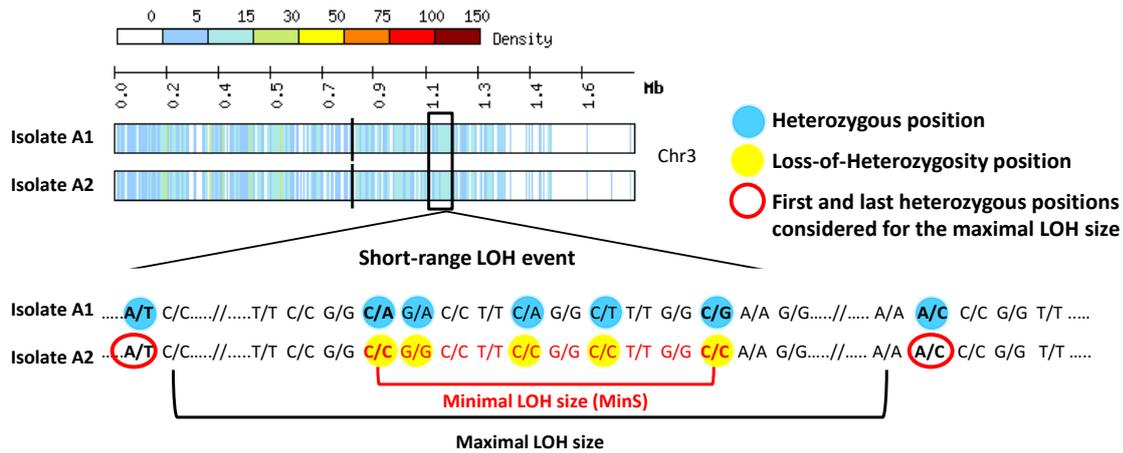


Figure 19 : Détermination de la taille des LOH. Exemple d'un petit événement de LOH observé entre les génomes des isolats A1 et A2. La partie supérieure de la figure représente les heatmaps de densité des SNPs hétérozygotes pour le Chr3 des 2 génomes (analysé sur des fenêtres de 1 Kb). La partie inférieure représente les séquences diploïdes correspondantes des deux génomes. Dans cet exemple, la taille minimale du LOH est de 10 bp, tandis que la taille maximale dépend du nombre de positions homozygotes de part et d'autre du LOH minimum (données personnelles).

1.4 Analyse des variations du nombre de copie (CNVs)

Plusieurs approches complémentaires sont utilisées pour l'analyse des CNVs. En effet, il est possible d'utiliser la profondeur de couverture du génome qui peut être représentée en \log_2 de la moyenne de celle-ci sur différentes portions du génome (fenêtres successives de 1 ou 10 kb). Les valeurs sont normalisées par la médiane de la profondeur de couverture du génome complet, ainsi une portion diploïde ($2n$) sera centrée sur 1 (**Figure 20, A**). Cette analyse permet de détecter des duplications segmentales sur de grandes portions chromosomiques ou des aneuploïdies de chromosomes entiers. En revanche, la répartition allélique ne peut pas être déterminée par cette approche. Pour cela une analyse complémentaire est nécessaire. Celle-ci consiste en l'étude des ratios alléliques de chaque SNP (ABhet) pour les différents chromosomes (**Figure 20, B**). La fréquence de chaque allèle peut être représentée sur un graphique pour chaque SNP détecté en comparaison avec la souche de référence SC5314. Une souche diploïde comportant des SNPs hétérozygotes aura une fréquence pour chaque allèle se situant en moyenne à 0,5.

Une souche triploïde peut contenir soit trois allèles identiques (fréquence allélique de 1) soit présenter deux allèles identiques et un différent (fréquence 0,66 et 0,33). Une souche tétraploïde peut présenter des fréquences alléliques de 0,5 (2x2 allèles identiques) ou de 1 (4 allèles identiques) ou de 0,25 et 0,75 (3 allèles identiques et 1 différent).

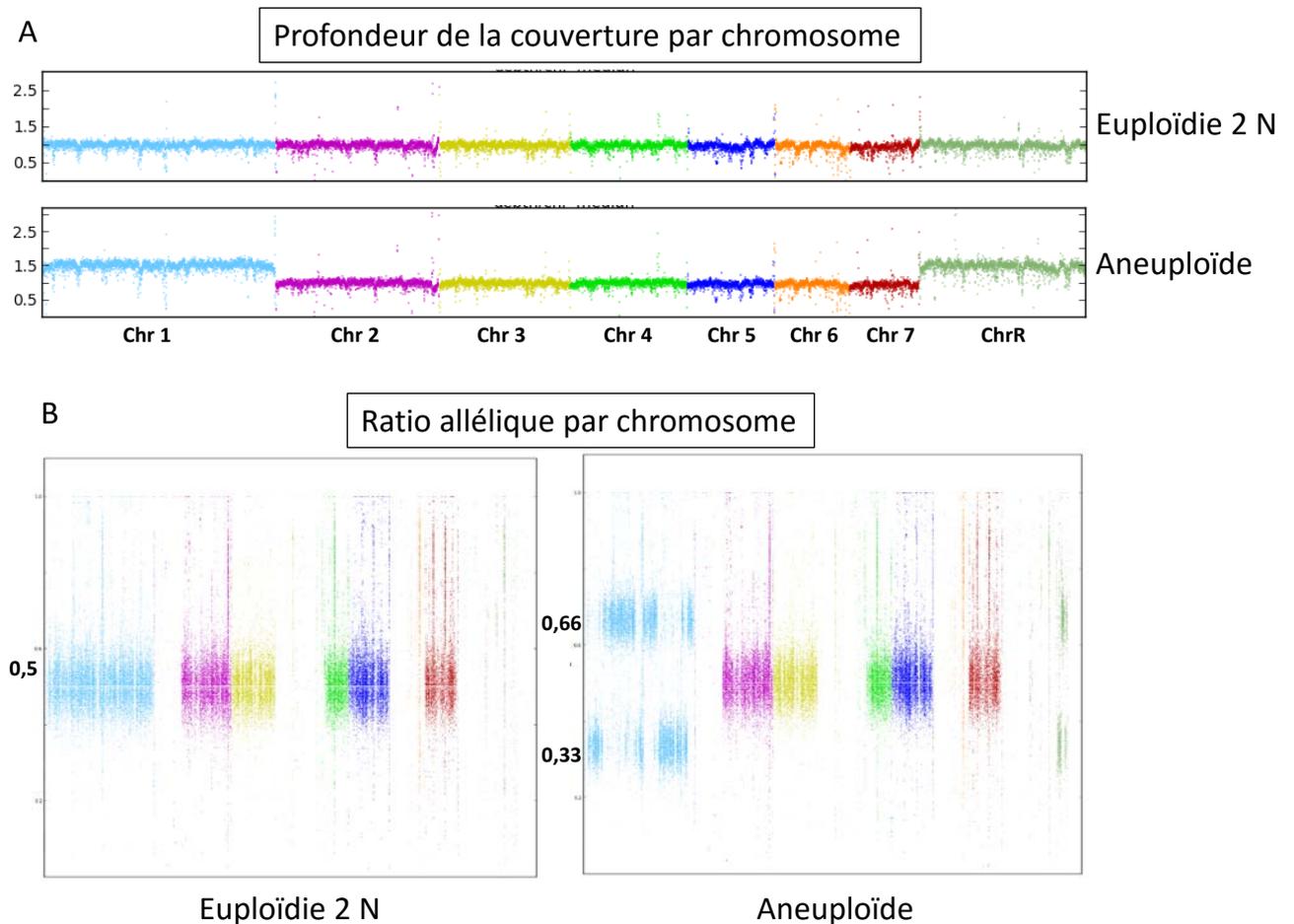


Figure 20 : Détection des aneuploïdies par l'analyse des variations du nombre de copies (CNVs). A. La profondeur moyenne de la couverture est calculée sur des portions de 10 kb puis normalisée par la médiane de celle du génome complet. Les résultats sont présentés en log2. Une souche diploïde se situe aux alentours de 1. Tandis que pour la souche aneuploïde on observe une trisomie des Chr1 et ChrR. B. Analyse des ratios alléliques. Le ratio pour chaque allèle est représenté sur ce graphe, montrant une répartition aux alentours de 0,5 des chromosomes de la souche euploïde et les Chr2 à 7 de la souche aneuploïde. En revanche on observe que ces ratios sont de 0,33 et 0,66 pour les Chr1 et ChrR de cette dernière. Dans cette représentation les portions homozygotes du génome sont détectées par l'absence de points sur le graphe (exemple du Chr6) (données provenant du laboratoire).

2 Etude de la diversité génétique de *C. albicans* au cours du commensalisme

Comme nous l'avons vu au cours de l'introduction, durant le commensalisme, les souches de *C. albicans* peuvent persister plusieurs années chez un même individu et évoluer par le biais de variations génétiques mineures appelées « microévolution » (Bougnoux et al., 2006; Jacobsen et al., 2008). Les souches dites « microévoluées » diffèrent le plus souvent par un ou deux LOH au niveau des loci étudiés en MLST, résultant de mutations ponctuelles ou d'événements de recombinaison homologue (Bougnoux et al., 2006; Diogo et al., 2009; Jacobsen et al., 2008; Odds et al., 2006; Takagi et al., 2013; Xu et al., 1999). Ces événements peuvent être concomitants à d'autres remaniements du génome tels que des altérations de la taille des chromosomes ou de la ploïdie, suggérant que le génome de *C. albicans* présente également une grande plasticité au cours du commensalisme (Diogo et al., 2009). Cependant, l'étendue de la diversité génétique au sein des populations cellulaires de *C. albicans* présentes chez un individu n'a jamais été évaluée à l'échelle du génome complet.

La connaissance du niveau basal d'hétérogénéité génétique des souches de *C. albicans* au cours du commensalisme est une information cruciale. En effet, elle conditionne l'interprétation des variations génétiques qui pourraient être observées dans les séries chronologiques de souches de *C. albicans* isolées au cours de la colonisation chronique et pathologique chez les patients atteints de CCMC. C'est pourquoi, nous avons souhaité déterminer et quantifier préalablement la diversité génétique des isolats de *C. albicans* chez le sujet « porteur sain », c'est-à-dire dans le cadre d'une simple colonisation commensale. Pour cela plusieurs isolats de *C. albicans* (n=3) issus d'un même prélèvement buccal de différents individus (porteurs sains, n=3) ont été séquencés et analysés via le pipeline développé dans notre laboratoire.

L'analyse comparative des génomes nous a permis d'identifier de façon exhaustive l'ensemble des SNPs qui différiaient les génomes entre eux et de cartographier les LOH. Bien que très proches (même profil MLST), les génomes des isolats issus d'un même prélèvement se sont différenciés par un nombre important de SNPs – allant de 549 à 5119 SNPs par comparaison. Ces SNPs étaient, en majorité, impliqués dans des événements de remaniement génomique de type LOH – allant de 78 à 297 LOH de différence par

comparaison. Cependant, la fréquence de survenue de ces événements était variable entre les isolats.

Ainsi, nous avons pu mettre en évidence qu'au sein d'un même prélèvement, plusieurs populations cellulaires de *C. albicans* pouvaient coexister.

Pour confirmer que ces événements n'étaient pas imputables à des erreurs de séquençage, ou liées au processus d'analyse ou à la reproduction clonale de la souche *in vitro* (lors des différents repiquages), nous avons quantifié et comparé le niveau de variabilité observé entre ces isolats et le niveau observé entre des isolats clonaux. Pour cela nous avons sélectionné des souches indépendantes de *C. albicans* et séquencé pour chacune d'entre elles plusieurs colonies obtenues lors du repiquage de la souche (isolats clonaux). La comparaison du nombre d'événements identifié entre les isolats clonaux était en moyenne de 400 SNPs et 69 LOH. Le niveau de diversité retrouvé entre les isolats buccaux était significativement supérieur au niveau de diversité observé entre les isolats clonaux (SNPs $p=0,007$, LOH $p=0,0013$), confirmant l'existence d'une hétérogénéité génétique au sein des souches de *C. albicans* présentes chez un individu. Nous avons également comparé la distribution des SNPs sur le génome et montré que la fréquence des mutations était significativement plus élevée dans les régions répétées, que ce soit dans les isolats clonaux ou buccaux. En revanche, chez les isolats buccaux la fréquence des mutations était significativement supérieure dans les régions intergéniques, et proche de la significativité pour les ORFs, par rapport à celle observée dans les isolats clonaux.

Dans cette première partie, nous avons prouvé qu'il existait un niveau élevé d'hétérogénéité génétique au sein des souches de *C. albicans* au cours du commensalisme et démontré que cette hétérogénéité n'était pas due à des problèmes inhérents à la technique et à l'analyse des données de séquençage à haut débit. Une telle diversité génétique pourrait favoriser l'adaptation rapide de *C. albicans* en condition de stress lié à un changement d'environnement. La connaissance de ces données était nécessaire pour pouvoir évaluer et interpréter les variations génétiques que nous pourrions observer au cours du temps entre les génomes provenant de nos séries de souches issues des patients atteints de CCMC.

Publication 1: *Within-Host Genomic Diversity of Candida albicans in Healthy Carriers: Implication for Microevolution Studies. (Manuscrit soumis à Scientific Reports; SREP-18-31879) (Annexe 1). Emilie Sitterlé, Corinne Maufrais, Matthieu Palayret, Natacha Sertour, Christophe d'Enfert, Marie-Elisabeth Bournoux*

Abstract

Genomic variations in *Candida albicans*, a major fungal pathogen of humans, have been observed upon exposure of this yeast to different stresses and experimental infections, possibly contributing to subsequent adaptation to these stress conditions. Yet, little is known about the extent of genomic diversity that is associated with commensalism, the predominant lifestyle of *C. albicans* in humans. In this study, we investigated the genetic diversity of *C. albicans* oral isolates recovered from healthy individuals, using multilocus sequencing typing (MLST) and whole genome sequencing. While MLST revealed occasional differences between isolates collected from a single individual, genomic sequencing showed that they differed by numerous single nucleotide polymorphisms, mostly resulting from short-range loss-of-heterozygosity events. These differences were shown to have occurred upon human carriage of *C. albicans* rather than subsequent *in vitro* manipulation of the isolates. Thus, *C. albicans* intra-sample diversity appears common in healthy individuals, higher than that observed using MLST. We propose that diversifying lineages coexist in a single human individual, and this diversity can enable rapid adaptation under stress exposure. These results are crucial for the interpretation of longitudinal studies evaluating the evolution of the *C. albicans* genome.

3 Diversité phénotypique et génomique de *C. albicans* au cours de l'interaction pathologique prolongée chez l'Homme

3.1 Introduction

C. albicans est, d'une part un commensal du tube digestif de l'Homme, qui constitue son principal réservoir, et, d'autre part un pathogène opportuniste responsable d'infections dont la gravité (superficielles à disséminées) est fonction de l'état physiologique et des défenses immunitaires de l'hôte. Bien que l'altération des fonctions immunitaires et la dysbiose demeurent les déterminants principaux de la survenue des infections opportunistes, l'espèce *C. albicans* dispose de capacités particulières, lui permettant de passer d'un état commensal à un état pathogène. Cependant, toutes les souches de *C. albicans* ne sont pas équivalentes. La variabilité génomique et phénotypique au sein de l'espèce peut également contribuer à des variations dans le développement et la sévérité de l'infection (Braunsdorf and LeibundGut-Landmann, 2018; Cavaliere et al., 2017). En effet, au cours de l'infection, *C. albicans* est confronté à divers stress environnementaux et immunitaires auxquels il doit s'adapter pour pouvoir croître et survivre (Forche, 2014). L'hétérogénéité génomique retrouvée au sein de la population cellulaire pourrait faciliter l'adaptation de cette levure à des changements environnementaux. La génération de variants génétiques est grandement facilitée par la plasticité de son génome. En effet, comme nous l'avons vu précédemment, en plus de l'hétérozygotie importante, des événements majeurs de remaniement des chromosomes peuvent survenir induisant une modification du nombre de copies de gènes situés sur le chromosome affecté. La variation du nombre de copies de ces gènes peut ainsi conférer un avantage sélectif dans certaines conditions (Selmecki et al., 2008). Cependant les aneuploïdies sont instables et donc souvent transitoires, l'état diploïde peut être retrouvé en l'absence de pression de sélection (Gerstein and Berman, 2015). De simples mutations ponctuelles peuvent également survenir dans certains gènes, dont la fonction permet aux cellules, de façon durable, de mieux tolérer l'environnement défavorable.

L'acquisition de mutations avantageuses peut être suivie d'une perte d'hétérozygotie (LOH) de l'allèle muté par des événements de recombinaison mitotique ou par perte du chromosome contenant l'allèle de type sauvage (Coste et al., 2006; Dunkel et al., 2008). La survenue de ces événements a largement été étudiée *in vitro* et *in vivo* essentiellement chez la souris, dans le cadre de l'adaptation de *C. albicans* en réponse aux antifongiques, mais également en réponse à diverses modifications environnementales (Forche et al., 2018, 2011, 2009; Hampe et al., 2017; Pande et al., 2013; Selmecki et al., 2010; Wartenberg et al., 2014). En revanche, hormis les études réalisées par des approches de typage moléculaire, très peu de travaux se sont portés sur la diversité génétique de *C. albicans* et les mécanismes génomiques adaptatifs survenant au cours de l'interaction pathologique chez l'Homme (Ford et al., 2015; Kim et al., 2015). Le défi étant d'identifier des variants génétiques qui se traduisent par des modifications phénotypiques impactant l'équilibre entre le commensalisme et la pathogénicité de *C. albicans* au cours de l'interaction avec l'Homme.

La Candidose Cutané-Muqueuse Chronique (CCMC) est un cas particulier d'interaction entre *C. albicans* et l'hôte qui se caractérise par des infections sévères, persistantes ou récurrentes de la peau, des ongles et des muqueuses par *C. albicans*. Au cours de la CCMC induite par une anomalie génétique primitive, les malades peuvent être porteurs de façon chronique d'une souche unique (même profil génotypique) de *C. albicans*. Ainsi, chaque souche peut évoluer chez un même patient durant des années, voire toute sa vie sous une forme « pathogène » normalement observée de façon fugace au cours des infections candidosiques invasives. La gravité des infections candidosiques des muqueuses, de la peau et des phanères chez ces patients impose pour la plupart d'être traités de façon prolongée ou répétée par divers antifongiques. L'utilisation de ces molécules sur une longue période peut être responsable de l'acquisition d'une résistance, contribuant à la difficulté de prise en charge de la CCMC. Cette pathologie constitue donc un modèle privilégié pour l'étude de l'adaptation de *C. albicans* chez l'Homme dans un contexte d'interaction pathologique prolongé.

L'objectif de cette partie étant d'utiliser ce « modèle », pour lequel nous suivons une importante cohorte de patients atteints de CCMC, pour explorer si *C. albicans* est capable d'adaptations génomiques spécifiques dans un contexte d'interaction à long terme avec un environnement de l'hôte modifié.

Cet environnement spécifique est induit d'une part par une réponse immunitaire et inflammatoire inappropriée de l'hôte, mais également par l'utilisation d'antifongiques chez ces patients durant une très longue période. Pour cela, nous avons constitué une collection de souches de *C. albicans* isolées sur plusieurs années chez des patients atteints de CCMC et leurs familles, sur laquelle nous avons réalisé une caractérisation phénotypique et génomique. Les résultats obtenus ont été comparés à ceux d'une collection « contrôle » de 144 souches cliniques pour lesquelles nous disposons des données phénotypiques et de séquençages.

3.2 Matériels et méthodes et Résultats

3.2.1 Constitution de la cohorte et caractérisation génotypique des souches de patients atteints de CCMC et de leur famille

3.2.1.1 Description des différentes collections de souches

Treize patients atteints de CCMC ainsi que 3 membres non malades de leur famille ont été inclus dans cette étude. La majorité des patients présentaient une mutation hétérozygote gain-de-fonction (GOF) du gène *STAT1* (n=9/13). Quatre patients présentaient un déficit autosomique récessif en *IL-17RA* (n=2) ou *ACT1* (n=2). La moyenne d'âge de ces patients était de 26 ans, allant de 2 à 64 ans.

Les 13 patients appartenaient à 8 familles différentes. En effet, chez 5 familles plusieurs membres malades et des individus sains ont pu être étudiés. Au total, parmi les 8 familles, 3 étaient composées de 2 membres malades (Famille F1 : P1 et P2 ; Famille F2 : P3 et P4 ; Famille F4 : P6 et P10), 1 famille (F5) était composée de 2 malades (P7 et P8) et 1 membre non malade (I3), et 1 famille était composée de 2 malades (P12 et P13) et 2 membres non malades (I1 et I2), enfin les 3 dernières familles (F3, F6 et F7), étaient composées d'un seul malade (P5, P9 et P11).

Nous avons constitué une collection de 63 souches de *C. albicans* - 60 isolées chez les 13 patients atteints de CCMC et 3 isolées chez des membres non malades de leur famille (**Figure 21**).

Les 63 isolats provenaient essentiellement de la bouche 65% (n=41), de selles 12% (n=8) mais également de la peau 11% (n=7), ongle n=3, vagin n=2 et 1 au niveau du nez (**Figure 21**). Une partie de ces isolats ont été obtenues chronologiquement chez 7 des 13 patients sur une période allant jusqu'à 9 ans (**Figure 21**). La vérification de l'identification de toutes les souches a été réalisée par spectrométrie de masse de type MALDI-TOF permettant de confirmer que les souches appartenaient bien à l'espèce *C. albicans stricto sensu*.

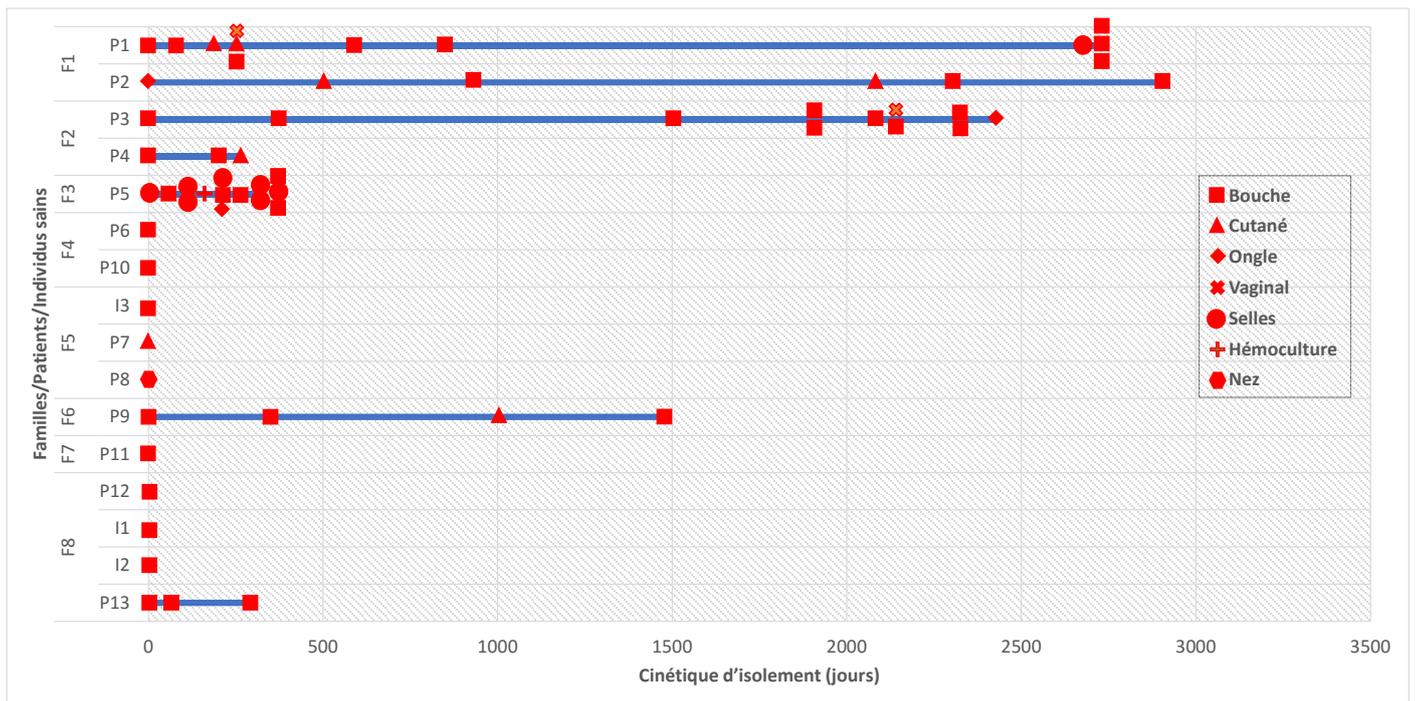


Figure 21 : Composition de la cohorte de patients atteints de CCMC et des membres non malades de leur famille et cinétique d'isolement des 63 souches de *C. albicans* de la collection. (F=Famille ; P=Patient ; I= Individu sain)

La collection de souches dite « contrôle » était constituée de 144 isolats cliniques de *C. albicans*, épidémiologiquement non liés, pour lesquels nous disposions des données de séquençage du génome mais également des données épidémiologiques, cliniques et phénotypiques (Ropars et al., 2018). Ces souches ont été isolées chez des individus sains ou des malades présentant une candidose systémique ou des muqueuses. Aucune des souches de cette collection ne provenait de sujets atteints de CCMC. Par ailleurs, toutes les souches étaient sensibles aux antifongiques. La souche de référence SC5314 a été utilisée comme contrôle.

3.2.1.2 Caractérisation génotypique des souches de patients CCMC et de leur famille

Afin de déterminer le lien génotypique des isolats entre patients mais également au cours du temps et au sein des différentes familles, nous avons dans un premier temps typé les 63 isolats obtenus par Multi Locus Sequence Typing (MLST) (Bougnoux et al., 2003). L'ensemble des résultats est détaillé dans le **Tableau 5**.

Les souches sont réparties dans 5 clades différents (clade : 1, 2, 4, 10 et 11) avec une majorité d'entre elles (71,4%, 45/63) issues du clade 11. Les clades 11, 1 et 4 sont des clades fréquents regroupant des souches à prédominance européennes. En revanche, aucune souche provenant des clades 3, 5, ou 6, connus comme étant enrichis en isolats oropharyngés, n'a été retrouvée.

- Caractérisation génotypique des isolats au sein d'un même patient

Chez 7 patients nous disposons de plusieurs isolats séquentiels, collectés sur des périodes plus ou moins longues, allant de 1 à 9 ans. Nous avons pu observer chez 5 d'entre eux (patients P2 à P5 et P13) la persistance d'une souche ayant un même profil MLST (même DST) et chez 2 patients (P1 et P9) l'apparition au cours du temps d'une souche dite « microévoluée » (modification du DST due à un ou deux événements de LOH aux loci étudiés). Enfin, chez ces 7 patients, seuls deux d'entre eux ont été infectés ponctuellement par un isolat de génotype MLST différent (P2 : 1/6 des isolats testés et P5 : 1/14) (**Tableau 5**).

La caractérisation des isolats provenant de différents sites anatomiques (bouche, ongle, peau, vaginal) a montré qu'un patient pouvait être colonisé ou infecté par une souche présentant un profil MLST unique ou microévoluée au niveau des différentes localisations. Ces souches peuvent donc persister et évoluer chez un même individu dans des niches très différentes (**Tableau 5**).

Famille	Individu / patient	Isolat (n) (date)	DST	MLST locus							Clade	
				AAT1	ACC1	ADP1	MPI	SYA1	VPS13	ZWF1		
FLO												
F1	P1	FLOJ (8) (2007-> 2009)	1327	62	12	21	1	6	YCRTGCT	30	CYRGGTWC	22
	P1	FLOJ (4) (2014)	2883	62	12	21	1	4	CCGTGCT	22	CTGGGTC	22
	P2	FLON (5) (2007->2015)	1327	62	12	21	1	6		30		22
	P2	FLON18 (1) (2009)	2884	2	5	5	4	2		24		97
BLO												
F2	P3	BLOV (12) (2008->2015)	912	62	12	21	1	6		30		20
	P4	BLOG (3) (2014)	912	62	12	21	1	6		30		20
HAM												
F3	P5	HAME (13) (2012->2013)	2886	62	12	21	5	6		30		4
		HAME41 (1) (2013)	2888	157	3	8	4	7		3		8
DES												
F4	P6	DESD (1) (2015)	3193	35	7	4	4	4		35		4
	P10	DESM (1) (2015)	3194	4	7	155	5	220		51		14
BEY												
F5	P7	BEYH (1) (2013)	1132	4	22	6	18	64		53		15
	P8	BEYM (1) (2013)	1132	4	22	6	18	64		53		15
	I3	BEYA (1) (2013)	2887	2	5	144	2	2		53		5
BOIN												
F6	P9	BOIN (2) (2009->2010)	69	2	5	5	2	2		6	CYARGKWY	5
		BOIN (2) (2012->2013)	24	2	5	5	2	2		24	CTAAGGTT	5
FOF												
F7	P11	FOFF (1) (2012)	2885	58	14	8	4	7		3		35
ARI												
F8	P12	ARIR (1) (2015)		58	14	8	4	7		10		8
	I1	ARIM (1) (2014)		58	14	8	4	7		10		8
	I2	ARIH (1) (2015)	504	58	14	8	4	7		10		8
	P13	ARIS (3) (2014->2015)		58	14	8	4	7		10		8

Tableau 5 : Caractéristiques MLST des 63 souches de *C. albicans* isolées chez les patients CCMC et leur famille. Pour chaque patient/individu le nombre de souches isolées au cours du temps est noté entre parenthèse (n) ainsi que la période d'isolement correspondante. Le profil MLST obtenu pour chaque allèle (7 loci testés) est indiqué dans la table, ainsi que le DST (<https://pubmlst.org/calbicans/>). Les souches dites « microévoluées » correspondent aux souches ayant subi un événement de LOH sur l'un des loci (en rouge dans le tableau). Une représentation simplifiée de la séquence nucléotidique est présentée (en bleu sont indiquées les positions polymorphiques). Les souches notées en gris présentent un profil MLST très différent des autres souches isolées chez le même individu. (Abréviations: n= nombre d'isolat, LOH = Loss-of-heterozygosity, F= Famille, P= Patient, I= Individu sain, DST= Diploid sequence typing, NC: Non cladable / = non testée).

- Caractérisation génotypique des isolats entre individus (sains et/ou malades d'une même famille)

Premièrement, la comparaison des souches entre patients CCMC d'une même famille (F1, F2, F4, F5 et F8) a montré que les malades d'une même famille partageaient la même souche sauf pour la famille (F4) où les deux patients (P6 et P10) étaient colonisés par des souches de *C. albicans* appartenant à un clade différent (**Tableau 5**).

Deuxièmement, concernant la comparaison d'individus sains et malades d'une même famille celle-ci a pu être réalisée pour deux familles (F5 et F8). Pour la famille F5, la souche provenant de l'individu sain (I3, souche #BEYA) avait un profil différent que celui retrouvé chez les deux malades (P5 et P6). Alors que dans la deuxième famille (F8) les deux membres sains (I2 et I3) partageaient la même souche que les deux malades (P12 et P13). Cette famille constitue une source d'isolats extrêmement intéressante pour la comparaison de l'évolution des souches chez un individu sain et un individu malade de l'évolution des souches chez un individu sain et un individu malade (**Tableau 5**).

Pour conclure, cette analyse génotypique a permis de montrer, par méthode conventionnelle de typage moléculaire, que les patients étaient colonisés ou infectés à long terme par une souche unique (même profil MLST sur une période allant jusqu'à 9 ans). La survenue chez un patient d'une souche différente est rare et semble transitoire. Nous avons également montré qu'au sein d'une même famille, les membres sains et malades pouvaient être colonisés par la même souche (même profil MLST).

Grâce à ces données de typage, nous avons sélectionné, au sein de notre collection, 37 souches d'intérêt (**Tableau 6**) afin de réaliser une caractérisation phénotypique étendue et le séquençage de leur génome. Les données phénotypiques et génomiques ont par la suite été utilisées dans les études comparatives suivantes :

- Comparaison des souches provenant de patients « CCMC » avec des souches contrôles (collection des 144 souches de *C. albicans*).
- Etude de l'évolution d'une souche identique ou microévoluée sur des séries de souches chronologiques chez un même patient.
- Comparaison de l'évolution d'une souche identique ou microévoluée chez un même patient mais isolée sur des sites anatomiques différents.
- Comparaison de l'évolution d'une même souche chez un patient atteint de CCMC et des membres sains de sa famille.

Famille	Individu/ patient	Nom de l'isolat	Date d'isolement	Site
FLO				
F1	P1	FLOJ22	mars-08	Bouche
		FLOJ23	févr.-09	Bouche
		FLOJ19	mars-08	Vaginal
		FLOJ20	mars-08	Cutané
	P2	FLOJ12	déc.-14	Bouche
		FLOJ13	juin-07	Ongle
		FLOJ17	nov.-08	Cutané
		FLOJ02	avr.-13	Cutané
	FLOJ07	sept.-13	Bouche	
	FLOJ68	juin-15	Bouche	
BLO				
F2	P3	BLOV27	oct-08	Bouche
		BLOV29	nov-09	Bouche
		BLOV32	nov-11	Bouche
		BLOV37	nov-12	Bouche
		BLOV56	sept-14	Vaginal
		BLOV53	sept-14	Bouche
		BLOV03	janv-14	Bouche
	P4	BLOV67	juin-15	Ongle
		BLOG54	sept-14	Bouche
		HAM		
F3	P5	HAME38	déc.-12	Bouche
		HAME10	août-13	Bouche
DES				
F4	P6	DESM70	oct-15	Bouche
	P10	DESD69	oct.-15	Bouche
BEY				
F5	P7	BEYH08	juil-13	Cutané
	P8	BEYM05	juil-13	Nez
BOIN				
F6	P9	BOIN28	oct.-09	Bouche
		BOIN30	avr.-10	Bouche
		BOIN34	juil.-12	Cutané
		BOIN06	nov.-13	Bouche
FOF				
F7	P11	FOFF33	mai-12	Bouche
ARI				
F8	P12	ARIR62	juin-15	Bouche
	I1	ARIM66	juil-14	Bouche
	I2	ARIH61	juin-15	Bouche
	P13	ARIS65	sept-14	Bouche
		ARIS64	juil-14	Bouche
		ARIS60	mai-15	Bouche

Tableau 6 : Liste des 37 souches de *C. albicans*, provenant de patients atteints de CCMC et de leur famille, sélectionnées au sein de notre collection pour le séquençage du génome. (F=Famille ; P=Patient ; I= Individu sain).

3.2.2 Caractérisation phénotypique des souches de patients atteints de CCMC et de leurs familles

Différentes conditions expérimentales pertinentes au regard de l'interaction avec l'hôte ont été testées *in vitro* et résumées dans le **Tableau 7** : i) sensibilité vis-à-vis de divers antifongiques ii) cinétique de croissance iii) sensibilité à différents stress (oxydatif, cationique, osmotique, thermique), iv) capacité à réaliser une transition levure-filament en réponse à différents stimuli et v) capacité à lyser les kératinocytes humains.

Condition tested	Supplementation	N*	Medium	Temperature	Method
Susceptibility to antifungal	fluconazole (Etest®)	1	RPMI	37°C	CMI determination (µg/ml)
	itraconazole (Etest®)	1	RPMI	37°C	
	voriconazole (Etest®)	1	RPMI	37°C	
	caspofungin (Etest®)	1	RPMI	37°C	
	anidulafugin (Etest®)	1	RPMI	37°C	
	5-fluorocytosine (Etest®)	1	RPMI	37°C	
Growth rate	without any supplementation	4	YPD	30°C	Doubling time (min)
	without any supplementation	4	SD	30°C	
Stress conditions	without any supplementation	2	SD	30°C ad 37°C	
	without any supplementation	4	YPD	30°C ad 37°C	
Temperature	Temperature - Low	1 / 4	SD and YPD	18°C	
	Temperature - High	1 / 4	SD and YPD	42°C	
Ion homeostasie	Lithium Chloride (300mM)	1	SD	30°C	
	ZnSO4 (5mM)	1	SD	30°C	
	EDTA (0,75 mM)	1	SD	30°C	
	NaCl (1M)	1	SD	30°C	
	NaCl (1M)	2	YPD	30°C and 37°C	
	CaCl2 (300 mM)	1	SD	30°C	
Cell wall integrity pathways	Caffeine (20 mM)	1	SD	30°C	Size of colonies normalized by control plate and SC5314 results
	Caffeine (15 mM)	2	YPD	30°C and 37°C	
	Calcofluor white (20 µM)	2	YPD	30°C and 37°C	
	Calcofluor white (20 µM)	1	SD	30°C	
	SDS (0.005% (w/v))	2	YPD	30°C and 37°C	
	SDS (0.01% (w/v))	1	SD	30°C	
	Congo red (15 µg/ml)	1	SD	30°C	
	Congo red (10 µg/ml)	2	YPD	30°C and 37°C	
	Congo red (20 µg/ml)	2	YPD	30°C and 37°C	
Osmotic stress	Sorbitol (1.5M)	1	SD	30°C	
Oxydative stress	H ₂ O ₂ (5mM)	1 / 1	SD and YPD	30°C and 37°C	
	H ₂ O ₂ (4,5mM)	2	YPD	30°C and 37°C	
	H ₂ O ₂ (5mM)	2	YPD	30°C and 37°C	
	Tert-Butyl-hydroperoxyde (5mM)	1	SD	30°C	
	Menadione (80 µM)	2	YPD	30°C and 37°C	
Carbon source	galactose (2% (w/v))	1	YPGal	30°C	
Antimycotics	5-Fluorocytosine (3.5µM)	2	SD	30°C and 37°C	
Morphology					
Hyphae induction	Spider medium	5	Spider	30°C and 37°C	Scoring ranged 0 to 4
	10% foetal calf serum	3	YPD	30°C and 37°C	
	Anaerobic Chamber (BBL GasPak)	1 / 1	SD and YPD	30°C and 37°C	
Invasion	Embedded condition	1 / 1	SD and YPD	30°C and 37°C	
Chlamydospore	N-Acetylglucosamine	1	N-Acetylglucosamine	25°C	
Keratinocyte damage					
	12h	3	TR146 human Keratinocytes monolayer infected for 24h		LDH release express in % damage relative to uninfected and high damage control (1% triton in F12)
	24h				

Tableau 7 : Caractérisation phénotypique des souches dans différentes conditions.
(N* : nombre de réplica)

3.2.2.1 Sensibilité des souches aux antifongiques

Nous avons déterminé la sensibilité des souches par mesure des Concentrations Minimales Inhibitrices (CMI) de 6 antifongiques par méthode Etest® (fluconazole (Fluco), itraconazole (Itra), voriconazole (Vori), 5-fluorocytosine (5-FC), caspofungine et anidulafungine (AND). Pour l'interprétation des CMI nous avons utilisé les breakpoints définis par l'EUCAST (V9.0, 2018).

Les résultats obtenus pour les 37 souches sont représentés dans la partie A de la **Figure 22**. Parmi les 37 souches testées, 22 (59,5%) étaient résistantes aux 3 antifongiques azolés (fluconazole, itraconazole et voriconazole) avec un haut niveau de résistance pour 18 d'entre elles (81,8%). Ces souches provenaient de 5 patients (P1 à P5). Pour un de ces patients (P3) le niveau de résistance au fluconazole a augmenté entre l'isolat de 2008 et celui de 2011 (augmentation de la CMI de 24 à 256 µg/ml).

Pour les échinocandines (caspofungine et anidulafungine), 10,8% (n=4) des souches avaient un faible niveau de résistance *in vitro*. Ces 4 souches ont été isolées chez 2 patients (P2 et P3) pour lesquels la résistance phénotypique des souches a été acquise au cours du temps. Enfin, concernant la 5-Fluorocytosine, 29,7% (n=11) des souches avaient un haut niveau de résistance (CMI > 32 µg/ml). Toutes les souches résistantes à cette molécule provenaient des patients P1 et P2 de la famille 1.

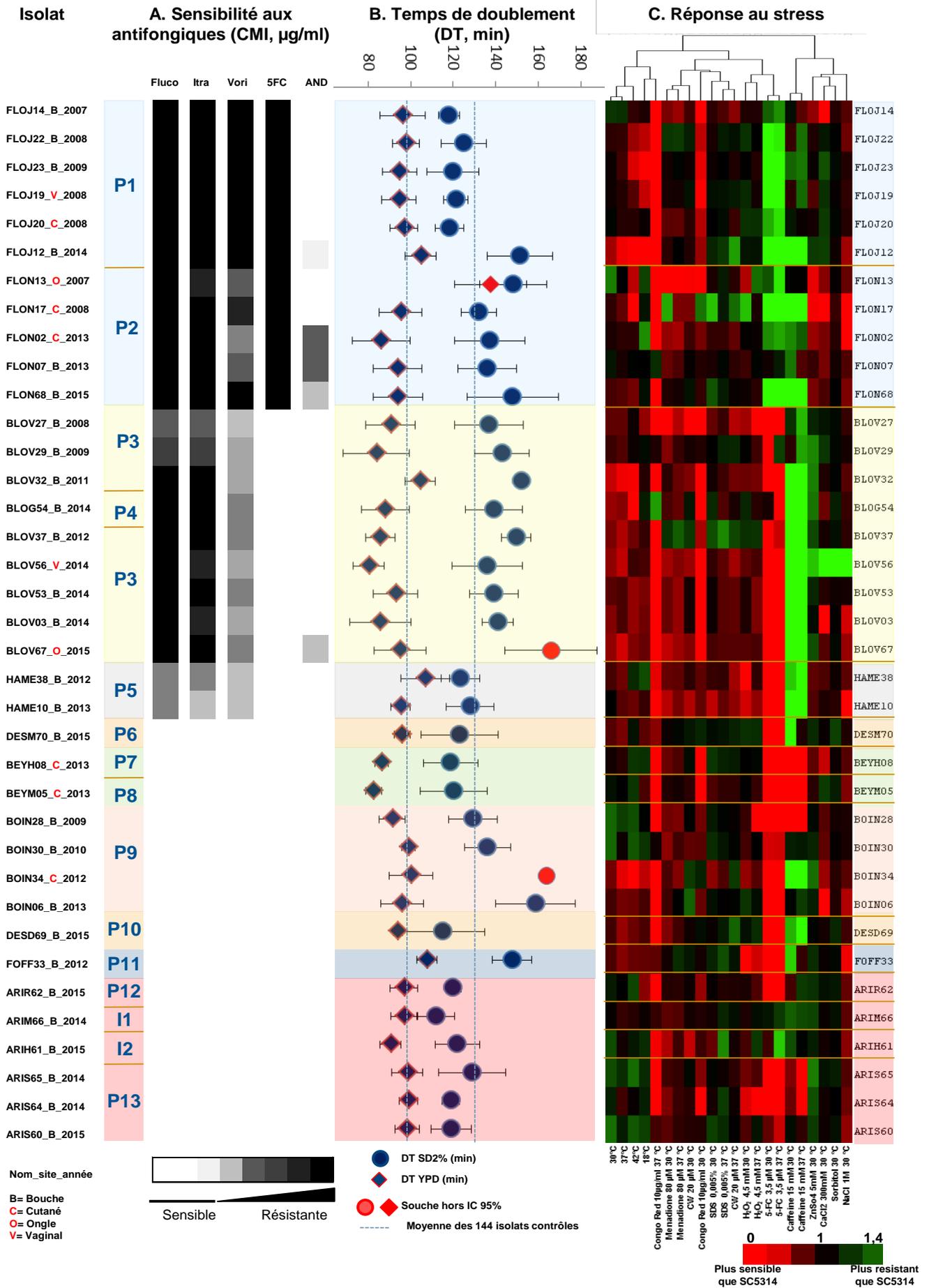


Figure 22 : Caractérisation phénotypique des 37 souches provenant des patients CCMC et de leur famille. A. Sensibilité des souches aux divers antifongiques. Les CMI ont été interprétées selon les breakpoints définis par l'EUCAST (V9.0, 2018) et représentées selon une échelle de niveau de gris (blanc, souche sensible ; gris à noir souche de sensibilité intermédiaire à haut niveau de résistance). (Fluco= fluconazole ; Itra= itraconazole ; Vori= voriconazole ; 5-FC= 5 Fluoro-cytosine ; AND = anidulafungine) B. Temps de doublement des souches (DT, min en YPD et SD2% à 30°C). La moyenne du DT obtenue pour les 144 souches contrôles est représentée en pointillée. Les DT des souches de patients CCMC sortant de l'intervalle de confiance à 95 % (IC95 YPD = 97 +/- 34,6 ; IC95 SD = 133 +/- 31) sont représentées en rouge. C. Réponse aux stress. Les scores normalisés sont représentés sous forme de heatmap. En niveaux de rouge sont représentées les souches plus sensibles que SC5314 (score normalisé < 0,8) et en niveaux de vert les souches plus résistantes (score normalisé > 1,2).

Au total sur les 13 patients CCMC, 5 (38%) étaient colonisés ou infectés par au moins une souche résistante aux antifongiques azolés. Parmi ces 5 patients, 2 étaient porteurs de souches résistantes à deux antifongiques (P1 à la 5-FC ; et P3 aux échinocandines) et 1 patient (P2) était porteur de souches résistantes aux 3 classes d'antifongiques testées. Bien que le clade 1 soit décrit comme enrichi en souches résistantes à la 5-FC, les souches de notre étude, résistantes à cette molécule appartenaient au clade 11 (Dodgson et al., 2004; Pujol et al., 2004). L'ensemble de ces résultats souligne le niveau élevé de résistance aux antifongiques retrouvé chez les patients atteints de CCMC.

3.2.2.2 Cinétique de croissance des souches

Les modifications du génome dans les souches de *C. albicans*, en particulier celles liées à l'apparition de variations de structure telles que les aneuploïdies, sont connues pour affecter leur capacité de prolifération. Afin d'évaluer la capacité des souches des patients CCMC à proliférer *in vitro*, nous avons déterminé pour chacune des 63 souches le temps de doublement (DT ; doubling time) à partir des courbes de croissance des souches. Pour cela la densité optique (DO) des souches cultivées à 30 °C en milieu liquide (YPD et SD2% ; milieu riche et pauvre respectivement) a été mesurée toutes les 10 min sur une période de 24h avec un lecteur de DO de type TECAN[®]. Le temps de doublement (en minutes) a été déterminé durant la phase exponentielle de croissance des souches et correspond à la moyenne au temps que met une souche pour passer d'une DO_{600nm} de 0,2 à 0,4 et 0,4 à 0,8.

Les résultats du DT des 37 souches sélectionnées ont été comparés à celui des souches contrôles et sont représentés dans **partie B de la Figure 22**.

Dans leur ensemble, les souches provenant des patients CCMC et de leur famille ne présentaient pas de défaut de croissance *in vitro*. En effet, le DT moyen était de 97 min en YPD (extrêmes : 80 - 138) et de 133 min en SD2% (extrêmes : 112 - 166). Bien que ces résultats ont montré une diversité importante dans la capacité des souches des patients CCMC à croître à 30°C, tant en YPD qu'en SD2%, le DT moyen dans cette population CCMC était comparable au DT moyen des souches contrôles (DT moy = 96 (extrêmes : 77 - 154) en YPD et 129 (extrêmes : 102 - 192) en SD2%). Dans cette analyse, 3 patients ont présenté une souche ayant un défaut de croissance : P1 (souche #FLON13) avec un DT= 138 min en YPD ; P3 (souche #BLOV67) et P9 (souche #BOIN34) avec un DT de 166 et 164 min en SD2%, respectivement (**Figure 22, partie B**). A noter que ces 3 souches provenaient de sites anatomiques particuliers (2 d'un ongle et une de la peau).

On note également, un « switch » phénotypique de certaines souches (augmentation du DT en SD2%) au cours du temps chez 3 patients (P1, #FLOJ12_B_2014 ; P3, #BLOV67_O_2015 et P9, #BOIN34_C_2012 et #BOIN06_B_2013) (**Figure 22, partie B**). Ces souches correspondent aux isolats les plus récents obtenus chez ces patients, suggérant une évolution phénotypique des souches au cours de l'interaction pathologique chez l'Homme.

3.2.2.3 Réponse aux stress environnementaux

Chez certains patients CCMC, les souches de *C. albicans* ont été isolées au niveau de plusieurs sites anatomiques variés (bouche, peau, ongle vagin). Ainsi, ces souches ont pu évoluer et diverger au cours du temps dans ces environnements extrêmement divers.

Afin de mieux caractériser phénotypiquement les souches des patients CCMC, nous avons réalisé un screening *in vitro* de leur sensibilité à différents stress (oxydatif, cationique, osmotique, thermique ou impactant l'intégrité de la paroi cellulaire) pertinents au regard des conditions environnementales qu'elles peuvent rencontrer au sein de cet environnement pathologique particulier que représente la CCMC. Les différentes conditions de stress testées sont détaillées dans le **Tableau 7**.

Les 37 souches de patients CCMC, les 144 souches contrôles ainsi que la souche SC5314, ont été préalablement cultivées 48 heures à 30°C en milieu liquide YPD à partir des stocks conservés à -80°C. Les cultures ont ensuite été diluées à une $DO_{600}=1$ dans de l'eau stérile puis inoculées en milieux solide plus ou moins additionnés des différentes molécules grâce au robot RoToR de chez Singer Instruments. Les plaques ont ensuite été incubées aux différentes températures indiquées pendant 7 jours.

Les conditions et concentrations informatives ont été déterminées dans un premier temps à J1, J3 et J7 de croissance. Les plaques ont été photographiées et la taille des colonies déterminée grâce au logiciel SGA (Wagih et al., 2013).

Après avoir sélectionné les conditions et concentrations discriminantes, les autres séries ont été réalisées au moins en dupliquât. Cette fois-ci les plaques ont été photographiées à l'aide du PhenoBooth de chez Singer Instruments avec une haute définition (4128 × 3096). C'est ensuite l'outil IRIS récemment publié par Kritikos *et al.* qui a été utilisé pour l'analyse de la taille des colonies (Kritikos et al., 2017). Celui-ci génère pour chaque colonie une valeur correspondant à un score de taille en pixel.

Pour comparer les expériences entre-elles, ainsi que les souches issues des différents patients entre eux, les valeurs obtenues pour la taille des colonies en conditions de stress ont été normalisées afin de s'affranchir d'un certain nombre de biais. Ces biais peuvent être liés aux caractéristiques intrinsèques de la souche (défaut de croissance en conditions standards) ou expérimentaux, incluant des variations dans les conditions de croissances des souches entre les différentes plaques ou au sein d'une plaque. En effet, entre différentes séries ou en fonction de l'emplacement de la colonie sur la plaque, la durée d'incubation, l'oxygénation ou la disponibilité locale des nutriments peut varier en rapport avec des variations stériques ou de consommation des nutriments d'une souche voisine. Ainsi, pour quantifier avec précision la croissance des colonies, ces variations sont corrigées par une double étape de normalisation. Premièrement, la taille de chaque colonie obtenue en condition de stress est normalisée par la taille de la même souche en condition standard (sans stress) puis celle-ci a été normalisée avec la valeur obtenue pour la souche SC5314 disposée au sein de chaque plaque.

Enfin, une confirmation de la réponse au stress de certaines souches a été effectuée en réalisant des dilutions successives.

Dans ce contexte, les cellules ont été cultivées 48h en YPD, quantifiées puis diluées à une concentration initiale de $DO_{600} = 0,1$ puis de 10 en 10 jusqu'à une $DO_{600} = 10^{-5}$.

Les résultats obtenus pour les 37 souches sont présentés dans partie C de la **Figure 22**. Celle-ci correspond à une classification (clustering) des données, réalisée sur les conditions testées grâce à l'outil Cluster 3.0 (de Hoon et al., 2004) en utilisant un clustering hiérarchique en lien complet (complete linkage clustering) avec une corrélation non paramétrique de rang de Spearman pour le calcul des distances. Le dendrogramme représenté dans la **Figure 22** a été obtenu avec le logiciel Java treeview (Saldanha, 2004). Les valeurs sont exprimées en \log_2 . En rouge sont représentées les souches plus sensibles que SC5314 (score normalisé $< 0,8$) et en vert les souches plus résistantes (score normalisé $> 1,2$).

A nouveau une importante diversité en termes de réponse aux différents stress a été observée entre les souches provenant des patients CCMC, mais également au sein des familles et des individus eux mêmes.

Aucun phénotype commun à toutes les souches provenant des patients CCMC n'a pu être mis en évidence avec les conditions testées. En revanche plusieurs conditions, dont deux jouant sur l'intégrité de la paroi fongique, ont semblé être particulièrement intéressantes pour l'étude de l'évolution phénotypique des souches au cours du temps. Ces conditions sont le rouge Congo (CR), la caféine et le chlorure de lithium (LiCl) (**Figure 22, partie C, Figure 23**). En effet, une majorité des souches (70,3% à 30°C et 83,8 % à 37°C) se sont avérées être impactées par l'usage du CR. Ces souches proviennent essentiellement de 8 patients (P1, P3, P4, P5, P7, P8, P9 et P12). La vérification de ce phénotype a été réalisée pour la dernière souche isolée de chaque patient avec des dilutions successives (**Figure 23**). Ceci nous a permis de confirmer ce résultat sauf pour la souche provenant du patient 5 qui ne semble pas présenter de défaut de croissance lors de cette analyse complémentaire.

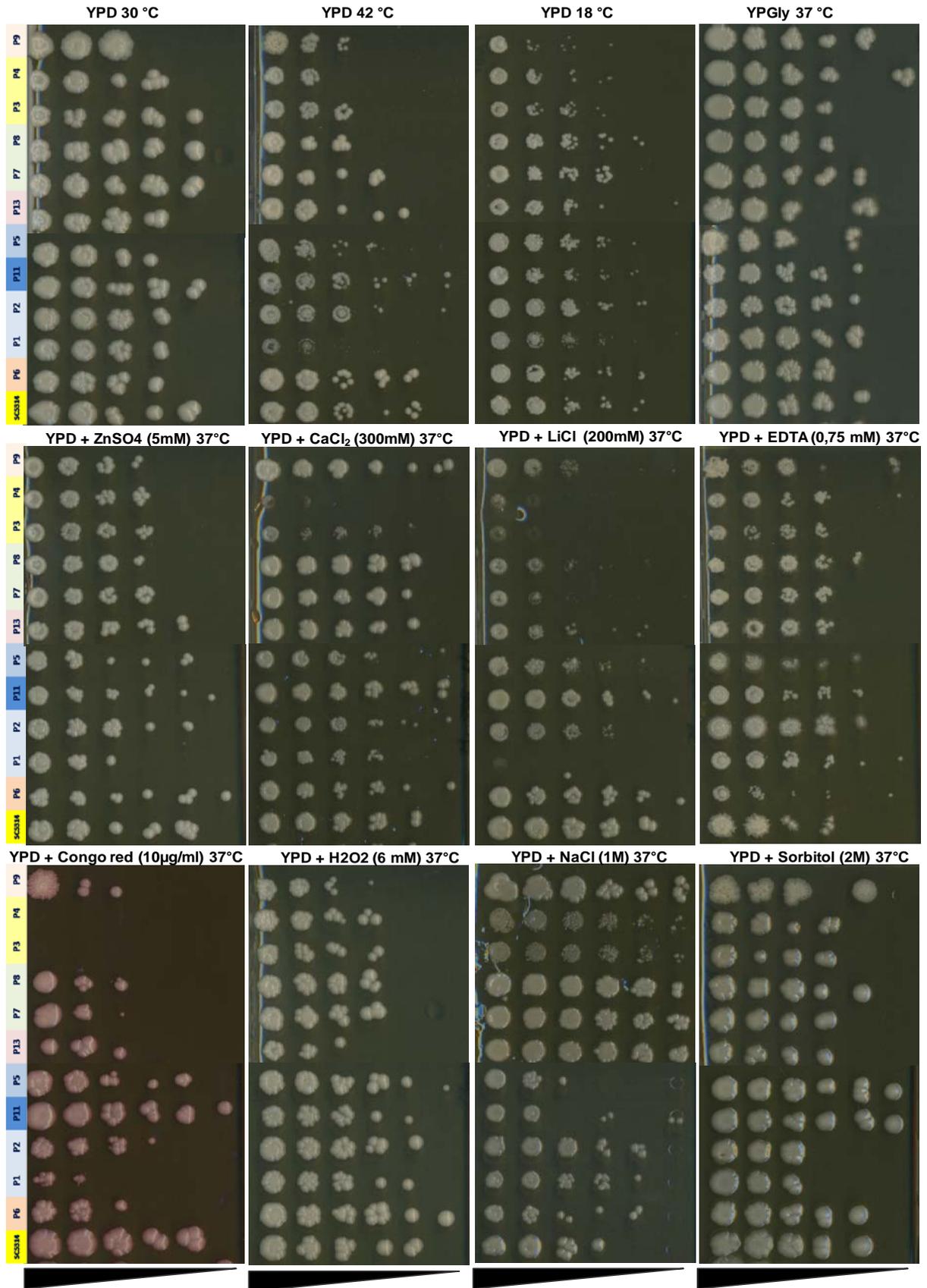


Figure 23 : Impact de l'exposition à différents stress sur les souches les plus récentes des patients atteints de CCMC.

(n=12, les patients P10 et P12 ont été inclus postérieurement à cette analyse). Douze conditions ont été testées en réalisant des dépôts des dilutions successives (de gauche à droite, $DO_{600} = 0,1$ à 10^{-5}) pour chacune des 12 souches de patients CCMC ainsi que la souche contrôle SC5314.

De façon intéressante nous avons observé le même profil de sensibilité de ces souches lors de leur exposition au chlorure de lithium (LiCl) **Figure 23**).

La deuxième molécule induisant un phénotype particulier en rapport avec l'intégrité de la paroi fongique, concerne la caféine. En effet, à l'opposé du CR, les souches provenant de 10/13 des patients (excepté P7, P8 et P13) ont globalement montré une meilleure résistance à cette molécule que la souche SC5314 (**Figure 22, partie C, Figure 24**). Et de façon très intéressante, on a observé une évolution de la résistance des souches à cette molécule au cours du temps chez l'ensemble des patients pour lesquels plusieurs souches ont été analysées chronologiquement (**Figure 22, partie C**). L'exemple le plus flagrant concerne le patient P1 avec une augmentation d'un facteur 3 du score normalisé (1,04 à 3,18) entre la souche de 2007 et celle de 2014. La vérification de ce phénotype a été réalisée pour la dernière souche isolée de chaque patient avec des dilutions successives (**Figure 24**). Ceci a permis de confirmer ce résultat sauf pour la souche provenant du patient 6 qui ne semble pas présenter d'augmentation de la résistance à la caféine dans cette analyse.

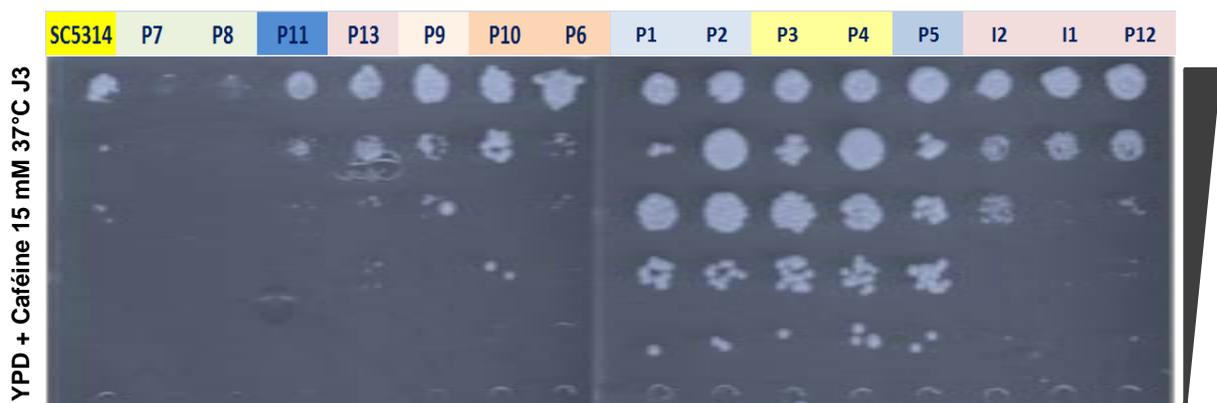


Figure 24 : Impact de la caféine à 15 mM sur les 13 souches les plus récentes des patients CCMC, les 2 souches d'individus sains, ainsi que la souche contrôle SC5314. Les souches des patients 7 et 8 sont sensibles à cette molécule. Les souches des patients 1 à 5 présentent un haut niveau de résistance à la caféine.

D'autres molécules, ayant un impact sur l'intégrité de la paroi fongique, ont été testées comme le Calcofluor White (CW) ou le SDS, mais ces molécules n'ont montré qu'un faible impact sur ces souches (**Figure 22, partie C**).

D'autres modifications phénotypiques ont pu être observées dans les séries chronologiques de souches (**Figure 22, partie C**). Parmi les 6 patients, pour lesquels nous disposons de plusieurs souches, nous avons identifié chez 4 d'entre eux, (P1, P2, P9 et P13) d'importantes modifications phénotypiques entre les premières et les dernières souches isolées (**Figure 22, partie C**). Par exemple, chez le patient P1, la souche la plus récente isolée en 2014 (souche #FLOJ12) présente une altération de la croissance aux différentes températures ainsi qu'une augmentation de la résistance au CaCl_2 , au ZnSo_4 au CW, au SDS et à la caféine (**Figure 22, partie C et Figure 23**). De la même façon, chez le patient P2 (Fils du patient P1) la souche la plus récente a montré une diminution de la sensibilité à la ménadione, au CW au CR et au SDS ainsi qu'une nette augmentation de la résistance à la caféine.

Par ailleurs, chez 2 patients on a noté l'apparition puis l'élimination de souches présentant un phénotype particulier. Par exemple, chez le patient P2, la souche isolée au niveau de la peau en 2008 (souche #FLON17) a globalement montré une meilleure résistance aux stress avec les autres souches isolées successivement (**Figure 22, partie C**). De la même façon chez le patient P3, deux souches (souches #BLOV29 et #BLOV37) ont présenté un profil atypique et n'ont pas été conservées au cours du temps (**Figure 22, partie C**).

Ces résultats montrent l'existence de différents variants phénotypiques chez un même individu dont certains persistent et d'autres sont éliminés.

3.2.2.4 Caractéristiques morphologiques et capacités d'invasion et de lyse cellulaire

Comme nous l'avons vu au cours de l'introduction, *C. albicans* est une levure polymorphique capable de se multiplier, en fonction des stimuli environnementaux auxquels elle est soumise, sous forme « levure », unicellulaire, mais aussi sous différentes formes multicellulaires, allant du pseudo-filament au véritable filament mycélien (ou hyphe) (**Figure 25**).

C'est sous la forme « levure » que *C. albicans* se multiplie et persiste dans le tube digestif en cas de portage commensal. En revanche, au cours des phénomènes pathologiques associés aux infections des muqueuses, *C. albicans* colonise et pénètre les cellules épithéliales sous forme de filaments mycéliens induisant ou non une lyse cellulaire (Sudbery, 2011). Cette plasticité morphologique est une caractéristique importante de la pathogénicité de *C. albicans*.

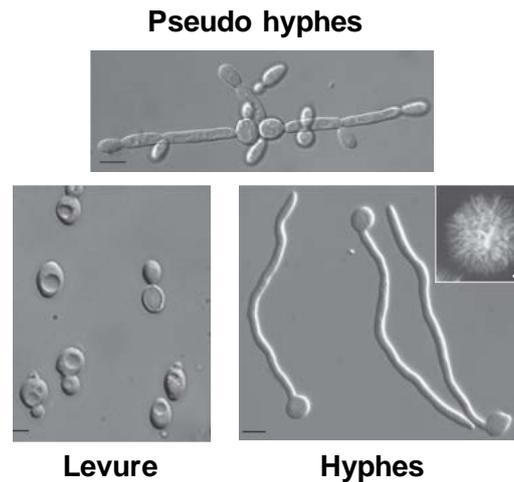


Figure 25 : Morphologie des formes levure, hyphe et pseudohyphe de *C. albicans*. Concernant la forme hyphe, l'encadré représente une colonie filamenteuse de *C. albicans* après 5 jours de culture sur milieu induisant la filamentation (Milieu Spider à 37°C) (d'après Sudbery *et al.* 2011).

De nombreux signaux environnementaux peuvent induire une croissance hyphale (comme par exemple la présence de sérum, un pH neutre, du CO₂ à 5%, l'hypoxie...), reflétant la diversité des microenvironnements que rencontre cette levure chez l'hôte.

En conditions de laboratoire, la croissance hyphale est classiquement induite par l'utilisation de milieux spécifiques, tels que le milieu « Lee » (qui contient un mélange d'acides aminés), le milieu « Spider » (milieu semi-synthétique à base de mannitol comme source de carbone) ou l'addition de sérum dans le milieu YPD. En règle générale, la croissance des filaments nécessite une température de 37°C (Sudbery, 2011).

Expérimentalement, deux étapes de la croissance hyphale peuvent être testées. La première correspond à la capacité de *C. albicans* à initier la filamentation. Cette étape correspond à la production du tube germinatif, suivit de la formation du filament à partir du stade levure. Elle s'effectue en milieu liquide à 37°C en quelques heures de conditions inductrices. La deuxième étape correspond à la capacité de *C. albicans* à maintenir une croissance hyphale, ce qui nous intéresse particulièrement dans notre modèle de patients CCMC. Celle-ci s'effectue en milieu solide (Spider ou YPD sérum), en évaluant la morphologie des colonies après 5 à 7 jours de culture. La formation des hyphes sur milieu solide se manifeste par des excroissances (plus ou moins nombreuses et longues) de la colonie principale. Une modification de l'aspect de la colonie peut également être observée, allant d'une colonie lisse à crénelée voire froissée.

Les colonies filamenteuses matures peuvent également envahir la gélose. Ce phénomène peut être testé en examinant la persistance des cellules dans la gélose après une culture sur milieu solide suivi du lavage de la surface de la gélose.

Dans cette étude, nous avons donc testé différentes conditions expérimentales permettant d'induire ou non une filamentation des souches en milieux solides, tels que le milieu Spider et le milieu YPD additionné de 10% de sérum de veau fœtal, ainsi qu'une condition de contrôle en YPD à 30°C et 37°C. Nous avons également testé la capacité des souches à envahir la gélose à 30 et 37 °C en milieu YPD et SD. Les différentes conditions d'induction de la filamentation testées sont résumées dans le **Tableau 7**.

Afin de comparer les souches entre-elles, un score a été appliqué à chaque colonie afin de distinguer les différents niveaux de modifications morphologiques (filamentation et aspect) et d'invasion de la gélose. La grille de score appliqué à l'analyse de la morphologie des colonies est représentée ci-dessous (**Figure 26**).

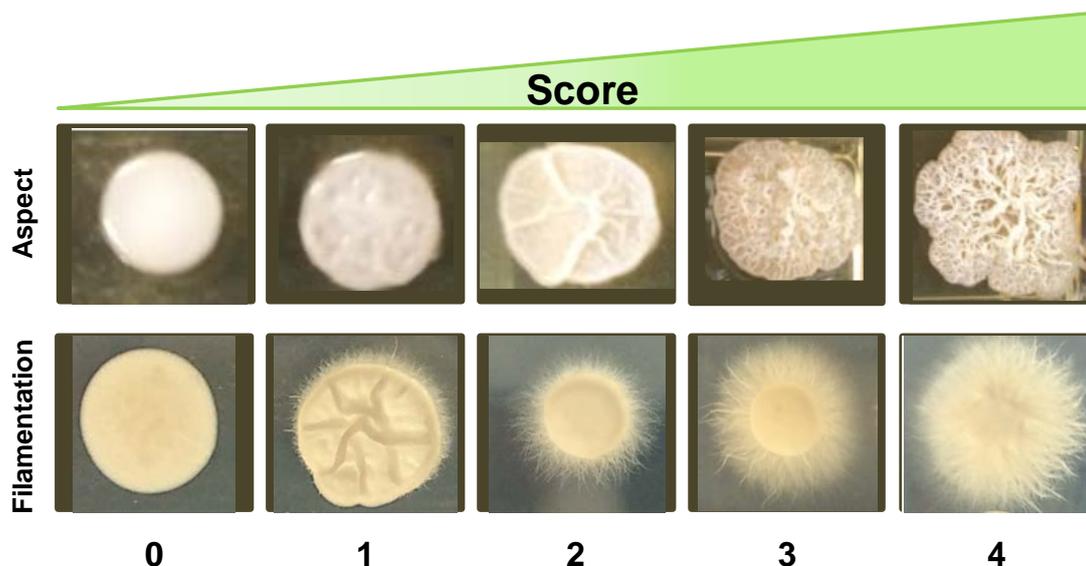


Figure 26 : Score appliqué à l'analyse de la morphologie des colonies (capacité de filamentation et aspect).

Les différentes conditions ont été appliquées aux 63 souches de la collection CCMC. Seuls les résultats obtenus pour les 37 souches sélectionnées dans les conditions les plus informatives (YPD 30°C et Spider 37°C pour la filamentation et l'aspect et YPD 37°C pour l'envahissement de la gélose) sont représentées dans les **parties A et C de la Figure 27**.

- Capacité de filamentation des souches

En condition non inductrice (YPD à 30°C) le score de filamentation devrait être $< 1,5$, les valeurs supérieures à ce score sont observées chez des souches présentant un phénotype anormal pouvant correspondre à une filamentation constitutive.

Au total, 21,6 % ($n=8/37$) des souches testées avaient un score anormal $\geq 1,5$. Ces souches provenaient de 3 patients (2 patients (P1 et P2) membres d'une même famille et le patient P5). A noter un score très élevé pour les souches du patient P2 (score 3 à 4, ($n=5$)) correspondant à un phénotype d'hyper filamentation (**Figure 27, partie C**, photos encadrées en vert). A titre de comparaison, seulement 11,8 % ($n=17/144$) des souches de la collection contrôle étaient capables de fil, amener en condition non inductrice, parmi lesquelles seules 2 souches avaient un score élevé de 3. L'une d'entre elles a été isolée au cours d'une infection invasive (CEC2020) et l'autre au cours du commensalisme (CEC3619).

Différentes conditions inductrices de la filamentation ont été testées (Milieux Spider et sérum à 30 et 37°C). De façon surprenante, 56,8% des souches (n=19) se sont révélées incapables de filamenter dans ces conditions et correspondaient aux souches provenant des patients P4, P6, P7, P8, P10 et P11 (**Figure 27, parties A et C**). Cependant, la proportion de souches incapables de filamenter dans ces conditions était comparable à celle observée pour les souches contrôles (64,6%, n=93).

Concernant l'analyse des séries de souches chronologiques chez nos patients CCMC, 4 (P3, P5, P9 et P13) ont présenté des souches ayant subi une importante modification phénotypique ou « switch phénotypique » au cours du temps (**Figure 27, parties A et C**). Chez ces patients, l'isolat le plus récent a présenté une meilleure capacité de filamentation en comparaison aux isolats antérieurs. Par exemple, chez le patient P3, les 8 souches isolées entre 2008 et 2014 étaient incapables de filamenter tandis que la souche isolée en 2015 a présenté un score moyen de 2. De la même façon, chez le patient P9, nous avons observé un important « switch phénotypique » entre les souches isolées en 2009/2010 et celles de 2012/2013.

A noter, qu'une souche a présenté un phénotype très particulier (souche # FLON17) (**Figure 27, partie C**). Celle-ci a montré : i. la disparition de sa capacité à filamenter en présence de sérum à 37°C (score moyen de 0). ii. la modification de son aspect morphologique (score 4, en Spider à 37°C) et, iii. une filamentation de type « aérienne » en Sérum à 30°C.

Ce phénotype n'a pas été observé chez d'autres isolats du même patient suggérant l'élimination de celui-ci au cours du temps.

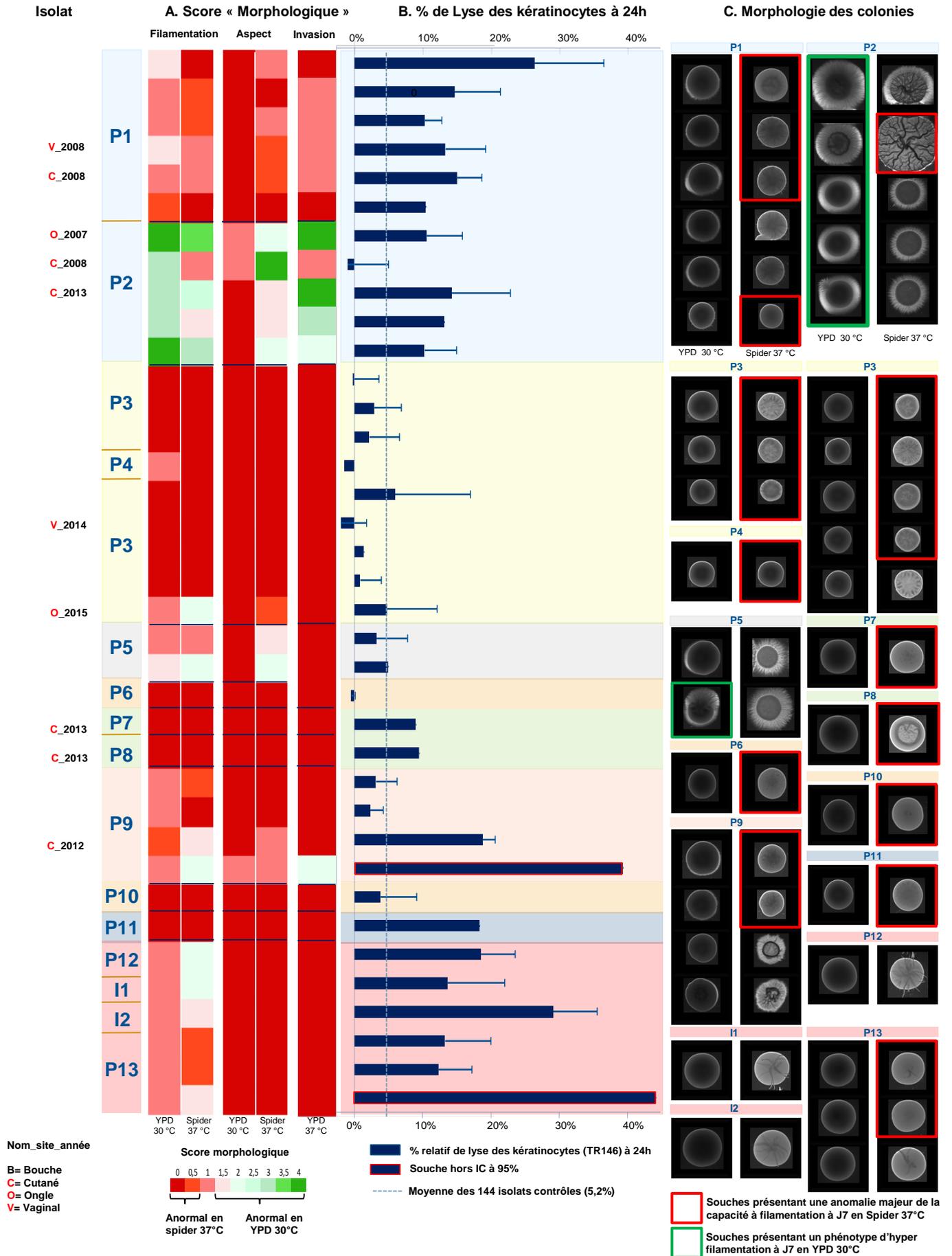


Figure 27 : Caractérisation morphologique et capacité de lyse cellulaire des 37 souches provenant de patients CCMC et de leur famille. A. Score morphologique. La moyenne des scores obtenue pour chacune des 37 souches est représentée sous forme d'échelle de couleur. Les résultats de différentes conditions testées sont représentés incluant le YPD à 30°C et le Spider à 37°C pour la filamentation et l'aspect des colonies ainsi que le YPD à 37°C pour l'envahissement de la gélose. **B. Capacité de lyse des Kératinocytes humains.** Le dommage causé aux cellules épithéliales TR146 a été évalué après 24h d'infection avec les différentes souches en quantifiant la libération de la lactate déshydrogénase (LDH) par ces cellules. Les résultats, représentés sous forme d'histogramme, sont exprimés en pourcentage de cellules lysées relatif à un témoin positif (lyse à 100 %, correspondant au traitement des cellules épithéliales avec 1% de Triton-X-100) et un témoin négatif (correspondant à la LDH libéré par les cellules non infectées). La moyenne obtenue pour les 144 souches contrôles est représentée en pointillés. Les résultats des souches de patients CCMC sortant de l'intervalle de confiance à 95 % (IC95 = 5,2 +/- 24,9) sont représentés avec un cadre rouge. **C. Représentation photographique du niveau de filamentation et de l'aspect des colonies.** Les colonies ont été photographiées après 7 jours de culture en YPD 30°C et spider 37°C. Les photos des colonies présentant un défaut de filamentation sont encadrées en rouge tandis que les colonies présentant un phénotype d'hyperfilamentation sont représentées en vert.

- Capacité des souches à envahir la gélose

Nous avons également testé, comme décrit ci-dessus, la capacité des souches à envahir la gélose (en YPD et SD à 30 et 37°C).

Dans les conditions testées, la majorité des souches (72,9 % ; 27/37) n'a pas montré d'envahissement de la gélose (**Figure 27, partie A**). En revanche, en parallèle avec l'importante filamentation des souches provenant du patient P2, celles-ci ont présenté un score élevé d'invasion (score 2 à 4), comparable à la souche SC5314. Chez ce même patient, la souche #FLON17, isolée en 2008 au niveau de la peau, s'est révélée incapable d'envahir la gélose à la différence des autres souches isolées en 2013 et 2015.

Par ailleurs, comme précédemment, une modification phénotypique a été observée avec la souche la plus récente isolée chez l'individu P9 en 2013. Celle-ci présentait un score d'invasion de 2 tandis que les souches précédentes étaient à 0, montrant à nouveau un phénomène d'évolution phénotypique des souches au cours de l'interaction avec l'hôte. Ces résultats n'ont été obtenus qu'une seule fois et sont donc à confirmer.

- Capacité des souches à lyser les kératinocytes humains

Grâce à une collaboration avec l'équipe de Salomé Leibundgut-Landmann de l'Université de Zurich nous avons déterminé la capacité des souches provenant des patients atteints de CCMC et de leur famille à lyser les kératinocytes humains.

En effet, une étude récente, réalisée par Schönherr *et al.* en collaboration avec notre équipe, a montré, dans un modèle d'infection oropharyngée chez la souris, que les souches de *C. albicans* présentaient des différences en termes de capacité à lyser les kératinocytes et de persister au niveau de la muqueuse buccale (Schönherr et al., 2017). Par ailleurs, il a été démontré une corrélation inverse entre la capacité d'une souche à persister au niveau de la muqueuse buccale et le degré d'activation de la réponse immunitaire de l'hôte (Schönherr et al., 2017). L'accumulation des neutrophiles et l'induction de l'inflammation dans les tissus infectés étaient variables selon la souche de *C. albicans* testée et en rapport avec la capacité de celle-ci à causer une lyse cellulaire.

Le dommage causé aux cellules épithéliales au cours de la co-culture *in vitro* entre *C. albicans* et les kératinocytes humains peut être évalué par la libération de la lactate déshydrogénase (LDH). Cette mesure semble être un bon reflet du niveau d'inflammation de la muqueuse buccale retrouvé *in vivo* chez la souris et donc de la pathogénicité des souches. Cependant, dans cette étude, aucun lien n'a été retrouvé entre la capacité de lyse des cellules et la capacité des souches à filamenter (Schönherr et al., 2017).

Au vu de la gravité des infections candidosiques dont souffrent les patients atteints de CCMC nous nous sommes demandés si celle-ci était uniquement dû au déficit immunitaire dans la voie de l'IL-17 de ces patients ou si les souches présentaient en plus un phénotype adaptatif particulier, leurs permettant soit de persister chez ces patients via une activation diminuée du système immunitaire par une diminution de la lyse cellulaire, soit à l'inverse une augmentation de la capacité de lyse des kératinocytes induisant une inflammation majeure responsable de la gravité des infections observées chez ces patients.

Pour cela, la capacité à lyser les cellules épithéliales des 37 souches de patients CCMC et de leur famille a été testée en utilisant des kératinocytes humains de la lignée TR 146.

Le dommage des cellules épithéliales TR146 a été évalué après 24h d'infection avec les différentes souches en quantifiant la libération de LDH par ces cellules.

Les résultats ont été exprimés en pourcentage de cellules lysées relatif à un témoin positif (lyse à 100 %, correspondant au traitement des cellules épithéliales avec 1% de Triton-X-100) et un témoin négatif (correspondant au LDH libéré par les cellules non infectées). A titre de comparaison, deux souches contrôles ont été rajoutées, la souche SC5314 induisant un dommage élevé aux cellules épithéliales et la souche clinique nommée « 101 » de *C. albicans* induisant un dommage faible aux cellules. Les résultats des 37 souches de patients CCMC et de leur famille sont représentés dans la **partie B de la Figure 27**.

En moyenne, pour les 37 souches, le pourcentage relatif de lyse des kératinocytes (TR146) était de 10,7 % (allant de -2 % à 44,1 %) avec à nouveau une importante diversité dans la capacité de lyse cellulaire entre les souches.

Parmi ces souches, 51,4 % (n=19/37) ont présenté un niveau élevé de lyse des kératinocytes avec un seuil arbitraire fixé à 10%. A titre de comparaison, la moyenne retrouvée pour les 144 souches contrôles était de 5,2 % (allant de -6% à 63,7%) avec seulement 21,5% (31/144) des souches induisant un niveau élevé de lyse.

Nous avons également observé des différences phénotypiques entre les familles et au sein des patients. En effet, les souches présentant un faible niveau de lyse (< 10%) sont globalement retrouvées chez les patients P3 à P8 et P10, tandis que les souches induisant un niveau élevé de lyse provenaient des patients P1, P2, P11, P12 et P13 (**Figure 27**)

Un « switch phénotypique » a été observé chez les souches de 3 patients (P2, P9 et P13), parmi lesquelles deux ont été maintenues au cours du temps (P9, souche #BOIN34, #BOIN06 et P13, souche #ARIS60) et une correspondait à une souche unique (P2, souche # FLON17) non maintenue au cours du temps.

A nouveau ces résultats confirment que les souches peuvent évoluer phénotypiquement au cours du temps dans le cadre d'une interaction pathologique à long terme chez l'homme. Ils montrent également que les patients peuvent être colonisés avec des souches présentant un phénotype complètement différent. Soit ces souches constituent des variants phénotypiques minoritaires au sein de la population cellulaire de *C. albicans*, soit elles présentent un phénotype défavorable, non conservé chez les patients. Ces résultats suggèrent qu'il existe une importante dynamique dans l'évolution des souches chez l'hôte. Ces résultats constituent donc une piste intéressante pour l'étude de l'évolution génomique des souches de *C. albicans* dans notre modèle de patients CCMC que nous allons voir maintenant.

3.2.3 Caractérisation génomique des souches de patients atteints de CCMC et de leur famille

En parallèle de la caractérisation phénotypique des souches, nous avons réalisé le séquençage du génome complet (séquençage à haut débit) des 37 souches sélectionnées de *C. albicans* provenant des patients CCMC (n=35) et des membres sains de leur famille (n=2) (**Tableau 6**).

Le lien phylogénétique, obtenu grâce aux données de séquençage du génome complet des 37 souches sélectionnées est représenté dans la **Figure 28**.

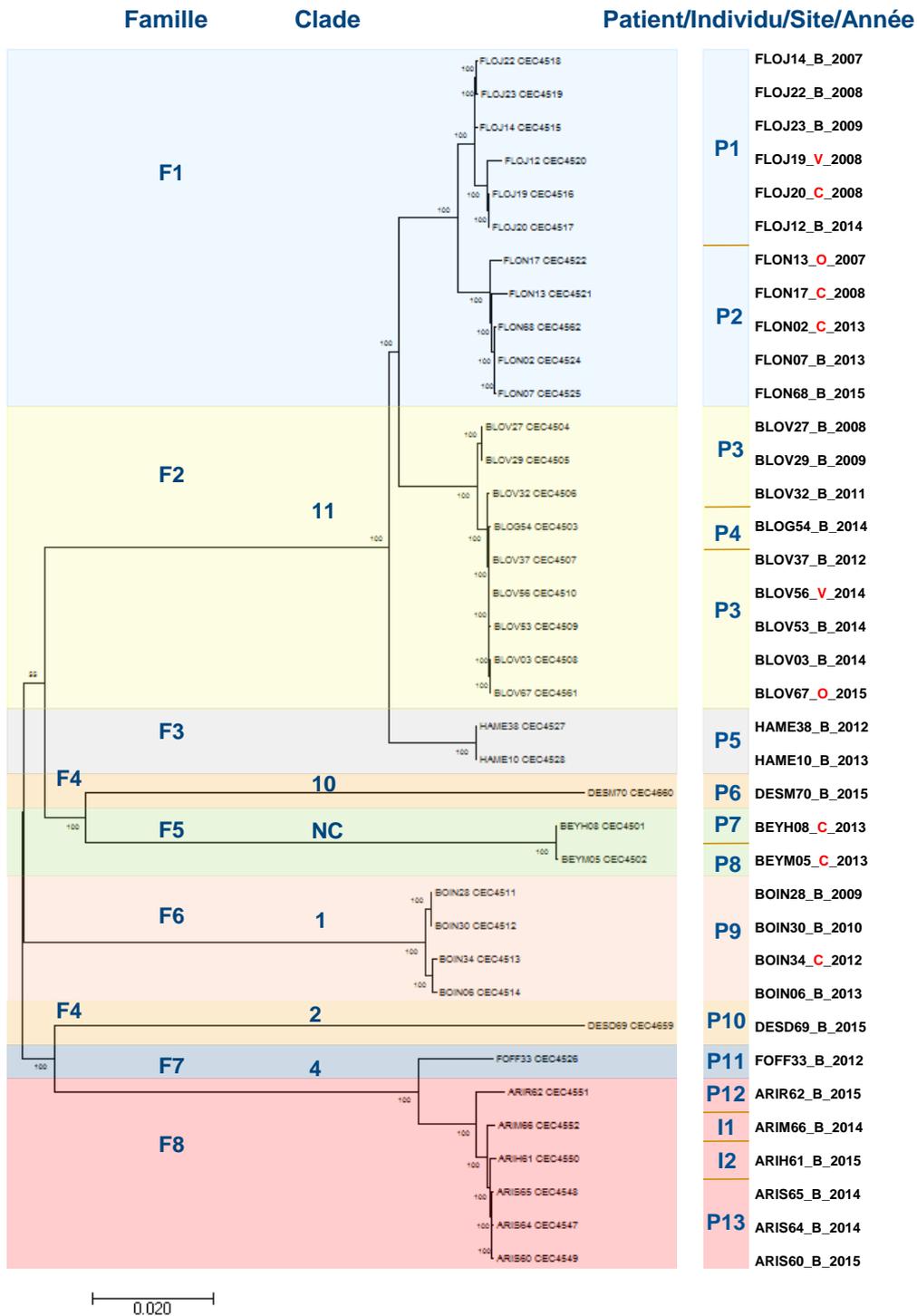


Figure 28: Arbre phylogénétique réalisé à partir des données de séquençage du génome complet des 37 souches de patients CCMC et de leur famille. Celui-ci a été réalisé grâce au logiciel MEGA7 (Kumar et al., 2016), en utilisant la méthode de Neighbor-Joining (Saitou and Nei, 1987). L'analyse de la robustesse a été déterminée sur 1000 bootstrap. Les distances évolutives ont été calculées à l'aide de la méthode du maximum de vraisemblance composite (Tamura et al., 2004). L'analyse a été réalisée sur 37 séquences nucléotidiques comprenant un total de 1 054 120 positions polymorphiques dans l'ensemble de données finales. (Origine des souches, B=Bouche, V= Vagin, C= Cutanée, O=Ongle).

Grâce à ces données, nous nous sommes intéressés aux remaniements génomiques tels que les variations du nombre de copies des gènes (CNVs) ainsi que les pertes d'hétérozygotie (LOH) mais également aux polymorphismes nucléotidiques (SNPs).

Nous avons utilisé la version A22-s06-m01-r01 de la séquence de référence de la souche SC5314 (Assemblage 22) pour l'analyse des génomes des 37 souches de *C. albicans*. La profondeur de séquençage était en moyenne de 126x (47x-313x) et le nombre moyen de positions polymorphes était de 124167 SNPs par génome. Les principales caractéristiques des données de séquençage sont résumées dans le **Tableau 8**.

	Nombre de séquences	Profondeur de séquençage	Couverture du génome	Nombre total de SNPs	SNPs hétérozygotes	SNPs homozygotes
Moyenne	9 315 874	126 X	99.5 %	124 167	75 803	48 130
Range	2 943 104-19 788 044	46-313	99-99,6	73 278-147 349	18 370-63 729	52 212-96 705

Tableau 8 : Résumé des principales caractéristiques des données de séquençages obtenues, en comparaison avec la souche de référence SC5314, pour les 37 génomes de patients CCMC et leur famille.

Les données obtenues ont ensuite été utilisées pour comparer les génomes (i) des souches de patients CCMC et des souches contrôles, (ii) des souches isolées chronologiquement chez un même patient CCMC et (iii) des souches isolées chez des individus sains et des malades d'une même famille qui partageaient le même génotype MLST.

3.2.3.1 Description des remaniements affectant le génome des souches isolées chez les patients CCMC et leur famille en comparaison avec les souches contrôles.

- Remaniements génomiques de type aneuploïdie et LOH

- **Aneuploïdies**

Dans notre collection de souches isolées chez les patients CCMC et leur famille, nous n'avons pas détecté d'événements d'aneuploïdies (comme des duplications de chromosomes entiers ou de segment de chromosomes).

Ces données sont concordantes avec les résultats que nous avons obtenus par l'analyse du contenu génique réalisé par cytométrie de flux (FACS). Cependant, ces résultats sont surprenants car dans notre collection, 59,5% (22/37) des souches étaient résistantes aux antifongiques azolés, suggérant que ce type d'événement n'est pas retrouvé ou conservé au cours des traitements de longue durée chez nos patients atteints de CCMC.

- **LOH distaux : comparaison entre les deux collections de souches (CCMC et contrôles)**

Au sein des souches de patients CCMC, nous avons observé de grands événements de LOH de type BIR (Break Induced Replication) sur les régions distales des différents chromosomes. La taille de ces grands LOH était variable selon les souches. Les souches des patients P1 à P4 ainsi que P6 et P9 étaient les plus impactées par ces événements. A titre d'exemple, nous avons représenté la taille des LOH distaux observés dans les souches les plus récentes isolées chez chaque patient (n=13) (**Figure 29**).

De façon intéressante, nous avons pu constater que l'étendue des LOH distaux chez les souches de patients CCMC était en moyenne plus importante que les souches de la collection contrôle (**Figure 30**). Cette différence était due, en partie, à la présence de larges LOH dans les Chr1, Chr3 et Chr5 des souches des patients CCMC. Il n'y a pas de corrélation entre la quantification des LOH distaux observés et l'appartenance des souches à un clade spécifique ou à leur phénotype de résistance aux antifongiques azolés.

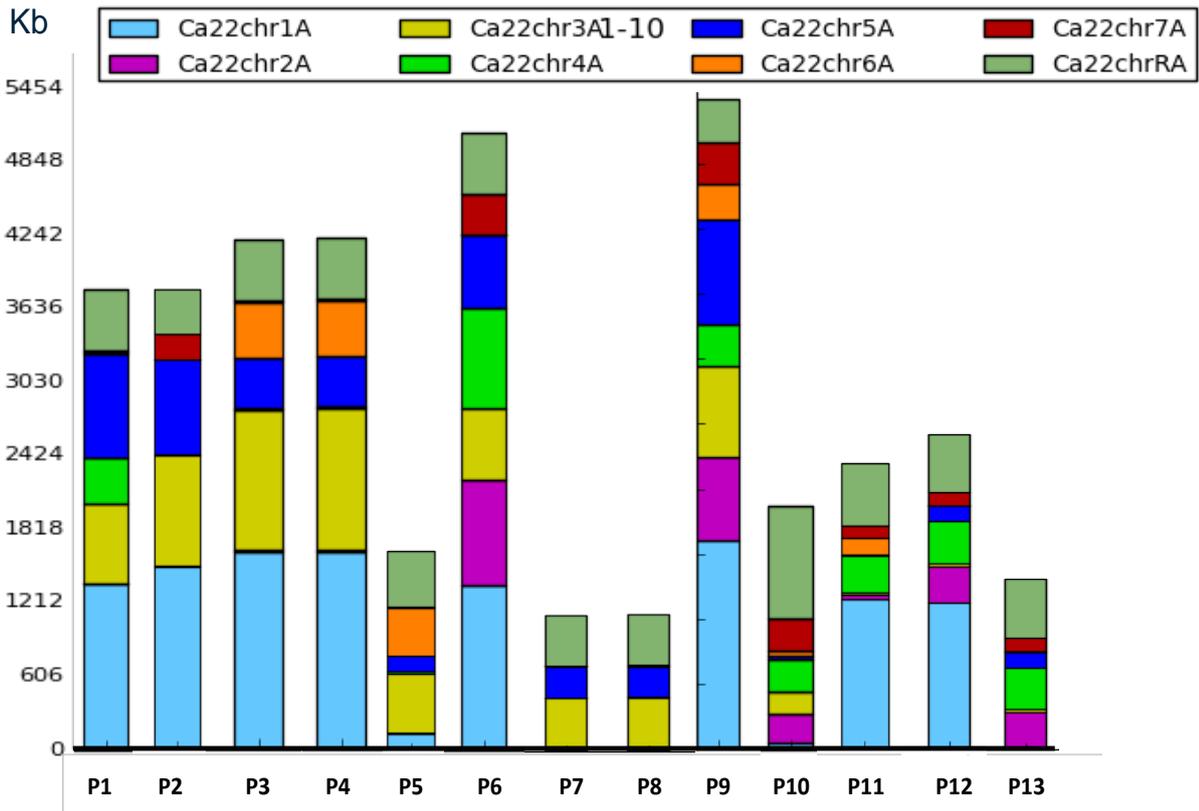


Figure 29 : Quantification en kb de la taille des événements de LOH distaux pour chacun des 8 chromosomes (représentés par les différentes couleurs). Ces événements sont calculés pour la souche la plus récente de chacun des 13 patients atteints de CCMC

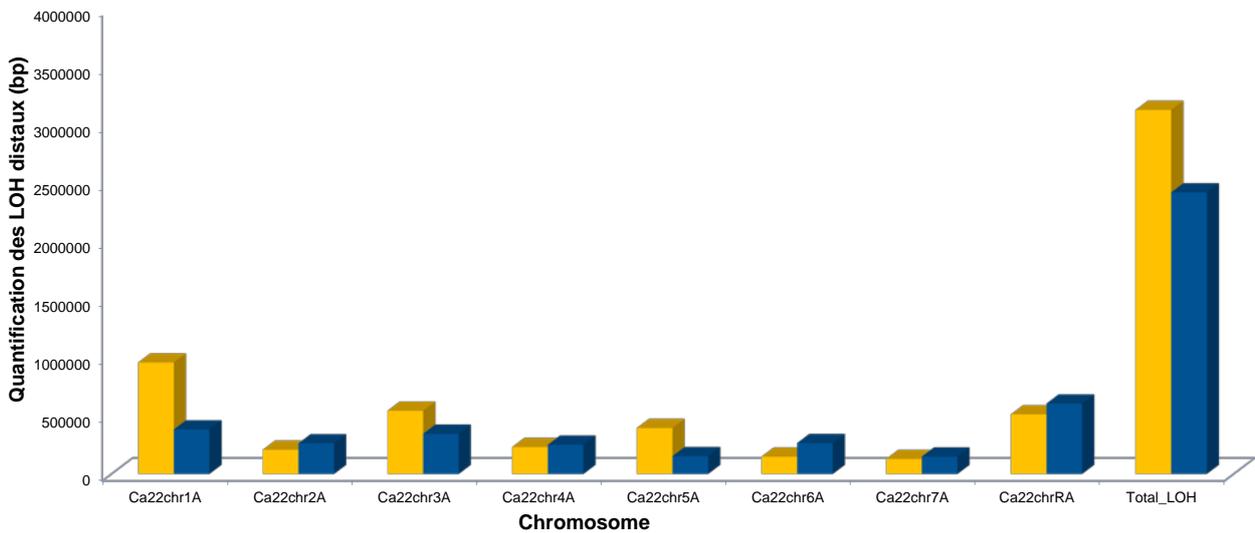


Figure 30 : Comparaison de la quantité moyenne (en pb) de LOH distaux détecté pour les 13 souches provenant de patients atteint de CCMC (en jaune) et les 144 souches contrôles (en bleu).

La comparaison des souches entre elles a montré que toutes les souches de patients CCMC présentaient un LOH au niveau du bras droit du ChrR et du Chr3 (**Figure 31**). Concernant l'homozygotie du bras droit du ChrR, celle-ci était connue et a été discutée dans la partie introductive de ce manuscrit tandis que le LOH au niveau du Chr3 été rapporté dans la souche SC5314. La comparaison de cette région (Chr3) avec les 144 souches contrôles a montré que ce LOH était également fréquent dans cette collection de souches.

Par ailleurs, des LOH distaux communs à plusieurs souches ont également été observés comme par exemple au niveau du bras droit du Chr1 (souches des patients P1-P4, P6, P9 et P11) ainsi qu'au niveau du bras gauche du Chr5 (souches des patients P1-P4 et P7-P9).

Sur le Chr5 se situent plusieurs gènes impliqués dans la résistance aux antifongiques azolés (*ERG11* et *TAC1*). Un LOH sur cette région peut conférer un avantage sélectif, pour les souches, en termes de réponse aux antifongiques azolés. Une corrélation entre la présence de cet événement et le niveau de résistance élevé des souches au fluconazole peut être observée pour les souches des patients P1 à P4. Concernant le LOH distal du Chr1, cet événement a également été rapporté, très récemment, par Forche *et al.* sur des souches de *C. albicans* isolées dans un modèle d'infection oropharyngée chez la souris (Forche *et al.*, 2018).

Une étude plus précise des gènes impactés par cet événement de LOH du Chr1 pourrait être intéressante dans le cadre de l'étude de l'évolution génomique de *C. albicans* au cours de l'interaction pathologique chez l'Homme.

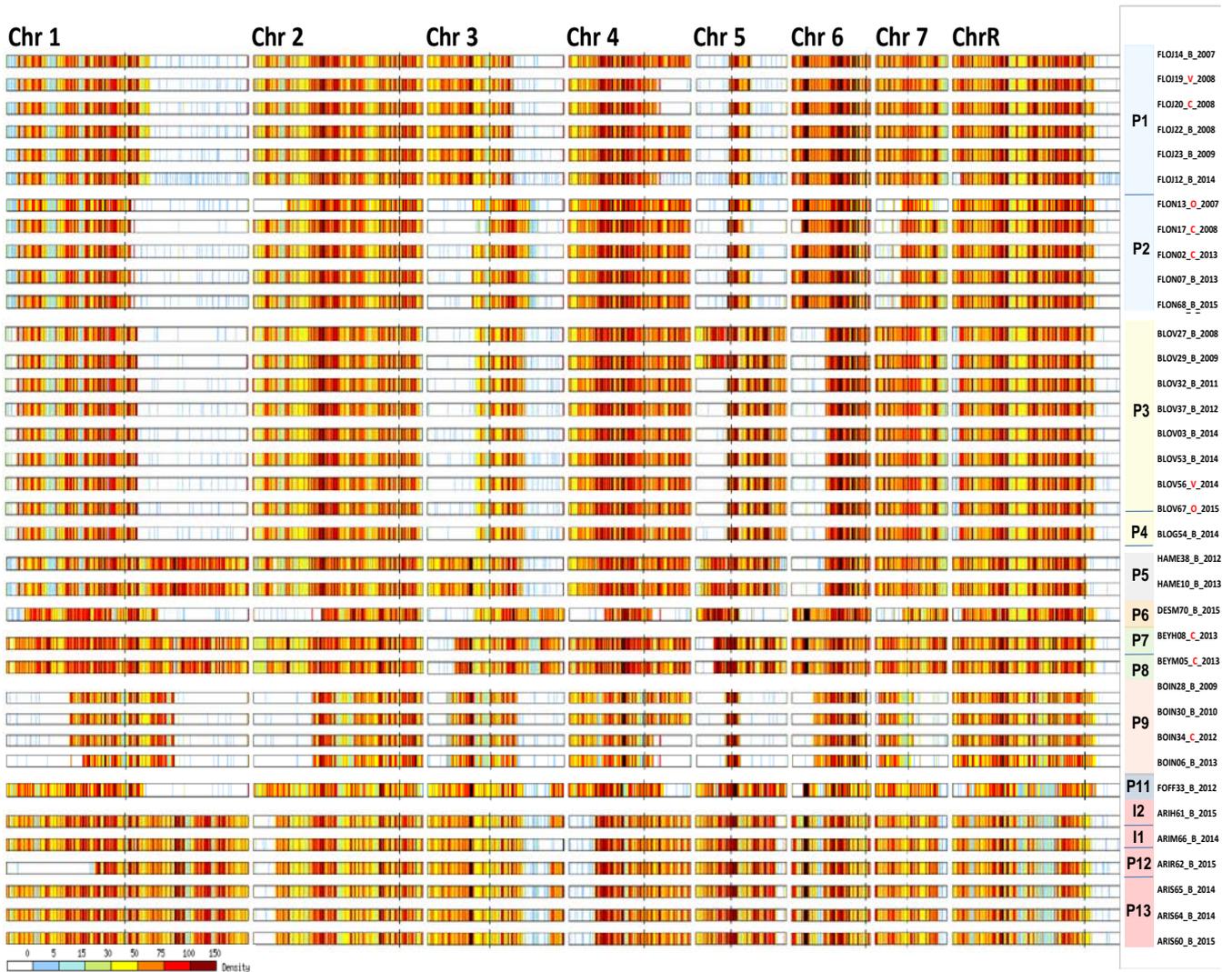


Figure 31 : Comparaison de la présence de grands événements de LOH de type BIR chez les souches de patients CCMC. Heatmaps représentant la densité de SNPs hétérozygotes sur les 8 chromosomes des différentes souches provenant des patients CCMC et de leur famille (n=36, la souche du patient P10 n'est pas représentée dans cette figure) (Fenêtre de 10 kb). Les zones blanches à foncées représentent respectivement des régions très homozygotes à très hétérozygotes.

- LOH distaux : analyse des séries de souches chronologiques chez les CCMC

Nous nous sommes ensuite intéressés aux événements de LOH survenant au cours du temps dans chaque série de souches chronologiques isolées chez les patients CCMC (**Figure 32**).

Plusieurs cas de figures peuvent se présenter, un LOH distal supplémentaire peut être identifié au sein d'une ou plusieurs souches d'une série chronologique. Ces souches peuvent soit persister soit disparaître, dans ce dernier cas elles n'ont été isolées que transitoirement chez le patient.

Par exemple, chez les patients P3 et P9, on note l'apparition d'un LOH (bras droit du Chr 5 et bras gauche du Chr4, respectivement) au cours du temps qui se fixe et persiste sur les isolats successifs. A l'opposé, chez le patient P1, le LOH du bras droit du Chr4 est retrouvé de façon transitoire sur des souches de différents sites d'isolement ou à différentes dates (**Figure 32**). Il en est de même pour les isolats du patient P2. Deux hypothèses peuvent être envisagées concernant la détection transitoire de LOH distaux au sein des souches chronologiques. Soit le LOH induit un désavantage sélectif (défaut de croissance de la souche) expliquant son élimination au cours du temps, soit celui-ci est présent de façon minoritaire au sein de la population cellulaire de *C. albicans* et n'est pas détecté lors de l'analyse d'une colonie unique pour chaque prélèvement. Ces résultats suggèrent qu'il existe également une hétérogénéité génétique de la population cellulaire de *C. albicans* au cours de l'interaction pathologique prolongée chez l'Homme.

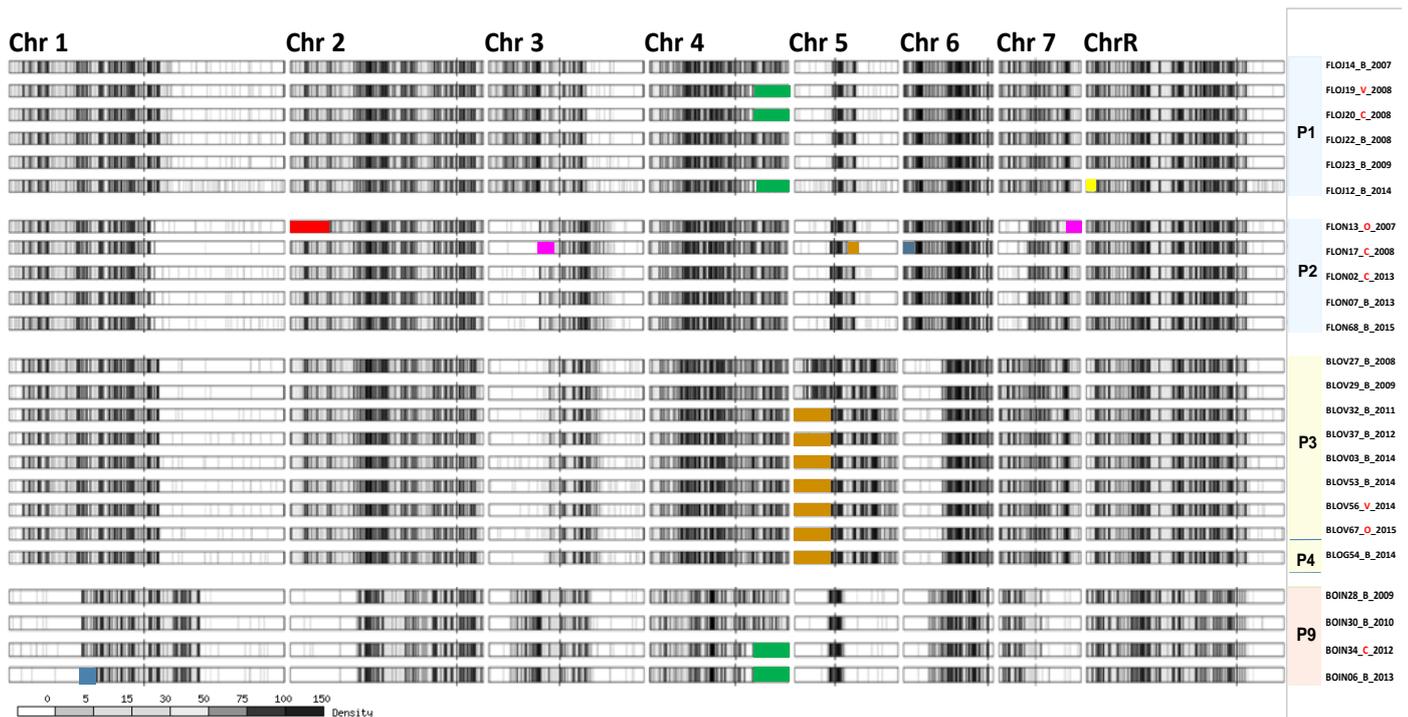


Figure 32 : Survenue d'événements de LOH au cours du temps sur les séries chronologiques de souches provenant des patients CCMC (P1 à P4 et P6). Dans ce contexte, les heatmaps, représentant la densité de SNPs hétérozygotes sur les 8 chromosomes, sont représentées en gradient de noir et blanc (Fenêtre de 10 kb). La survenue de nouveaux événements est figurée par un rectangle de couleur.

- Remaniements génomiques de type SNPs

Nous avons ensuite identifié les mutations (SNPs) situées dans des gènes pouvant être impliqués dans l'adaptation de *C. albicans* à son environnement.

Deux hypothèses, non mutuellement exclusives, peuvent être envisagées afin d'identifier les mutations les plus intéressantes.

Dans la première hypothèse, nous considérerons que les souches ont continué d'évoluer entre le premier et le dernier isolat étudiés chez les patients CCMC. Dans ce cas c'est la survenue de nouveaux SNPs qui nous intéresse (cette hypothèse sera explorée dans un paragraphe dédié à l'évolution des souches au cours du temps).

Dans la deuxième hypothèse, nous considérerons que les souches sont adaptées à leur environnement au moment du premier isolement. Dans ce contexte, ce sont les SNPs communs entre les génomes des souches des différents patients CCMC qui nous intéressent. Afin d'étudier cette deuxième hypothèse, nous avons identifié les SNPs communs aux génomes des souches CCMC et absents des génomes de souches contrôles (**Figure 33**).

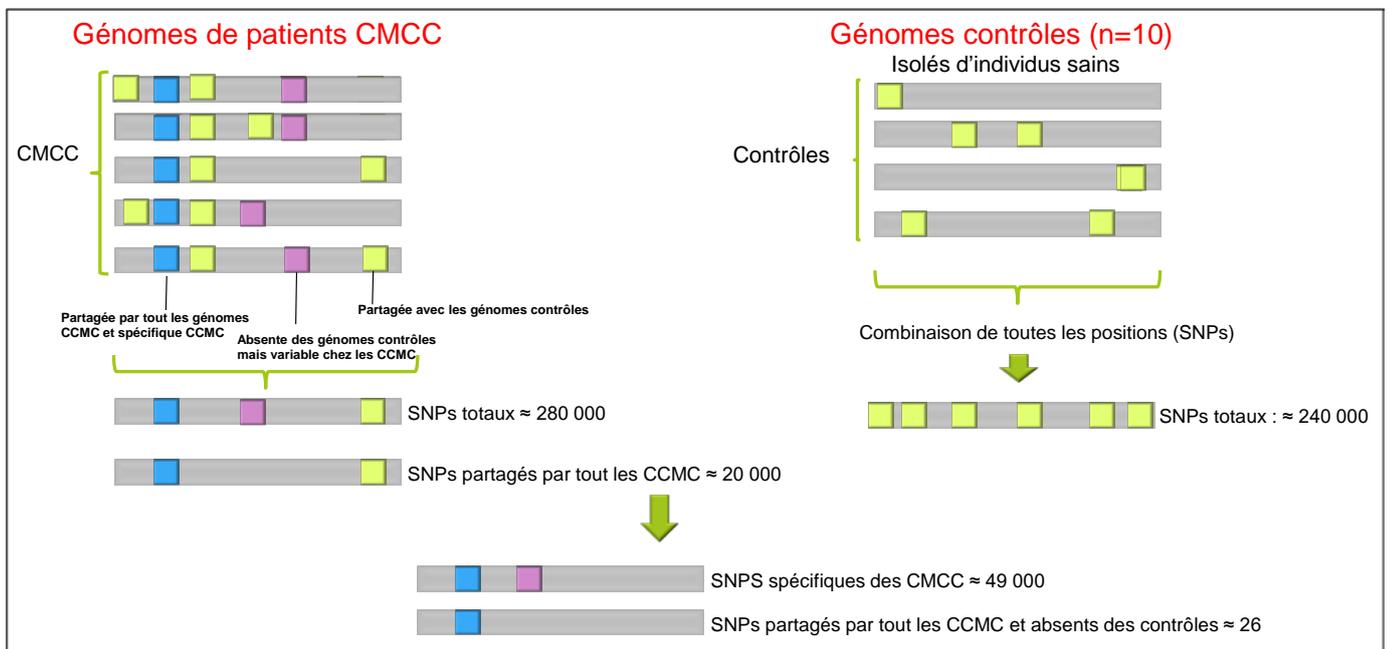


Figure 33 : Algorithme employé pour détecter les SNPs spécifiques aux génomes des souches isolées chez les patients atteints de CCMC (ici n=9) en comparaison avec des génomes de souches contrôles (n=10).

- Détection des SNPs communs aux génomes des souches de patients CCMC

L'approche décrite dans la **Figure 33** a permis d'identifier 26 SNPs d'intérêt, parmi lesquels, 9 étaient non synonymes. Ces 9 SNPs étaient localisés dans des gènes dont certains étaient non caractérisés chez *C. albicans*, tels que le gène *DAL5* (n=2 SNPs) codant pour une protéine transmembranaire de type perméase et le gène *CIRT4B* (n=2) codant pour une transposase et dans 5 gènes non annotés (*C3_03940C_A*; *C5_02090W_A*; *C6_02350C_A*; *C7_02600C_A*, *C7_03460W_A*).

Parmi ces gènes identifiés, le gène *DAL5* peut constituer une piste intéressante dans l'étude de l'adaptation génomique de *C. albicans* au cours de l'interaction pathologique de longue durée. En effet, les protéines transmembranaires sont souvent essentielles à la survie des cellules fongiques. Chez *S. cerevisiae*, *DAL5* constitue une sous-famille de gènes dont l'histoire évolutive complexe (duplications, délétions et transferts horizontaux de ces gènes), suggère une évolution particulière et une spécialisation (Hellborg et al., 2008). Sur le plan fonctionnel, ce gène est surexprimé chez *S. saccharomyces* en présence de caféine, ce qui semble indiquer qu'il est impliqué dans la voie mTOR (Kuranda et al., 2006).

Chez *C. albicans*, ce gène code pour une protéine transmembranaire de type perméase, dont la régulation dépend de la disponibilité en source d'azote elle-même directement régulée par le facteur de transcription Gat1 (Limjindaporn et al., 2003; Singh et al., 2011). Ce gène peut donc être intéressant à explorer, dans le cadre de notre étude, car il peut être directement ou indirectement impliqué dans les phénotypes observés chez nos souches de patients CCMC.

Nous avons également élargi notre sélection en examinant les SNPs communs à une majorité des génomes des souches de patients CCMC. Dans ce contexte, 209 mutations non-synonymes ont été identifiées impactant 50 gènes différents. Une analyse du lien ontologique entre ces gènes (GO, Gene Ontology) a été réalisée grâce à l'outil disponible sur CGD. Cette analyse permet de décrire les liens entre les gènes selon leurs fonctions moléculaires, les compartiments cellulaires et les processus biologiques auxquels ils appartiennent.

Parmi les 50 gènes identifiés, 17 ont un rôle dans la filamentation ($p=0,00059$) et la croissance des cellules fongiques ($p=0,00087$) (**Tableau 9**). Plusieurs gènes sont également impliqués dans l'adhésion cellulaire et la formation de biofilm. Ces processus biologiques sont essentiels dans le cadre de l'interaction hôte-pathogène.

Par exemple, parmi les gènes identifiés, nous nous sommes intéressés au gène *ECE1* (candidalysine). Ce gène code pour une toxine nommée candidalysine, récemment décrite comme étant impliquée dans la lyse cellulaires de *C. albicans* (Moyes et al., 2016).

Ce gène nous a particulièrement intéressés, au vu des « switch phénotypiques » (capacité de lyse des kératinocytes humains) observées entre les souches des différents patients et entre les séries de souches chronologiques. La séquence protéique du peptide III d'Ece1p (spécifiquement impliqué dans la lyse cellulaire) a donc été comparée entre les souches des différents patients CCMC. Bien que plusieurs profils protéiques aient été identifiés, aucune corrélation n'a pu être mise en évidence entre le pourcentage de lyse des kératinocytes par les souches et le profil protéique observé pour le peptide III d'Ece1p. Par ailleurs, aucune modification n'a été observée au cours du temps au sein des séries chronologiques. Ces résultats sont concordants avec la récente étude réalisée par Schönherr et al. ayant comparé, sur un grand nombre de souches de *C. albicans*, l'expression du gène *ECE1* et la capacité de lyse des souches (Schönherr et al., 2017).

GO_term	Corrected P-value	Gene(s) annotated to the term
filamentous growth	0.00059	SOG2:HOG1:EFH1:FGR23:TEC1:ADR1:ARG83:HWP2:FGR28:FGR3:TSC2:SEC9:ALS4:ALS2:FGR6-1: CDC4:RFG1
growth	0.00087	SOG2:HOG1:EFH1:FGR23:TEC1:ADR1:ARG83:HWP2:FGR28:FGR3:TSC2:SEC9:ALS4:ALS2:FGR6-1: CDC4:RFG1
biological adhesion	0.00135	TEC1:ALS7:HWP2:ALS9:ALS4:ALS2:STE2
cell adhesion	0.00214	TEC1:ALS7:ALS9:ALS4:ALS2:STE2
symbiosis, encompassing mutualism through parasitism	0.01683	EFH1:TEC1:ALS7:HWP2:ALS9:ALS4:ALS2
interspecies interaction between organisms	0.01862	EFH1:TEC1:ALS7:HWP2:ALS9:ALS4:ALS2
single-species submerged biofilm formation	0.01957	TEC1:ALS7:ECE1:HWP2:ALS9:ALS2:STE2
single-species biofilm formation	0.02160	TEC1:ALS7:ECE1:HWP2:ALS9:ALS2:STE2
biofilm formation	0.03771	TEC1:ALS7:ECE1:HWP2:ALS9:ALS2:STE2

Tableau 9 : Lien ontologique entre les 50 gènes communs altérés dans le génome de la majorité des souches de patients CCMC. Seuls les liens significatifs sont représentés dans ce tableau.

- Détection des SNPs communs aux génomes de souches de patients CCMC par datamining

Nous avons ensuite utilisés des approches de « datamining » qui appliquent des algorithmes statistiques spécifiques pour l'analyse des données volumineuses « big-data ».

En effet, les statistiques « classiques » visent habituellement à affirmer avec un risque limité une hypothèse selon laquelle les valeurs de certains paramètres mesurés diffèrent dans une partie de la population étudiée. Dans ce contexte, la population étudiée (l'échantillon) est toujours plus importante que le nombre de paramètres, par exemple l'âge ou le sexe ratio d'une population d'individus.

Dans le cadre de la génomique, il nous intéresse, par exemple, de savoir si certains événements mutationnels sont reliés à une population particulière. Dans ce contexte, le nombre de paramètres mesurés (par exemple les SNPs), est largement plus important que la taille de l'échantillon. Ainsi, d'autres méthodes d'analyses sont utilisées sous le nom de « datamining » (extraction de connaissances à partir de données ou exploration de données). Ces approches peuvent employer deux méthodologies pour parvenir aux résultats. On parle d'approche « supervisée » ou « non supervisée ».

Dans les approches non supervisées, l'ensemble des paramètres (SNPs dans notre contexte) est utilisé pour permettre un « clustering » des échantillons, sans fournir d'hypothèse initiale. Les données correspondantes aux clusters sont ensuite analysées. Si l'on parvient à isoler certaines populations présentant le même profil, il est alors possible de les comparer entre elles, sur la base de données cliniques ou biologiques (même famille de patient, malade/non malade...) à l'aide de statistiques classiques. Si une hypothèse est validée, il est possible de revenir aux paramètres (SNPs) ayant guidé le clustering, et ainsi déceler une éventuelle association. Cependant, les profils génétiques guidant le clustering sont souvent difficiles à exploiter en raison de nombreux facteurs confondants interférant dans l'analyse.

La deuxième approche utilisée en datamining consiste à définir au préalable une hypothèse d'étude (par exemple un phénotype particulier). On parle alors d'analyse supervisée. Dans cette approche, on part du principe que l'hypothèse posée est vraie, puis on recherche les paramètres (SNPs) qui permettent au mieux d'expliquer cette hypothèse. Le set de données est alors réduit à quelques mutations qui peuvent être analysées par des tests statistiques standards. Dans notre étude nous avons testé les deux approches de datamining décrites ci-dessus.

Comparaison des génomes par approche non supervisée :

Dans ce contexte nous avons utilisé l'ensemble des SNPs détectés pour chacun des génomes (CCMC n=37 et contrôles n=144) afin d'établir un arbre phylogénétique (**Figure 34**). Celui-ci a montré une clusterisation des souches selon leur appartenance aux différents clades. Cette approche ne nous a donc pas permis d'isoler un profil génétique spécifique des génomes des souches CCMC. Cependant on observe avec cette analyse que les souches issues d'un même patient ou les souches provenant d'une même famille clusterisent ensemble confirmant l'importante proximité génétique des souches entre elles.

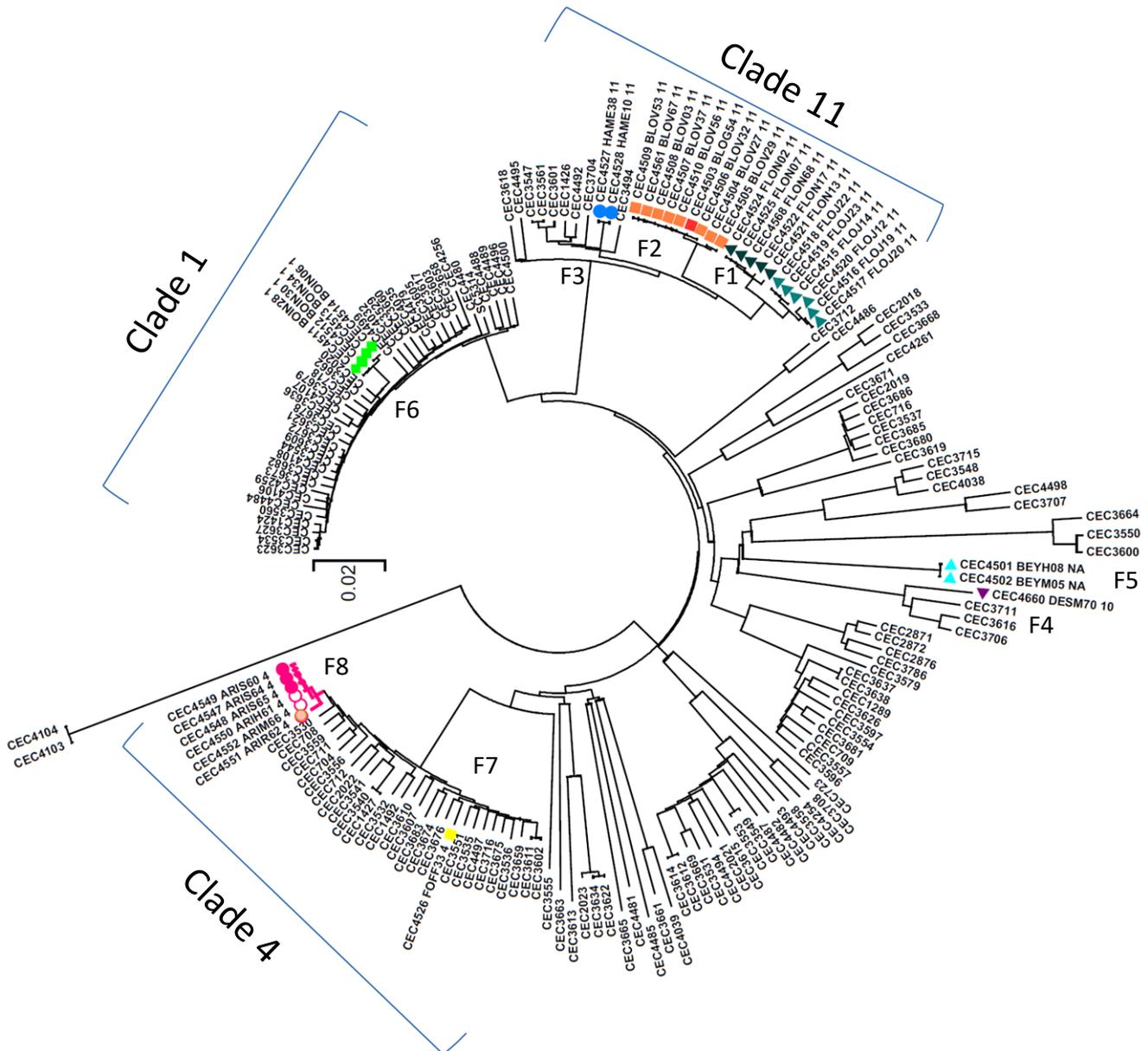


Figure 34 : Arbre phylogénétique réalisé à partir des données de séquençage des génomes des souches de patients CCMC et de leur famille en comparaison avec les 144 génomes contrôles. Celui-ci a été réalisé grâce au logiciel MEGA7 (Kumar et al., 2016), en utilisant la méthode de Neighbor-Joining (Saitou and Nei, 1987). Les distances évolutives ont été calculées à l'aide de la méthode du maximum de vraisemblance composite (Tamura et al., 2004). L'analyse a été réalisée sur 180 séquences nucléotidiques comprenant un total de 537 126 positions polymorphiques. Les souches de patients CCMC sont identifiées avec les différentes couleurs. Le numéro d'appartenance aux différentes familles est également indiqué (F= famille).

Comparaison des génomes par approche supervisée :

L'approche par datamining de type supervisée a été utilisée grâce à la mise en place de la méthodologie suivante (Rodriguez et al., 2015).

Brièvement, des algorithmes de classification dichotomiques (Random Forest et Decision Tree) et de régression linéaire (Lasso ; Least Absolute Shrinkage and Selection Operator) inclus au sein d'un script R (packages : randomForest, rpart, MASS, ROCR, cluster, pvclust, glmnet et pROC) ont été utilisés grâce à l'outil RGui (v2.14.1) (Robin et al., 2011; Tibshirani, 1996).

Ces algorithmes ont permis de mettre en évidence des gènes particulièrement intéressants (plus ou moins mutés chez les CCMC en comparaison avec les 144 génomes contrôles). Un test statistique standard (test Z ou t corrigé) a été réalisé pour déterminer la significativité des gènes identifiés.

Différentes hypothèses ont été testées successivement avec ces outils :

Premièrement, nous avons testé une approche par « mutation ». Nous avons posé l'hypothèse que certains SNPs provenant des génomes CCMC (n=11) étaient différents de ceux des génomes contrôles (n=144). Par ce biais, aucune position n'a été retrouvée avec une p-value significative.

Deuxièmement, nous avons testé une approche par « gène ». Nous avons posé l'hypothèse que certains gènes étaient plus fréquemment mutés chez les génomes des CCMC en comparaison avec les génomes contrôles. Pour cela, nous avons comptabilisé le nombre total de SNP par gène que nous avons comparé entre les 2 populations. Cette approche ne nous a pas permis d'identifier des gènes intéressants. Pour augmenter notre puissance statistique, nous avons réduit nos données de SNPs en ne comptabilisant que le nombre de SNPs non synonymes et donc pouvant avoir un impact biologique. Par cette approche nous avons identifié plusieurs gènes pour chacun des deux modèles utilisés (classification et régression linéaire).

Le modèle de régression linéaire (Lasso) a permis d'obtenir une liste de gènes associés à un Odds ratio et une p-value (test Z) ainsi qu'une représentation graphique des principaux résultats (**Tableau 10 et Figure 35**). Dans cet exemple, le gène *OCA1 (C2_10190C_A)* est retrouvé significativement plus muté chez les CCMC que chez les contrôles. En effet, ce gène est retrouvé significativement plus muté chez les CCMC (64%, soit 7/11) que chez les contrôles. (9,7%, soit 14/143).

Avec le modèle de classification (Random Forrest), nous avons également obtenu une liste de gènes associés à un score correspondant au coefficient de Gini (mesure statistique de la dispersion d'une distribution dans une population donnée) (**Figure 36**). Dans cet exemple, plusieurs gènes ont été identifiés. Les gènes, *THI6* et *OCA1*, également mis en évidence par l'approche de régression linéaire, ainsi que d'autres gènes non caractérisés chez *C. albicans* (**Figure 36**).

	Estimate	Std. Error	z value	Pr(> z)	OR	lower	upper	
C2_10190C_A	0,59453214	0,28410303	2,09266387	0,036379171	1,8121829	1,03842012	3,16250313	*
C3_02860W_A	0,19265094	0,47317408	0,40714602	0,683900747	1,2124595	0,47962381	3,0650231	NS
CR_03520C_A	0,67883197	0,4850051	1,39963882	0,161621502	1,97157352	0,7620367	5,10093824	NS
CR_08900C_A	1,95437504	192,727901	0,01014059	0,991909117	7,05950571	6,29E-164	7,93E+164	NS

Tableau 10 : Caractéristiques statistiques des gènes identifiés par approche de régression linéaire de type Lasso.

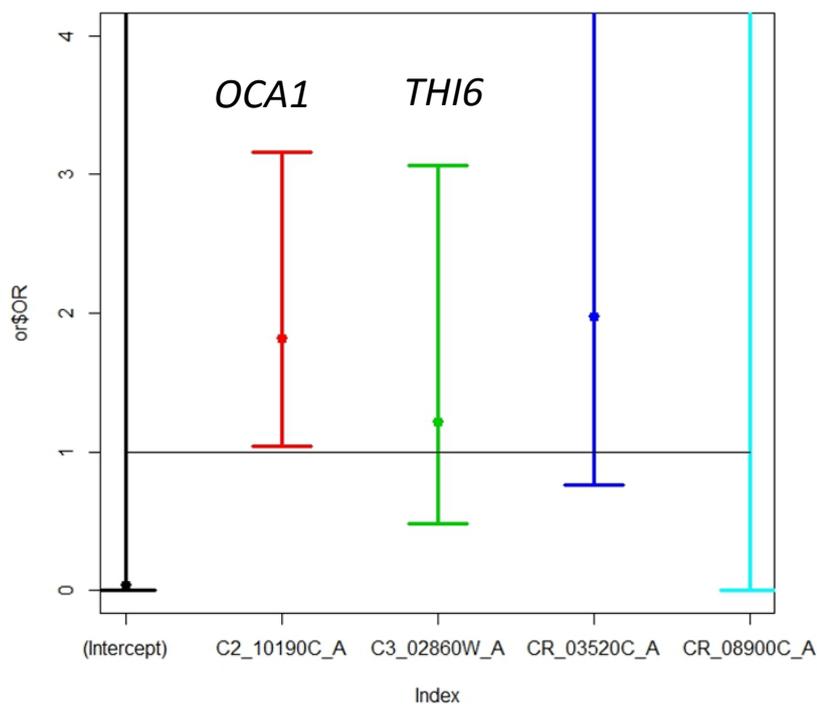


Figure 35 : Représentation graphique des Odds ratio obtenus pour les gènes identifiés par l'approche de régression linéaire de type Lasso.

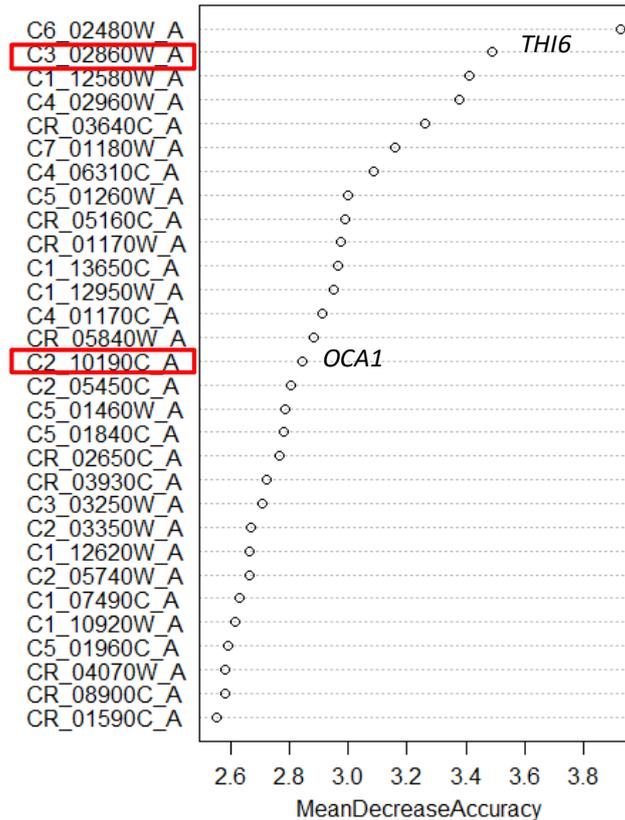


Figure 36 : Représentation graphique des scores obtenus (coefficient de Gini normalisé) pour les gènes identifiés par l'approche de type classification (Random Forrest).

L'approche « par gène », comptabilisant le nombre de SNPs non-synonymes, a donc été sélectionnée et testée avec plusieurs jeux de données. Chaque jeu correspondant à différentes hypothèses (n=10) dans lesquelles le nombre de génomes comparé a été modifié. Par exemple, n=33 ou n=11 pour les génomes des CCMC (correspondant soit à l'ensemble des génomes séquencés pour les CCMC soit à 1 génome par patient). De la même façon, nous avons progressivement réduit le set de données des 144 génomes contrôles afin de cibler au mieux l'hypothèse testée et donc les gènes identifiés. Par exemple, en ne sélectionnant que les génomes provenant d'isolats commensaux (n=27 ou n=7) ou en retirant du set de données certains génomes provenant d'isolats responsables d'infections superficielles tels que les infections vaginales (n=131 ou n=104). Ainsi, différentes combinaisons ont été testées générant à chaque analyse une liste de gènes candidats par modèle. L'ensemble des données obtenues a été regroupé dans une table afin de ne sélectionner que les gènes les plus pertinents.

Au total, par la méthode de régression linéaire (Lasso) nous avons identifié (pour les 10 hypothèses testées) 36 gènes différents, dont 14 ont été trouvés $\geq 2X$ (*OCA1*, *THI6*, *MRP20*, *MRP20*, *GND1*, *UBI4*, *EMP24*, *ACS1*...). Parmi les gènes identifiés, 3 ont été trouvés avec une p-value significative (*OCA1* p=0,036 ; *CDC34* p=0,014 ; *C4_02470C_A* p=0.041). Seul le gène *OCA1* est caractérisé chez *C. albicans*.

Par la méthode de classification (Random Forrest) nous avons identifié 93 gènes (pour les 10 hypothèses testées) dont 41 ont été trouvés $\geq 2X$ (*THI6*, *RPA135*, *MRP20*, *IDP2*, *UBI4*, *ZCF31*, *EMP24*, *TAC1*, *ERG11*, *FAV1*, *PGA41*, *GND1*, *MDM12*, *RTG3*, *MDM10*, *MNNN11*, *ACS1*, *RPL17B*, *PMT4*, *DCK1*, *TRP2*, *BMT6*, *OCA1*...). Les gènes *TAC1* et *ERG11*, impliqués dans la résistance aux antifongiques azolés, ont également été identifiés par cette méthode, montrant la capacité du datamining à identifier des gènes candidats potentiellement impliqués dans l'adaptation génomique de *C. albicans* en réponse aux stress dont l'implication reste à démontrer. Au final, 19 gènes ont été identifiées de façon commune par les deux méthodes (**Tableau 11**), parmi lesquelles seuls 6 ont été caractérisés chez *C. albicans* (*UBI4*, *EMP24*, *GND1*, *RTG3*, *OCA1* et *CAS5*).

ORF	Type	Name	p-Value (Z test from Lasso)	Description
C3_07270C_A	Verified	UBI4		Ubiquitin precursor (polyubiquitin) contains 3 tandem repeats of the ubiquitin peptide that are processed to individual units; transcript induced by stress; mRNA in yeast and mycelia at similar abundance; Spider biofilm induced
CR_07590W_A	Verified	EMP24		COPII-coated vesicle component
C1_13860C_A	Verified	GND1		6-phosphogluconate dehydrogenase; soluble in hyphae; farnesol, macrophage-induced protein; antigenic in mice; dual localization to cytosol and peroxisomes depends on alternative splicing; rat catheter and Spider biofilm repressed
C1_10990C_A	Verified	RTG3		Putative transcription factor with bZIP DNA-binding motif; mutants are sensitive to cations and antifungal drugs, and defective in filamentous growth
C2_10190C_A	Verified	OCA1	0,036	Putative protein phosphatase of the PTP family (tyrosine-specific); ortholog of <i>S. cerevisiae</i> Oca1; mutant is viable; mutant shows virulence defect
C4_01190W_A	Verified	CAS5		Transcription factor; cell wall damage response; required for adherence, response/resistance to caspofungin; repressed in core stress response; mutants have reduced CFU in mice, hyphal defect in <i>C. elegans</i> infection; Spider biofilm induced
CR_03640C_A	Uncharacterized			Ortholog(s) have role in protein targeting to mitochondrion and mitochondrial outer membrane localization
C3_02860W_A	Uncharacterized	THI6		Putative thiamin-phosphate pyrophosphorylase, hydroxyethylthiazole kinase; fungal-specific; Spider biofilm induced
C7_00570W_A	Uncharacterized	RPA135		Putative RNA polymerase I subunit A135; repressed by prostaglandins
C1_01600W_A	Uncharacterized	MRP20		Component of mitochondrial ribosome; decreased expression in hyphae compared to yeast-form cells
C1_00700W_A	Uncharacterized			Protein of unknown function; repressed by nitric oxide
C7_03270W_A	Uncharacterized			Putative adhesin-like protein
C2_10350C_A	Uncharacterized	ACS1		Acetyl-CoA synthetase; induced by human neutrophils; fluconazole-repressed; regulated by Nrg1/Mig1; colony morphology-related gene regulation by Ssn6; only in stationary phase cultures; rat catheter biofilm repressed, Spider biofilm induced
C6_02480W_A	Uncharacterized			Similar to alcohol dehydrogenases; induced by benomyl treatment, nitric oxide; induced in core stress response; oxidative stress-induced via Cap1; Spider biofilm repressed
C3_07050W_A	Uncharacterized			Ortholog(s) have enzyme activator activity
C2_10200W_A	Uncharacterized			Predicted olichyl-diphosphooligosaccharide-protein glycotransferase; role in protein N-linked glycosylation; Spider biofilm repressed
CR_03520C_A	Uncharacterized		0,055	Putative actin cytoskeleton component; protein present in exponential and stationary growth phase yeast cultures
C1_02010C_A	Uncharacterized	ZCF24		Predicted Zn(II)2Cys6 transcription factor; caspofungin induced; Hap43-repressed
C1_10940C_A	Uncharacterized	ERP5		Protein involved in ER to Golgi transport; rat catheter and Spider biofilm repressed

Tableau 11 : Liste des 19 gènes identifiés par les deux approches (Lasso et Random Forrest).

Notre attention s'est portée sur le gène *OCA1* pour lequel une p-value significative a été retrouvée. Ce gène code pour une putative tyrosine phosphatase, nécessaire à l'arrêt du cycle cellulaire (en phase G1) en réponse aux dommages de l'ADN causés par un stress oxydatif de type lipidique (lipides oxydés).

Les lipides représentent la base structurale des membranes cellulaires et sont donc des constituants essentiels des cellules. Les acides gras polyinsaturés (AGPI) sont un groupe de lipides jouant un rôle majeur dans le maintien de la fluidité membranaire chez les organismes procaryotes et eucaryotes. Ces AGPI, sont les précurseurs d'une classe de molécules appelées oxylipines (correspondant à l'oxydation des AGPI).

Ces molécules ont un rôle physiologique majeur mais sont également impliquées dans des processus pathologiques. Chez les champignons, il a été montré que les oxylipines étaient impliquées dans la régulation de la morphogenèse et du cycle reproductif (Kock et al., 2007). Les oxylipines ont également des propriétés immunomodulatrices de l'hôte et semble être importantes pour la pathogénicité de *C. albicans*. Par exemple, la production de l'oxylipin 3,18-diHETE se produit au sein des hyphes et lors de la formation du biofilm, deux acteurs importants dans la pathogénicité de *C. albicans* (Deva et al., 2001, 2000).

L'hydro peroxyde d'acide linoléique (LoaOOH) est également un lipide oxydé. Celui-ci est formé à la suite de l'attaque des acides gras insaturés à longue chaîne par des radicaux libre (Evans et al., 1998). Chez *S. cerevisiae* cette molécule est toxique même à faible concentration, et la protéine Oca1p est nécessaire à la croissance des cellules en présence de cette substance (Alic et al., 2001; Evans et al., 1998). Cependant, Oca1p n'est pas essentielle à la survie immédiate des cellules (Alic et al., 2001). La suppression du gène *OCA1* ou de ses homologues (*SIW14*, *OCA2*) résulte en une augmentation de la sensibilité à l'LoaOOH, à la ménadione et de façon intéressante à la caféine.

Chez *C. albicans* Oca1p est également une phosphatase putative de la famille des PTP (tyrosine spécifique) ; le seul phénotype connu pour du mutant KO de ce gène consiste en un défaut de virulence chez *Galleria mellonella* (Hanaoka et al., 2008).

Dans le cadre de notre projet, nous avons regardé la répartition des mutations sur le gène *OCA1* pour les CCMC et les contrôles. La position la plus fréquemment retrouvée, que ce soit chez les CCMC (5/11) ou les contrôles (7/144), était commune aux souches provenant du clade 11. Afin d'écartier un biais en rapport avec l'origine des souches nous avons comparé phénotypiquement nos souches à différentes souches contrôles (provenant ou non du clade 11 et ayant ou non la mutation). Cependant nous n'avons pas trouvé de corrélation entre la présence/absence de cette mutation sur le gène *OCA1* et la réponse aux différents stress testés tels que la caféine, le peroxyde d'hydrogène et la ménadione.

3.2.3.2 Evolution phénotypique et génotypique des souches de patients CCMC et pistes futures

Dans notre collection de souches provenant des 13 patients CCMC, nous disposons de séries de souches (isolées sur des périodes allant de 1 à 9 années pour 6 d'entre eux (P1, P2, P3, P5, P9 et P13). L'analyse phénotypique et génomique de ces souches nous a apporté de nombreuses informations en termes d'évolution des souches au cours du temps dans notre modèle de patients CCMC. En effet, comme nous l'avons vu précédemment, que ce soit pour l'analyse des phénotypes ou des génomes, d'importantes variations entre souches ont été observées au cours du temps. Nous avons donc regardé, à travers plusieurs exemples, si la survenue de certains « switch phénotypiques » était associée à des modifications génomiques (LOH, SNPs).

- Evolution des souches chez le patient P1

La souche la plus récente isolée en 2014 (souche #FLOJ12) a présenté plusieurs phénotypes différents de ceux observés chez les 5 autres souches isolées chez ce patient : i. une absence de filamentation en YPD à 30°C ii. un défaut de croissance en SD2%, iii. une augmentation de la CMI à l'anidulafungine, iv. une sensibilité accrue au stress thermique et v. une augmentation de la résistance à la caféine. Ces résultats sont résumés dans la **Figure 37**.

En parallèle, l'analyse comparative des 4 génomes a montré l'apparition de 2 LOH sur le génome de cette même souche (souche #FLOJ12). Ces LOH sont localisés au niveau du bras droit du Chr4 et à l'extrémité gauche du ChrR. Au total 6285 SNPS ont été identifiés entre la souche de 2009 et celle de 2014, dont la moitié (n=2794) étaient impliqués dans les 2 LOH de la souche de 2014. Nous avons retrouvé 629 mutations non synonymes (NS) distribuées sur 429 ORFs et 10 codons stop (10 ORFs). Nous nous sommes intéressé à la survenue d'un codon stop sur le gène *TPS2* car il pourrait expliquer le défaut de croissance, aux différentes températures, de la souche #FLOJ12. En effet, ce gène code pour une tréhalose-6-phosphate (Tre6P) phosphatase impliquée dans le métabolisme central de la cellule. Chez *C. albicans* des mutants homozygotes pour ce gène ont montré un défaut de croissance, une augmentation de la sensibilité au choc thermique, une sensibilité au stress oxydatif et une diminution de la virulence dans un modèle murin. Afin d'explorer l'implication de ce gène dans le(s) phénotype(s) observé(s), nous avons transformé la souche de 2009 (souche #FLOJ23) avec un allèle du gène *TPS2* contenant le codon stop.

Nous avons comparé les courbes de croissance, en SD2% et YPD à 42°C, du mutant avec celles des différentes souches isolées chez ce patient (souche #FLOJ23_2009 et #FLOJ12_2014). Cependant, l'introduction de la mutation n'a pas induit de modifications phénotypiques. De nombreux autres gènes pourraient être testés afin d'identifier celui impliqué dans le(s) phénotype(s) observé(s).

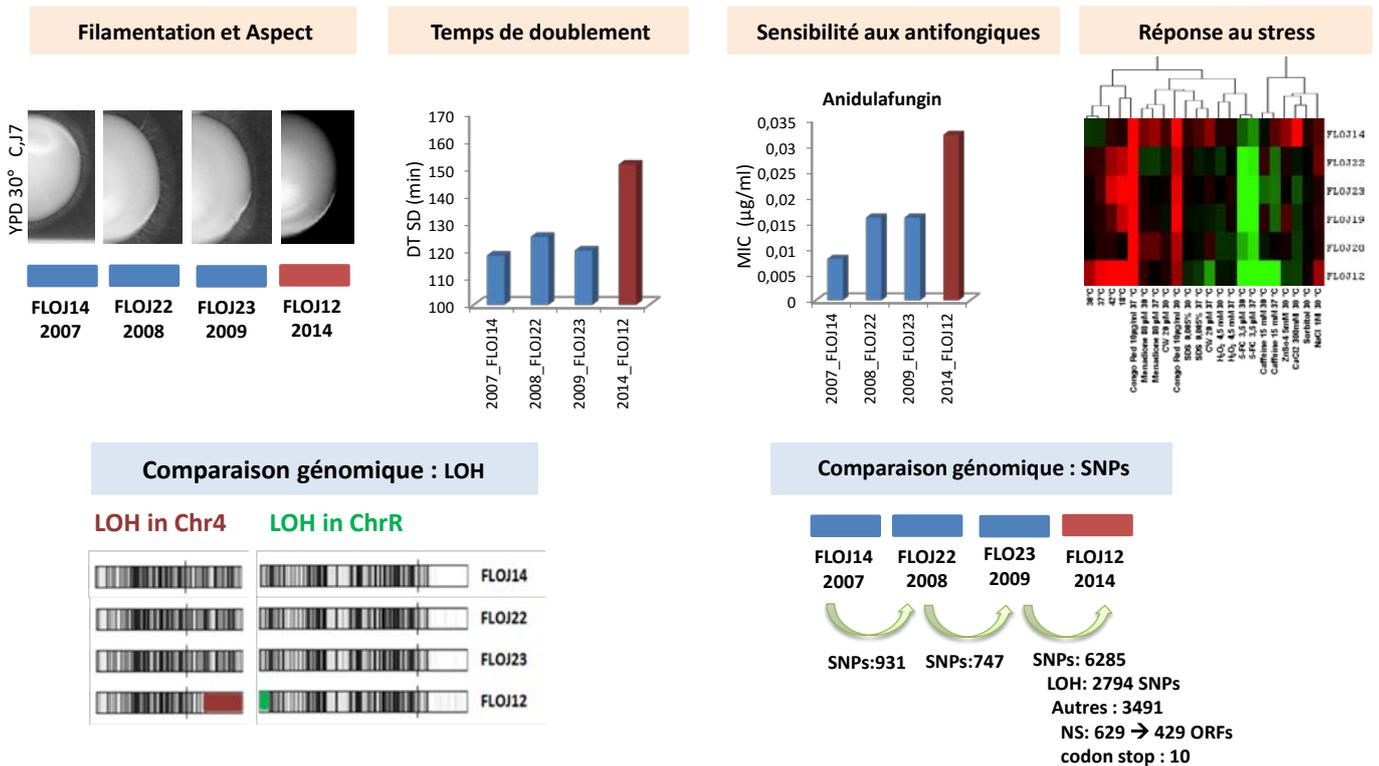


Figure 37 : Evolution phénotypique et génomique de 4 souches provenant du patient P1. Dans cette figure sont représentés les « switch phénotypiques » ainsi que les modifications génomiques majeures survenues au cours du temps entre 4 des 6 souches isolées chez le patient P1 entre 2007 et 2014. La souche #FLOJ12 isolée en 2014 montre de nombreuses variations phénotypiques et génomiques en comparaison avec les autres souches. (Chr= chromosome; LOH = Loss Of Heterozygosity; SNP=Single Nucleotide Polymorphism; ORF: Open Reading Frame; NS= SNP Non Synonyme).

- Evolution des souches chez le patient P9

Un autre exemple provient du patient P9. Chez ce patient les souches isolées en 2012 et 2013 (souche #BOIN34 et #BOIN06) ont également présenté des « switch phénotypiques » tels que : i. l'acquisition de la capacité à filamenter, ii. un défaut de croissance, iii. une modification de la réponse à plusieurs stress incluant la caféine iv. une augmentation de la lyse des kératinocytes humains. Ces résultats sont résumés dans la **Figure 38**.

En parallèle, l'analyse comparative des 4 génomes a montré l'apparition et la persistance d'un large LOH sur le bras droit du Chr4 sur les souches isolées à partir de 2012 (souche #BOIN34 et #BOIN06). Un LOH additionnel a été observé au niveau du bras gauche du Chr1, sur la souche la plus récente (souche #BOIN06).

La comparaison du nombre de SNPs a montré un nombre important de différences entre les génomes de 2010 et 2012 (souche #BOIN30 et #BOIN34) (n=3229) (Figure 38). Parmi ces SNPs, un grand nombre d'entre eux sont la conséquence du LOH situé sur le Chr4 (533 SNPs (NS) impactant 124 gènes différents). Parmi les gènes identifiés, nous avons retrouvé à l'état homozygote plusieurs mutations NC dans le gène *GAT1*. Ce gène est un facteur de transcription impliqué dans la régulation de l'azote et donc dans la voie de signalisation mTOR (répression du catabolisme de l'azote)(Beck and Hall, 1999; Homann et al., 2009; Limjindaporn et al., 2003). De façon intéressante, les mutants de délétion de ce gène ont montré une résistance importante à la caféine ainsi qu'une diminution de la virulence dans un modèle d'infection disséminée chez la souris (Homann et al., 2009; Limjindaporn et al., 2003). D'autres mutations homozygotes NS induites par le LOH ont été retrouvées dans des gènes d'intérêt, tels que *ZCF5* (Codant pour un facteur de transcription impliqué dans la régulation de gènes jouant un rôle dans la morphologie des colonies (Vandeputte et al., 2011)) et *IFF5* (codant pour une adhésine de la famille Iff impliqué dans l'adhésion cellulaire (Bates et al., 2007)) et sont donc également intéressants à explorer.

Par ailleurs, en plus des SNPs liés à l'existence du LOH au niveau du Chr4, 60 SNPs NS ont été identifiés impactant 37 gènes différents. Parmi ces gènes, 3 codent pour des adhésines de la famille des Als (*ALS1*, *ALS6* et *ALS9*). Ces 3 adhésines sont impliquées dans l'adhésion cellulaire au cours de la formation du biofilm (Klotz et al., 2007). Deux autres gènes, non caractérisés chez *C. albicans*, étaient également porteurs de mutations NS homozygotes.

Le gène *ZCF32*, codant pour un facteur de transcription impliqué dans la régulation de la formation du biofilm (Kakade et al., 2016). Les mutants de délétions de ce gène ont un phénotype d'hyperfilamentation. Le gène *SEN1*, non caractérisé, code pour une hélicase putative réprimée en présence de prostaglandines (Levitin and Whiteway, 2007). Les prostaglandines sont des hormones de l'hôte impliquées dans la réponse inflammatoire et dans la défense contre de nombreux pathogènes (Levitin and Whiteway, 2007). L'ensemble de ces gènes peut être intéressants à étudier dans le cadre de l'évolution des souches de *C. albicans* dans un contexte pathologique de longue durée.

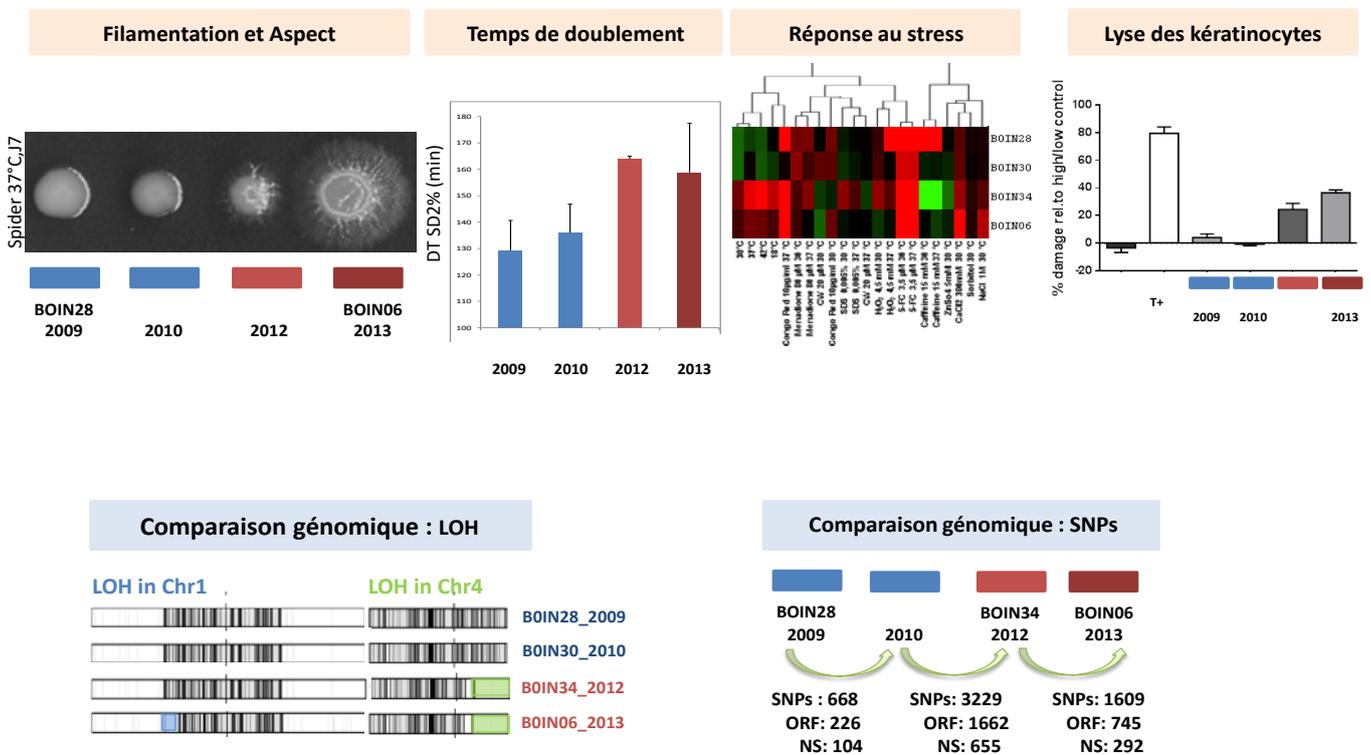


Figure 38 : Evolution phénotypique et génomique des 4 souches provenant du patient P9. Dans cette figure sont représentés les « switch phénotypiques » ainsi que les modifications génomiques majeures survenues au cours du temps entre les 4 souches isolées chez le patient P9 entre 2009 et 2013. Les souches #BOIN34 et #BOIN06 montrent de nombreuses variations phénotypiques et génomiques en comparaison avec les 2 autres souches. (Chr= chromosome; LOH = Loss Of Heterozygosity; SNP= Single Nucleotide Polymorphism; ORF: Open Reading Frame; NS= SNP Non Synonyme).

- Evolution des souches entre individu malade et sains de la famille F8

Un autre aspect extrêmement intéressant a consisté à étudier comment une même souche de *C. albicans* (même profil MLST) pouvait évoluer chez deux individus différents d'une même famille dont l'un est malade et l'autre sain. En effet, notre cohorte de patients CCMC inclut une famille unique pour laquelle nous disposons de souches provenant de deux individus malades (P12 (n=1) et P13 (n=3) et 2 individus non malades (I1 (n=1) et I2 (n=1)).

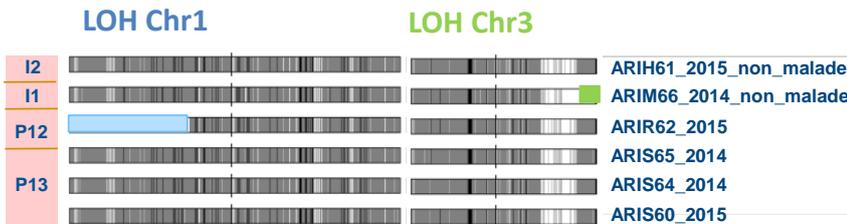
Les 6 souches partageaient le même profil MLST. L'analyse comparative des 6 génomes a montré des résultats très intéressants. Premièrement, bien que l'ensemble des 6 souches soit extrêmement proches en termes d'hétérozygotie, nous avons observé un LOH au niveau du bras droit du Chr3 de la souche isolée chez un individu non malade I1 (souche #ARIM66) ainsi qu'un LOH au niveau du bras gauche du Chr1 de la souche du patient P12 (souche #ARIR62) (**Figure 39, A**). En revanche, les 3 souches provenant du patient P13 et de l'individu non malade I2 (souche #ARIH61) étaient identiques en termes de LOH (**Figure 39, A**). La comparaison de 3 souches identiques en terme de LOH (#ARIH61 de l'individu sain I2 et les deux souches #ARIS60 et #ARIS65 du patient P13) a montré un nombre important de SNPs.

Un total de 2547 SNPs, dont 579 NS, impactant un total de 174 gènes a été identifié entre la souche de l'individu sain I2 (souche #ARIH61) et celle du patient P13 (souche #ARIS60) (**Figure 39, B**). Parmi les 174 gènes, seuls 53 ont été caractérisés dans le génome de *C. albicans*. Une seule mutation homozygote NS a été retrouvée entre ces deux génomes. Celle-ci était localisée au sein du gène *ARP4* qui code pour une sous-unité du complexe NuA4 HAT (Histone acétyltransférase). Il a été montré que ce complexe était nécessaire pour l'induction de gènes spécifiques de la forme Hyphale via Efg1 (Lu et al., 2008).

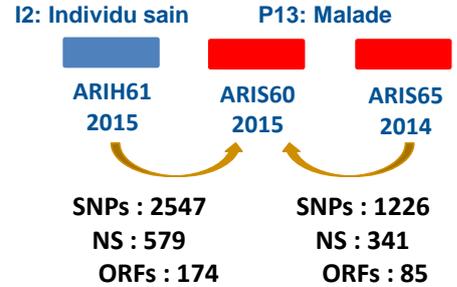
De façon intéressante, une analyse du lien ontologique entre les 53 gènes caractérisés a permis de montrer un lien significatif entre ces gènes, dont les principaux sont résumés dans la **Figure 39, partie C**. Ces résultats montrent une différence en termes d'évolution des gènes en fonction de l'environnement (individu sain vs malade) dans lequel progressent les souches de *C. albicans* à long terme. En effet dans cette analyse, de nombreux gènes identifiés sont impliqués dans des processus essentiels dans le cadre de l'interaction hôte-pathogène tels que i. la capacité des souches à croître, ii. à filamento, mais également iii. à répondre aux stress.

Ces résultats suggèrent que l'interaction pathologique entre *C. albicans* et l'Homme pourrait induire une modification de l'évolution génomique des souches et donc une possible adaptation spécifique à leur environnement.

A. Comparaison génomique : LOH



B. Comparaison génomique : SNPs



C. Comparaison génomique : Ontologie

Term from	GO_term	Cluster frequency	Corrected P-value	Gene(s) annotated to the term	
Component	cell periphery	24 out of 53 genes, 45.3%	2.48e-08	GPM2:GSL1:SPA2:RBE1:HGT6:DCK1:PNG2:TOR1:SIT1:C2_10010C_A:ALS7:MYOS:GUT2:HYR3:HSL1:ALS1:ALS4:ALS2:CSA1:RBR3:SLN1:GEF2:ALS3:ELF1	
	cell surface	12 out of 53 genes, 22.6%	4.85e-06	RBE1:SIT1:FGR23:ALS7:HYR3:ALS1:ALS4:ALS2:CSA1:RBR3:PHO113:ALS3	
	cell wall	10 out of 53 genes, 18.9%	5.42e-05	GPM2:RBE1:ALS7:HYR3:ALS1:ALS4:ALS2:CSA1:RBR3:ALS3	
Process	filamentous growth	24 out of 53 genes, 45.3%	7.78e-09	SPA2:DCK1:RPB4:YAK1:FGR51:FGR12:FGR23:C3_02760C_A:PEP8:C4_02420C_A:FGR28:MYOS:HSL1:APM1:C5_05510C_A:C6_01220C_A:ALS1:ALS4:CDC39:ALS2:L	
	growth	24 out of 53 genes, 45.3%	1.28e-08	SPA2:DCK1:RPB4:YAK1:FGR51:FGR12:FGR23:C3_02760C_A:PEP8:C4_02420C_A:FGR28:MYOS:HSL1:APM1:C5_05510C_A:C6_01220C_A:ALS1:ALS4:CDC39:ALS2:L	
	cellular response to stimulus	29 out of 53 genes, 54.7%	2.34e-06	CNS1:GSL1:GAR1:DCK1:RPB4:YAK1:FGR51:FGR12:GRE2:TOR1:SIT1:FGR23:C3_02760C_A:PEP8:FGR28:DCR1:MYOS:YBL053:HSL1:MEC1:APM1:C5_05510C_A:ARP4:ALS1:CDC39:LIP5:PLC1:SLN1:IME2	
	cellular response to starvation	14 out of 53 genes, 26.4%	4.30e-06	DCK1:YAK1:FGR51:FGR12:FGR23:C3_02760C_A:PEP8:FGR28:MYOS:HSL1:APM1:C5_05510C_A:CDC39:LIP5	
	cellular response to nutrient levels	14 out of 53 genes, 26.4%	2.50e-05	DCK1:YAK1:FGR51:FGR12:FGR23:C3_02760C_A:PEP8:FGR28:MYOS:HSL1:APM1:C5_05510C_A:CDC39:LIP5	
	filamentous growth of a population of unicellular organisms in response to starvation	12 out of 53 genes, 22.6%	4.96e-05	DCK1:FGR51:FGR12:FGR23:C3_02760C_A:PEP8:FGR28:MYOS:APM1:C5_05510C_A:CDC39:LIP5	
	cell communication	17 out of 53 genes, 32.1%	7.17e-05	DCK1:YAK1:FGR51:FGR12:TOR1:FGR23:C3_02760C_A:PEP8:FGR28:MYOS:HSL1:APM1:C5_05510C_A:CDC39:LIP5:PLC1:SLN1	
	cellular response to stress	20 out of 53 genes, 37.7%	0.00073	DCK1:RPB4:YAK1:FGR51:FGR12:GRE2:FGR23:C3_02760C_A:PEP8:FGR28:MYOS:YBL053:HSL1:MEC1:APM1:C5_05510C_A:ARP4:CDC39:LIP5:SLN1	
	Function	protein kinase activity	6 out of 53 genes, 11.3%	0.01435	YAK1:TOR1:HSL1:MEC1:SLN1:IME2
		transferase activity, transferring phosphorus-containing groups	9 out of 53 genes, 17.0%	0.02381	RPB4:YAK1:TOR1:C4_02420C_A:DCR1:HSL1:MEC1:SLN1:IME2

Figure 39 : Comparaison des génomes de souches de *C. albicans* isolées chez un individu sain et un malade. A. Comparaison de la survenue des événements de LOH entre les 6 souches provenant des individus sains et malades de la famille F8 (I1, n=1 ; I2 n=1, P12 n=1 ; P13 n=3). B. Comparaison du nombre de SNPs. Seules les souches : #ARIH61 de l'individu sain I2 et les deux souches #ARIS60 et #ARIS65 du patient P13 sont représentées car elles ne diffèrent que par des événements de SNPs non impliqués dans un LOH. C. Lien ontologique entre les gènes, impactés par des mutations NS différenciant la souche #ARIH61 de l'individu sain I2 et la souche #ARIS60 du patient P13. Cette analyse a permis de montrer plusieurs liens significatifs entre ces gènes en rapport avec leur fonction moléculaire, leur compartiment cellulaire et leur fonction biologique. (Chr= chromosome; LOH = Loss Of Heterozygosity; SNP= Single Nucleotide Polymorphism; ORF: Open Reading Frame; NS= SNP Non Synonyme, GO=Gene Ontology).

3.3 Discussion

La Candidose Cutané-Muqueuse Chronique (CCMC) est un cas particulier d'interaction entre *C. albicans* et l'Homme constituant un modèle privilégié pour l'étude de l'évolution génomique de *C. albicans* chez l'hôte naturel. En effet, au cours de cette pathologie, la persistance ou récurrence des infections survenant chez ces patients induit un environnement inflammatoire chronique provoquant un stress permanent sur les souches. Ce modèle permet ainsi d'étudier les mécanismes adaptatifs mis en place par *C. albicans* au cours de l'interaction pathologique prolongée chez ces malades. De plus, la gravité des infections dues à *C. albicans* chez ces patients nécessite le recours à un traitement prolongé ou répété par des antifongiques, essentiellement de la classe des azolés ou des échinocandines. L'utilisation de ces molécules sur une longue période peut être responsable de l'émergence de variants génétiques résistants aux antifongiques à l'origine de l'échec thérapeutique chez ces patients.

L'objectif de cette partie du travail était d'utiliser ce modèle afin d'explorer si *C. albicans* était capable d'adaptations spécifiques dans un contexte d'interaction à long terme avec son hôte, dans un environnement hostile, d'une part en raison d'une réponse inflammatoire exacerbée mais également en raison de l'utilisation d'antifongiques.

Pour cela, nous avons constitué une collection unique de souches de *C. albicans* isolées sur plusieurs années chez des patients atteints de CCMC et leur famille, sur laquelle nous avons réalisé une caractérisation phénotypique et génomique. La collection inclut 63 souches - 60 isolées chez 13 individus CCMC issus de 8 familles et 3 isolées chez des membres non malades de leur famille.

Dans un premier temps, nous avons typé les souches par Multi Locus Sequence Typing (MLST). Cette analyse a permis de montrer que les patients étaient colonisés ou infectés à long terme (suivi allant jusqu'à 9 ans) par une souche unique ayant un même profil MLST. Nous avons également montré que la détection chez un patient d'une souche différente (profil MLST différent) était rare et transitoire. Notre étude a également révélé qu'au sein d'une même famille, les membres sains et malades pouvaient être colonisés par la même souche (même profil MLST).

Grâce à ces données de typage, nous avons sélectionné, au sein de notre collection, 37 souches pour lesquelles nous avons réalisé une caractérisation phénotypique étendue et réalisé le séquençage du génome complet.

La caractérisation phénotypique des souches a été effectuée dans différentes conditions expérimentales pertinentes au regard de l'interaction avec l'hôte : (i) sensibilité vis-à-vis de divers antifongiques, (ii) cinétique de croissance, (iii) sensibilité à différents stress (oxydatif, cationique, osmotique, thermique), (iv) capacité des souches à réaliser la transition levure-filament en réponse à différents stimuli et (v) capacité à lyser les kératinocytes humains. L'analyse de ces phénotypes a révélé d'importantes variations phénotypiques entre les souches des patients CCMC.

En effet, comme nous nous y attendions, aucun phénotype commun à toutes les souches des patients CCMC n'a pu être identifié dans les conditions testées. Cependant, plusieurs phénotypes intéressants ont pu être mis en évidence. Par exemple, plusieurs conditions, jouant sur l'intégrité de la paroi fongique, telles que l'effet du rouge Congo (CR) et de la Caféine, ont révélé des variations phénotypiques importantes chez ces souches. Une majorité des souches se sont avérées être impactées par l'usage du CR tandis qu'une majorité des souches ont montré une meilleure résistance à la caféine en comparaison avec la souche SC5314. Concernant la caféine, nous avons observé une augmentation de la résistance des souches au cours du temps chez l'ensemble des patients pour lesquels plusieurs souches ont été étudiées chronologiquement. De façon intéressante, les souches ayant présentées le plus haut niveau de résistance à la caféine ont présenté également une sensibilité diminuée aux antifongiques azolés (patients P1 à P5), pouvant suggérer un lien entre ces deux phénotypes. Bien que ces résultats doivent être confirmés par des analyses complémentaires, ils sont particulièrement intéressants et constituent une piste majeure pour une exploration détaillée de l'évolution des souches dans un contexte pathologique. C'est pourquoi, dans la suite de cette discussion, nous avons porté un intérêt particulier à la description des voies de signalisation impliquées dans l'intégrité de la paroi fongique, en particulier l'impact de la caféine, afin d'identifier un potentiel lien entre les phénotypes observés et un gène spécifique qui aurait pu être la cible d'altération chez nos patients.

En effet, comme nous l'avons vu au cours de l'introduction, la paroi fongique joue un rôle central dans l'adaptation de la levure à son environnement. Bien que l'architecture de la paroi cellulaire soit globalement conservée entre toutes les espèces fongiques (composée d'un réseau polysaccharidique incluant principalement de la chitine, des glucanes, et du mannane/galactomannane), sa plasticité et sa composition dépendent de mécanismes de régulation actifs impliqués dans sa biosynthèse et sa restructuration (Dichtl et al., 2016). Ainsi, de nombreuses voies de signalisations sont impliquées dans l'intégrité de la paroi fongique en réponse à différentes conditions perturbatrices telles que celles vues précédemment (Dichtl et al., 2016).

La caféine, qui a induit une réponse phénotypique particulière chez nos souches, est une molécule de la famille des méthylxanthines qui agit comme analogue de faible affinité de l'adénosine. Cette molécule a la capacité d'inhiber de nombreuses enzymes incluant l'AMPc phosphodiesterase, et la protéine kinase mTOR (mammalian Target Of Rapamycin). La protéine kinase mTOR est un régulateur central de la croissance cellulaire en réponse à la disponibilité des nutriments aux facteurs de croissance et aux stimuli environnementaux (Hood-DeGrenier, 2011). L'interférence de la caféine avec la voie de signalisation « TOR » se fait par inhibition directe du complexe TORC1, ce qui entraîne de façon artificiel un signal de carence nutritive à la cellule (Homann et al., 2009). En plus du rôle de mTOR dans l'assimilation des nutriments, cette kinase est également impliquée dans la régulation de l'expression de plusieurs gènes spécifiques impliqués dans l'adhésion cellulaire et la formation des hyphes (Bastidas et al., 2009). En effet, il a été montré que Tor1 pouvait réguler négativement l'expression d'adhésines et par conséquent l'adhésion cellulaire ainsi que la croissance des hyphes en modulant l'activité de Bcr1, Efg1 et les répresseurs Nrg1 et Tup1 (Bastidas et al., 2009). En effet, chez *C. albicans*, l'expression des adhésines *ALS1*, *ALS3*, *HWP1* et *ECE1* est régulée par les facteurs de transcription Bcr1 et Tec1 et les répresseurs Nrg1 et Tup1. Ces gènes sont exclusivement exprimés au niveau des hyphes dont la régulation implique la voie de signalisation AMPc-PKA (protein kinase-A). Cette voie régule la morphogénèse de la levure par l'intermédiaire du facteur de transcription Efg1 en réponse à divers facteurs nutritionnels et environnementaux (Bastidas et al., 2009). En conditions nutritives favorables, Tor1 contrôle ainsi l'expression des adhésines par un double

mécanisme impliquant la répression des facteurs de transcriptions activateurs (Bcr1, Efg1) et l'activation des répresseurs (Nrg1 et Tup1) (Bastidas et al., 2009).

Enfin, le complexe TOR interagit également avec une autre voie essentielle, celle de la calcineurine. Celle-ci est importante dans la régulation de la réponse cellulaire impliquée dans la survie de *C. albicans* en présence d'inhibiteurs exogènes (Sanglard et al., 2003). En effet, la calcineurine est impliquée dans la morphogénèse des cellules, la virulence mais également dans la tolérance à différents toxiques tels que les antifongiques azolés (Mount et al., 2018; Sanglard et al., 2003). La stabilité de cette molécule est assurée par la protéine Hsp90 qui agit en aval sur divers effecteurs (Khandelwal et al., 2018). L'inhibition de l'activité d'Hsp90 par l'intermédiaire de Tor1 entraîne une hypersensibilité aux antifongiques azolés chez *C. albicans* (Shekhar-Guturja et al., 2016). Ainsi, l'hyperactivation de la voie de signalisation TOR peut être un mécanisme utilisé par *C. albicans* pour contourner la toxicité des antifongiques azolés, par l'intermédiaire de la voie de la calcineurine (Khandelwal et al., 2018). Ainsi, les phénotypes particuliers observés suggèrent qu'il pourrait y avoir un lien entre la voie de signalisation mTOR et l'évolution phénotypique des souches au cours du temps dans un contexte pathologique. L'exploration de cette voie de signalisation fera donc l'objet d'analyses complémentaires.

Afin d'identifier les altérations du génome qui pourraient avoir un lien avec les phénotypes observés chez nos patients, nous avons réalisé le séquençage des génomes complets de 37 souches provenant des patients CCMC et de leur famille que nous avons comparé aux génomes de souches contrôles. Pour cela, nous avons recherché les variations du nombre de copies des gènes (CNVs), les événements de pertes d'hétérozygotie (LOH), ainsi que les polymorphismes nucléotidiques (SNPs) pour l'ensemble des souches. Ces données ont ensuite été utilisées pour comparer (i) les génomes provenant des souches de patients CCMC avec les souches contrôles, (ii) les génomes de souches isolées sur plusieurs années chez un même individu et (iii) les génomes de souches isolées d'individus sains et malades d'une même famille qui partagent le même génotype MLST.

Cette analyse a révélé une importante diversité génomique entre les souches de patients CCMC mais également entre les souches d'un même individu.

En effet, bien que les souches soient extrêmement proches génotypiquement, plusieurs milliers de SNPs incluant des événements de LOH, ont été identifiés entre les génomes des souches issues d'un même malade. La diversité observée entre les souches était cependant du même ordre que celle observée entre les isolats issus d'un même prélèvement chez un individu sain. Une moindre diversité génomique entre les séries de souches séquentielles isolées dans notre contexte de patient atteint de CCMC était prévisible. En effet, nous posons l'hypothèse, dans ce travail, qu'un environnement pathologique de longue durée associé à l'utilisation d'antifongiques pouvait induire une pression de sélection et possiblement une adaptation des souches pour survivre dans ce contexte.

Ce travail nous a néanmoins permis d'identifier des altérations du génome pouvant être la conséquence d'une adaptation spécifique des souches dans un contexte d'infection chronique. En effet, de nombreux gènes impliqués dans l'adhésion cellulaire ou dans des processus faisant intervenir des protéines de paroi ont été identifiés au cours de cette étude tels que des gènes de la famille des *ALS*, *ECE1*, *DAL5*, *OCA1*, *GAT1*... . Parmi ces gènes, 3 ont montré un lien direct ou indirect avec la voie de signalisation mTOR et par conséquent le phénotype de résistance à la caféine. Ces gènes constituent des pistes intéressantes pour l'étude de l'évolution génomique de *C. albicans* au cours de l'interaction pathologique prolongée chez l'Homme. Une étude plus approfondie sur l'imputation de certains gènes dans l'adaptation de *C. albicans* fera l'objet de la suite de ce travail.

Enfin, nous avons également cherché à identifier des corrélations entre la survenue de certains phénotypes et la présence de génotypes particuliers. Cependant, il n'a pas été possible de rapprocher les données génomiques des phénotypes étudiés tels que la filamentation ou la réponse au stress impactant l'intégrité de la paroi fongique. Une des explications pourrait résider dans la complexité de ces phénomènes multifactoriels, multigéniques et impliquant de nombreuses voies de régulations. Par exemple, la capacité des souches à filamenter dépend non seulement du milieu et de la température d'incubation, mais également de différents facteurs tels que la disponibilité des nutriments, la pression en oxygène, la production de CO₂, la production/détection et réponse aux molécules de quorum sensing (Sudbery, 2011). Par exemple, une mutation, qui n'affecte pas directement le processus de croissance de l'hyphe, peut néanmoins l'affecter indirectement à la suite d'une modification dans l'un de ces signaux micro-environnementaux.

Par ailleurs, d'autres mécanismes tels que des modifications transcriptionnelles ou post transcriptionnelles voire d'épigénétique peuvent également être impliquées dans les variations phénotypiques observées (Freire-Benítez et al., 2016). La complexité de l'intervention de ces différents facteurs a rendu difficile la mise en évidence de corrélations génotype-phénotype mais constitue un objectif primordial pour l'étude de l'évolution de la pathogénicité de ces souches au cours de l'infection à long terme chez l'Homme.

En revanche, un phénotype extrêmement bien caractérisé dans la littérature a été exploré plus en détail et constitue la dernière partie de ce travail. Celui-ci concerne la survenue de la résistance phénotypique des souches aux antifongiques azolés et échinocandines.

4 Analyse des mécanismes de résistance aux antifongiques

Un autre volet intéressant de l'analyse de l'évolution génomique des souches provenant de patients atteints de CCMC consiste en l'étude approfondie des mécanismes de résistance aux antifongiques de ces souches. En effet, les patients présentant une CCMC sont pour la plupart traités de façon prolongée et répétée par divers antifongiques (azolés ou échinocandines). L'utilisation de ces molécules engendre une augmentation progressive de la résistance de ces isolats et peut être responsable d'échec thérapeutique chez ces patients. Parmi les 37 génomes séquencés, 22 provenaient d'isolats résistants à un ou plusieurs antifongiques. Ces isolats ont été obtenus chronologiquement chez 5 de nos patients atteints de CCMC (P1 à P5) (**Tableau 12**). Grâce aux données de séquençage à haut débit nous avons pu caractériser les remaniements génomiques potentiellement impliqués dans la résistance de ces souches aux principaux antifongiques utilisés en clinique.

De façon surprenante, aucun événement d'aneuploïdie (Trisomie, duplication segmentale ou isochromosome) n'a été observé chez ces souches. En revanche 18/22 (82%) des souches ont présenté un grand événement de LOH impliquant la quasi-totalité du bras gauche du chromosome 5 sur lequel se trouvent les deux principaux gènes impliqués dans la résistance aux azolés (*TAC1* et *ERG11*). Ces souches provenaient des patients P1 à P4 et correspondaient aux souches présentant un haut niveau de résistance aux fluconazole (CMI > 256 µg/ml). Chez le patient P3, la survenue de cet événement de LOH au cours du temps s'est accompagnée d'une augmentation importante des CMI au fluconazole (**Tableau 12**). Cependant, de façon intéressante, un grand événement de LOH au niveau du Chr5 a également été observé sur les 4 souches, sensibles aux antifongiques azolés, isolées chez le patient P9 (**Figure 31**).

Afin de caractériser en détail les mécanismes impliqués dans la résistance phénotypique de ces souches, nous avons analysé les séquences des principaux gènes impliqués dans la résistance aux azolés (*ERG11*, *TAC1*, *MRR1* et *UPC2*) et aux échinocandines (*FKS1*). Cependant, ces gènes présentent un niveau élevé de polymorphisme.

Afin de différencier les mutations reflétant le polymorphisme naturel de celles conférant une résistance, nous avons préalablement réalisé un répertoire des mutations de polymorphisme naturel observées dans ces différents gènes d'intérêts, en utilisant les données des séquences génomiques de 151 souches (collection de souches contrôles) de *C. albicans* sensibles aux différents antifongiques. Ce répertoire a été utilisé pour l'analyse des 22 souches résistantes aux azolés et/ou aux échinocandines issues des patients CCMC. Cette démarche nous a permis d'identifier, en plus du polymorphisme naturel, 19 mutations non-synonymes dans les gènes *ERG11* (n = 11), *TAC1* (n = 5), *UPC2* (n = 2) et *FKS1* (n = 1) (**Tableau 12**). Neuf mutations avaient déjà été décrites dans la littérature, dont une très récemment, comme associées à un phénotype de résistance aux azolés ou aux échinocandines (Coste et al., 2009; Flowers et al., 2012; Lackner et al., 2014; Löffler et al., 1997; Sanglard et al., 1998; Sanglard and Bille, 2002; Wang et al., 2009; White, 1997; Wu et al., 2017). Concernant les 10 nouvelles mutations, 6 ont été introduites par mutagenèse dirigée (5 dans *TAC1* et 1 dans *UPC2*) dans le génome d'une souche sauvage. Nous avons pu confirmer pour 3 d'entre elles, localisées sur le gène *TAC1* (induisant les substitutions protéiques H263Y, H838Y et P971S), leur implication dans la résistance phénotypique (augmentation significative de la CMI au fluconazole allant de 5,3 à plus de 85 fois par rapport à l'introduction de l'allèle sauvage). Concernant les 4 nouvelles mutations détectées dans le gène *ERG11*, leur implication potentielle dans la résistance *in vitro* a été évaluée en utilisant différents logiciels de prédiction d'impact mutationnel (PoPMuSiCv3.1, HoTMuSiCv0.5 ainsi que SNPMuSiC). Parmi ces quatre mutations, deux (Y64H, P236S) ont été prédites comme ayant un impact sur la structure de la protéine et peuvent donc être également impliquées dans la résistance *in vitro* de ces souches.

La création de ce répertoire réalisé à partir d'un grand nombre d'isolats sensibles de *C. albicans* a facilité l'analyse des séquences de souches résistantes aux antifongiques (azolés et échinocandines). Ainsi, l'utilisation de ce répertoire a permis de sélectionner un nombre restreint de nouvelles mutations potentiellement impliquées dans la résistance aux azolés, que nous avons testées par mutagenèse et pour lesquelles plusieurs ont pu être confirmées.

D'autres gènes, potentiellement impliqués dans la tolérance/résistance aux antifongiques azolés (tels que *ERG3*, *ERG6*, *RGD1*, *PEP8*, *NRP2*, *HSP90...*), sont en cours d'exploration afin de caractériser au mieux les mécanismes de résistance survenant chez les patients traités à long terme par ces molécules.

Patient	Isolat	CMI (µg/ml)		Substitution en acide aminé				Remaniement chromosomique
		FLZ	AND	Erg11p	Tac1p	Upc2p	Fk1p	
P1	FLOJ14_B_2007	>256	0,008	Y132H, G450E	H263Y, P971S	G648D ^h		LOH Chr5
P1	FLOJ19_V_2008	>256	0,012	Y132H, G450E	H263Y, P971S			LOH Chr5
P1	FLOJ20_C_2008	>256	0,016	Y132H, G450E	H263Y, P971S			LOH Chr5
P1	FLOJ22_B_2008	>256	0,016	Y132H, G450E	H263Y, P971S			LOH Chr5
P1	FLOJ23_B_2009	>256	0,016	Y132H, G450E	H263Y, P971S			LOH Chr5
P1	FLOJ12_B_2014	>256	0,032	Y132H, G450E	H263Y, P971S			LOH Chr5
P2	FLOJ13_O_2007	>256	0,004	Y132H, R467I	H263Y, P971S	D654N ^h		LOH Chr5
P2	FLOJ17_C_2008	>256	0,008	Y132H, R467I	H263Y, P971S	D654N		LOH Chr5
P2	FLOJ02_C_2013	>256	0,125	Y64H , T123I, Y132H, R467I	H263Y, P971S	D654N ^h	R647G	LOH Chr5
P2	FLOJ07_B_2013	>256	0,19	Y64H , T123I, Y132H, R467I	H263Y, P971S	D654N ^h	R647G	LOH Chr5
P2	FLOJ68_B_2015	>256	0,047	Y64H , T123I, Y132H, R467I	H263Y, P971S	D654N ^h	R647G	LOH Chr5
P3	BLOV27_B_2008	24	0,008	R157K ^h	H263Y, H838Y, C858Y			
P3	BLOV29_B_2009	48	0,008	R157K ^h	H263Y, H838Y, C858Y			
P3	BLOV32_B_2011	>256	0,008	R157K	H263Y, H838Y, C858Y			LOH Chr5
P3	BLOV37_B_2012	>256	0,012	R157K	H263Y, H838Y, C858Y			LOH Chr5
P3	BLOV03_B_2014	>256	0,006	R157K	H263Y, H838Y, C858Y			LOH Chr5
P3	BLOV56_V_2014	>256	0,012	R157K	H263Y, H838Y, C858Y			LOH Chr5
P3	BLOV53_B_2014	>256	0,008	R157K	H263Y, H838Y, C858Y			LOH Chr5
P3	BLOV67_O_2015	>256	0,064	R157K	H263Y, H838Y, C858Y			LOH Chr5
P4	BLOG54_B_2014	>256	0,008	D9V ^h , R157K	H263Y, H838Y, C858Y			LOH Chr5
P5	HAME38_B_2012	8	0,003	G448E	A736T, N823K, C858Y			
P5	HAME10_B_2013	8	0,004	G448E	A736T, N823K, C858Y			

Tableau 12 : Caractéristiques des 22 souches de *C. albicans* issues des patients CCMC présentant une résistance acquise aux azolés et/ou aux échinocandines. Description des résultats des CMI (Concentration Minimales Inhibitrices) au fluconazole (FLZ) et à l'anidulafungine (AND) associées aux remaniements chromosomiques et aux substitutions en acide aminé dans les facteurs de transcription Tac1p et Upc2p et les protéines Erg11p et Fks1p. Seules les mutations précédemment décrites comme étant impliquées dans la résistance aux antifongiques et les nouvelles mutations (notées en gras) détectées grâce au répertoire de polymorphisme naturel sont inscrites dans cette table. En rouge sont notées les mutations qui ont été confirmées, dans ce travail, par mutagenèse. Le h correspond aux mutations retrouvées uniquement à l'état hétérozygote.

Publication 2 : *A Useful Tool to Highlight New Resistance Mutations in Genes Involved in Candida albicans Resistance to Azoles and Echinocandines. (Manuscrit en cours de finalisation) (Annexe 2). Emilie Sitterlé, Alixe Coste, Thomes Obadia, Corinne Maufrais, Murielle Chauvel, Natacha Sertour, Anne Puel, Dominique Sanglard, Christophe d'Enfert, Marie-Elisabeth Bougnoux.*

Abstract

Background: Genome of *Candida albicans* is highly variable inducing a high level of gene polymorphism. Point mutations in genes involved in resistance to antifungal may either confer phenotypic resistance or be natural polymorphisms.

Objectives: Establish a comprehensive repertoire of natural polymorphisms in genes involved in azoles and echinocandins resistance to help identify resistance mutations in clinical isolates.

Methods: We took advantage of genome sequencing data from 151 unrelated *C. albicans* strains susceptible to antifungal agents. We analyzed the sequences of genes involved in resistance to azoles (*ERG11*, *TAC1*, *MRR1*, *UPC2*) and echinocandins (*FKS1*) and compared them to the reference sequences of SC5314 strain. We tested then the capacity of the repertoire to highlight resistance mutations in 10 resistant isolates. Selected putative new resistant mutations were analyzed by transcript level measurements of *CDR2* and *ERG11* with RT-qPCR and by mutagenesis experiments or mutational prediction effect.

Results: The repertoire includes 126 amino-acid substitutions. The final rate of new mutations appearance per strain was less than one. Using this repertoire, we identified from the 10 resistant strains, 22 additional substitutions. Among them, 10 have already been associated to resistance. The remaining 12 substitutions affected Tac1p (n=6/7), Erg11p (n=4/11) and Upc2p (n=2/3). Eight were tested by mutagenesis (6 in Tac1p and 2 in Upc2p). Four substitutions in Tac1p (H263Y, A790V, H839Y and P971S) confer a high fold change increase of the fluconazole MIC. Additionally, 2 substitutions in Erg11p (Y64H, P236S) were predicted as deleterious.

Conclusions: This repertoire is a powerful tool helping to rapidly unveil mutations linked to antifungal resistance.

Conclusions-Perspectives

Au cours de l'interaction avec l'hôte, *C. albicans* est confronté à un grand nombre de stress environnementaux et immunitaires imposant des capacités d'adaptation rapides pour survivre à ces environnements changeants. La plasticité importante du génome de *C. albicans* est un des éléments clé lui permettant de générer de façon rapide de la diversité génomique. De nombreux événements de remaniements génomiques, incluant des aneuploïdies ou des pertes d'hétérozygotie (LOH), ont été observés au sein de diverses souches de *C. albicans*. La survenue de ces événements a largement été étudiée *in vitro* et *in vivo*, essentiellement chez la souris, en réponse à diverses modifications environnementales. En revanche, hormis les études réalisées par des approches de typage moléculaire, très peu de travaux se sont portés sur la diversité génétique de *C. albicans* et les modifications du génome survenant de façon naturelle au cours de l'interaction chez l'Homme, que ce soit durant le commensalisme ou au cours de l'infection. Dans ce travail, nous avons souhaité approfondir nos connaissances concernant la diversité génétique de *C. albicans* chez l'individu sain mais également au cours de l'interaction pathologique de longue durée.

Dans une première partie de ce travail nous avons pu montrer qu'il existait un niveau élevé d'hétérogénéité génétique dans la population cellulaire de *C. albicans* dans des prélèvements buccaux de porteurs sains et démontré que cette hétérogénéité n'était pas le reflet de problèmes inhérents à la technique et à l'analyse des données de séquençage à haut débit. Une telle diversité génétique pourrait favoriser l'adaptation rapide de *C. albicans* en condition de stress lié à un changement d'environnement. L'évaluation et la quantification de la diversité génétique basale de *C. albicans* au cours du commensalisme était une information à prendre en compte pour les études d'évolution génomique qui sont habituellement réalisées à partir d'une colonie unique par échantillon. En effet, dans la suite de notre projet nous avons souhaité évaluer dans quelle mesure *C. albicans* était capable d'évoluer phénotypiquement et génétiquement, cette fois-ci dans un contexte pathologique de longue durée grâce à l'étude de souches provenant de patients atteints de CCMC.

La connaissance du niveau basal d'hétérogénéité au sein d'un individu nous a ainsi permis d'évaluer et d'interpréter la quantification des variations génétiques observées au cours du temps entre les génomes proches provenant de nos séries chronologiques d'isolats.

Bien que la caractérisation génotypique des souches par typage MLST ait montré que les patients étaient colonisés par une souche a priori unique de *C. albicans* (même profil MLST ou très proche), l'analyse phénotypique des souches dans de nombreuses conditions a montré une importante diversité. Celle-ci a été observée entre les souches des différents patients mais également au sein d'un même patient. En effet, nous avons montré que certains patients pouvaient être colonisés avec des souches présentant des phénotypes très différents dont certaines peuvent persister au cours du temps ou être éliminées. La mise en évidence aléatoire de certains phénotypes peut résulter de la présence en faible proportion de l'isolat atypique au sein de la population cellulaire de *C. albicans*, ou alors s'expliquer par l'élimination de l'isolat du fait d'un désavantage sélectif (défaut de croissance par exemple). Ces résultats suggèrent qu'il existe une importante dynamique dans l'apparition et l'élimination de certains variants phénotypiques au cours de l'interaction pathologique à long terme chez l'hôte. Par ailleurs, nous avons pu observer chez certains patients la survenue de modifications phénotypiques importantes des souches pouvant persister au cours du temps. Ces résultats montrent que les souches sont également capables d'évoluer phénotypiquement au cours de l'interaction pathologique chez l'Homme. Une étude phénotypique des isolats obtenus au cours du commensalisme pourrait permettre d'évaluer la pertinence de ces observations.

En parallèle, l'analyse comparative des génomes des souches de patients CCMC a montré un nombre important de variations génomiques entre les souches des différents patients mais également au sein des séries de souches chronologiques. Nous avons montré que des modifications génomiques à grande échelle (LOH) pouvaient survenir de façon aléatoire sur les génomes mais également se fixer et ainsi être conservées au cours du temps. Bien qu'un grand nombre de ces variants puisse être en rapport avec la diversité des souches existante au sein d'un échantillon, la survenue puis la persistance d'événements spécifiques (LOH ou SNPs) peuvent être la conséquence de l'adaptation des souches au cours de l'interaction pathologique chronique.

Nous avons donc, au cours de cette étude, identifié les gènes impactés par les modifications génomiques (LOH et SNPs) retrouvées dans le génome des souches de différents patients ou chez un même patient. Il est cependant nécessaire de déterminer l'imputation de ces gènes dans la survenue des « switch phénotypiques » observés dans cette étude telles que l'augmentation de la résistance des souches aux antifongiques ou à la caféine ainsi que les modifications de la capacité à filamenter ou lyser les kératinocytes. Un certain nombre de gènes (*DAL5*, *OCA1*, *TPS2*, *ECE1*, *TAC1*, *ERG11*, *UPC2*, *FKS1*...) ont déjà été analysés plus spécifiquement. Cependant, seules les mutations dans les gènes impliqués dans la résistance aux antifongiques ont permis d'établir de vraies corrélations génotypes/phénotypes pour nos souches. L'analyse des gènes connus pour être impliqués dans la résistance aux antifongiques azolés et aux échinocandines nous a permis de décrire plusieurs nouvelles mutations dans les gènes *ERG11* et *TAC1*. Parmi celles-ci, nous avons pu prouver que 3 sont impliquées dans la résistance aux antifongiques azolés.

Cependant, l'analyse de gènes supplémentaires dont la mutation pourrait expliquer la survenue des « switch phénotypiques », fera l'objet d'explorations ultérieures. Celle-ci sera associée à des approches d'études de corrélation génotype/phénotype à grande échelle de type GWAS (Genome-Wide Association) nécessitant un grand nombre de souches associé à un grand nombre de phénotypes pour parvenir à de telles corrélations.

Références bibliographiques

- Abbey, D., Hickman, M., Gresham, D., Berman, J., 2011. High-Resolution SNP/CGH Microarrays Reveal the Accumulation of Loss of Heterozygosity in Commonly Used *Candida albicans* Strains. *G3* 1, 523–530. <https://doi.org/10.1534/g3.111.000885>
- Abbey, D.A., Funt, J., Lurie-Weinberger, M.N., Thompson, D.A., Regev, A., Myers, C.L., Berman, J., 2014. YMAP: a pipeline for visualization of copy number variation and loss of heterozygosity in eukaryotic pathogens. *Genome Med.* 6. <https://doi.org/10.1186/s13073-014-0100-8>
- Alic, N., Higgins, V.J., Dawes, I.W., 2001. Identification of a *Saccharomyces cerevisiae* gene that is required for G1 arrest in response to the lipid oxidation product linoleic acid hydroperoxide. *Mol. Biol. Cell* 12, 1801–1810. <https://doi.org/10.1091/mbc.12.6.1801>
- Angebault, C., Djossou, F., Abélanet, S., Permal, E., Ben Soltana, M., Diancourt, L., Bouchier, C., Woerther, P.-L., Catzeflis, F., Andreumont, A., d'Enfert, C., Bournonville, M.-E., 2013. *Candida albicans* is not always the preferential yeast colonizing humans: a study in Wayampi Amerindians. *J. Infect. Dis.* 208, 1705–1716. <https://doi.org/10.1093/infdis/jit389>
- Arnaud, M.B., Costanzo, M.C., Skrzypek, M.S., Binkley, G., Lane, C., Miyasato, S.R., Sherlock, G., 2005. The *Candida* Genome Database (CGD), a community resource for *Candida albicans* gene and protein information. *Nucleic Acids Res.* 33, D358-363. <https://doi.org/10.1093/nar/gki003>
- Bastidas, R.J., Heitman, J., Cardenas, M.E., 2009. The Protein Kinase Tor1 Regulates Adhesin Gene Expression in *Candida albicans*. *PLoS Pathog.* 5, e1000294. <https://doi.org/10.1371/journal.ppat.1000294>
- Bates, S., de la Rosa, J.M., MacCallum, D.M., Brown, A.J.P., Gow, N.A.R., Odds, F.C., 2007. *Candida albicans* Iff11, a secreted protein required for cell wall structure and virulence. *Infect. Immun.* 75, 2922–2928. <https://doi.org/10.1128/IAI.00102-07>
- Beck, T., Hall, M.N., 1999. The TOR signalling pathway controls nuclear localization of nutrient-regulated transcription factors. *Nature* 402, 689–692. <https://doi.org/10.1038/45287>
- Bennett, R.J., 2015. The parasexual lifestyle of *Candida albicans*. *Curr. Opin. Microbiol.* 28, 10–17. <https://doi.org/10.1016/j.mib.2015.06.017>
- Bennett, R.J., Forche, A., Berman, J., 2014. Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harb. Perspect. Med.* 4. <https://doi.org/10.1101/cshperspect.a019604>
- Bennett, R.J., Johnson, A.D., 2003. Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. *EMBO J.* 22, 2505–2515. <https://doi.org/10.1093/emboj/cdg235>
- Bennett, S., 2004. *Sox1*. *Pharmacogenomics* 5, 433–438. <https://doi.org/10.1517/14622416.5.4.433>
- Berman, J., Forche, A., 2017. Haplotyping a Non-meiotic Diploid Fungal Pathogen Using Induced Aneuploidies and SNP/CGH Microarray Analysis. *Methods Mol. Biol. Clifton NJ* 1551, 131–146. https://doi.org/10.1007/978-1-4939-6750-6_7
- Berman, J., Hadany, L., 2012. Does stress induce (para)sex? Implications for *Candida albicans* evolution. *Trends Genet.* 28, 197–203. <https://doi.org/10.1016/j.tig.2012.01.004>
- Binkley, J., Arnaud, M.B., Inglis, D.O., Skrzypek, M.S., Shah, P., Wymore, F., Binkley, G., Miyasato, S.R., Simison, M., Sherlock, G., 2014. The *Candida* Genome Database: the new homology information page highlights protein similarity and phylogeny. *Nucleic Acids Res.* 42, D711-716. <https://doi.org/10.1093/nar/gkt1046>

- Blignaut, E., Molepo, J., Pujol, C., Soll, D.R., Pfaller, M.A., 2005. Clade-related amphotericin B resistance among South African *Candida albicans* isolates. *Diagn. Microbiol. Infect. Dis.* 53, 29–31. <https://doi.org/10.1016/j.diagmicrobio.2005.03.013>
- Boisson, B., Wang, C., Pedergnana, V., Wu, L., Cypowyj, S., Rybojad, M., Belkadi, A., Picard, C., Abel, L., Fieschi, C., Puel, A., Li, X., Casanova, J.-L., 2013. An ACT1 Mutation Selectively Abolishes Interleukin-17 Responses in Humans with Chronic Mucocutaneous Candidiasis. *Immunity* 39, 676–686. <https://doi.org/10.1016/j.immuni.2013.09.002>
- Bouchonville, K., Forche, A., Tang, K.E.S., Selmecki, A., Berman, J., 2009. Aneuploid chromosomes are highly unstable during DNA transformation of *Candida albicans*. *Eukaryot. Cell* 8, 1554–1566. <https://doi.org/10.1128/EC.00209-09>
- Bougnoux, M.-E., Aanensen, D.M., Morand, S., Théraud, M., Spratt, B.G., d’Enfert, C., 2004. Multilocus sequence typing of *Candida albicans*: strategies, data exchange and applications. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 4, 243–252. <https://doi.org/10.1016/j.meegid.2004.06.002>
- Bougnoux, M.-E., Diogo, D., François, N., Sendid, B., Veirmeire, S., Colombel, J.F., Bouchier, C., Van Kruiningen, H., d’Enfert, C., Poulain, D., 2006. Multilocus sequence typing reveals intrafamilial transmission and microevolutions of *Candida albicans* isolates from the human digestive tract. *J. Clin. Microbiol.* 44, 1810–1820. <https://doi.org/10.1128/JCM.44.5.1810-1820.2006>
- Bougnoux, M.-E., Morand, S., d’Enfert, C., 2002. Usefulness of multilocus sequence typing for characterization of clinical isolates of *Candida albicans*. *J. Clin. Microbiol.* 40, 1290–1297.
- Bougnoux, M.-E., Pujol, C., Diogo, D., Bouchier, C., Soll, D.R., d’Enfert, C., 2008. Mating is rare within as well as between clades of the human pathogen *Candida albicans*. *Fungal Genet. Biol. FG B* 45, 221–231. <https://doi.org/10.1016/j.fgb.2007.10.008>
- Bougnoux, M.-E., Tavanti, A., Bouchier, C., Gow, N.A.R., Magnier, A., Davidson, A.D., Maiden, M.C.J., d’Enfert, C., Odds, F.C., 2003. Collaborative Consensus for Optimized Multilocus Sequence Typing of *Candida albicans*. *J. Clin. Microbiol.* 41, 5265–5266. <https://doi.org/10.1128/JCM.41.11.5265-5266.2003>
- Braun, B.R., van Het Hoog, M., d’Enfert, C., Martchenko, M., Dungan, J., Kuo, A., Inglis, D.O., Uhl, M.A., Hogues, H., Berriman, M., Lorenz, M., Levitin, A., Oberholzer, U., Bachewich, C., Harcus, D., Marcil, A., Dignard, D., Iouk, T., Zito, R., Frangeul, L., Tekaia, F., Rutherford, K., Wang, E., Munro, C.A., Bates, S., Gow, N.A., Hoyer, L.L., Köhler, G., Morschhäuser, J., Newport, G., Znaidi, S., Raymond, M., Turcotte, B., Sherlock, G., Costanzo, M., Ihmels, J., Berman, J., Sanglard, D., Agabian, N., Mitchell, A.P., Johnson, A.D., Whiteway, M., Nantel, A., 2005. A human-curated annotation of the *Candida albicans* genome. *PLoS Genet.* 1, 36–57. <https://doi.org/10.1371/journal.pgen.0010001>
- Braunsdorf, C., LeibundGut-Landmann, S., 2018. Modulation of the Fungal-Host Interaction by the Intra-Species Diversity of *C. albicans*. *Pathog. Basel Switz.* 7. <https://doi.org/10.3390/pathogens7010011>
- Brown, G.D., Denning, D.W., Gow, N.A.R., Levitz, S.M., Netea, M.G., White, T.C., 2012. Hidden Killers: Human Fungal Infections. *Sci. Transl. Med.* 4, 165rv13-165rv13. <https://doi.org/10.1126/scitranslmed.3004404>
- Butler, G., Rasmussen, M.D., Lin, M.F., Santos, M.A.S., Sakthikumar, S., Munro, C.A., Rheinbay, E., Grabherr, M., Forche, A., Reedy, J.L., Agrafioti, I., Arnaud, M.B., Bates, S., Brown, A.J.P., Brunke, S., Costanzo, M.C., Fitzpatrick, D.A., de Groot, P.W.J., Harris, D., Hoyer, L.L., Hube, B., Klis, F.M., Kodira, C., Lennard, N., Logue, M.E., Martin, R., Neiman, A.M., Nikolaou, E., Quail, M.A., Quinn, J., Santos, M.C., Schmitzberger, F.F., Sherlock, G., Shah, P., Silverstein, K.A.T., Skrzypek, M.S., Soll, D., Staggs, R., Stansfield, I., Stumpf, M.P.H., Sudbery, P.E., Srikantha, T., Zeng, Q., Berman, J., Berriman, M., Heitman, J., Gow, N.A.R., Lorenz, M.C., Birren, B.W., Kellis, M., Cuomo, C.A., 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* 459, 657–662. <https://doi.org/10.1038/nature08064>
- Calderón-Noreña, D.M., González-Novo, A., Orellana-Muñoz, S., Gutiérrez-Escribano, P., Arnáiz-Pita, Y., Dueñas-Santero, E., Suárez, M.B., Bougnoux, M.-E., Del Rey, F., Sherlock, G., d’Enfert, C., Correa-Bordes, J., de Aldana,

- C.R.V., 2015. A single nucleotide polymorphism uncovers a novel function for the transcription factor Ace2 during *Candida albicans* hyphal development. *PLoS Genet.* 11, e1005152. <https://doi.org/10.1371/journal.pgen.1005152>
- Carr, L.L., Gottschling, D.E., 2008. Does age influence loss of heterozygosity? *Exp. Gerontol.* 43, 123–129. <https://doi.org/10.1016/j.exger.2007.10.010>
- Cavaliere, D., Di Paola, M., Rizzetto, L., Tocci, N., De Filippo, C., Lionetti, P., Ardizzoni, A., Colombari, B., Paulone, S., Gut, I.G., Berná, L., Gut, M., Blanc, J., Kapushesky, M., Pericolini, E., Blasi, E., Peppoloni, S., 2017. Genomic and Phenotypic Variation in Morphogenetic Networks of Two *Candida albicans* Isolates Subtends Their Different Pathogenic Potential. *Front. Immunol.* 8, 1997. <https://doi.org/10.3389/fimmu.2017.01997>
- Chandrasekar, P.H., Sobel, J.D., 2006. Micafungin: a new echinocandin. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 42, 1171–1178. <https://doi.org/10.1086/501020>
- Chávez-Galarza, J., Pais, C., Sampaio, P., 2010. Microsatellite typing identifies the major clades of the human pathogen *Candida albicans*. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* 10, 697–702. <https://doi.org/10.1016/j.meegid.2010.03.007>
- Cheng, S.-C., Joosten, L.A.B., Kullberg, B.-J., Netea, M.G., 2012. Interplay between *Candida albicans* and the mammalian innate host defense. *Infect. Immun.* 80, 1304–1313. <https://doi.org/10.1128/IAI.06146-11>
- Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G., Hirschman, J.E., Hitz, B.C., Karra, K., Krieger, C.J., Miyasato, S.R., Nash, R.S., Park, J., Skrzypek, M.S., Simison, M., Weng, S., Wong, E.D., 2012. *Saccharomyces* Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* 40, D700–705. <https://doi.org/10.1093/nar/gkr1029>
- Chibana, H., Beckerman, J.L., Magee, P.T., 2000. Fine-resolution physical mapping of genomic diversity in *Candida albicans*. *Genome Res.* 10, 1865–1877.
- Chibana, H., Magee, B.B., Grindle, S., Ran, Y., Scherer, S., Magee, P.T., 1998. A physical map of chromosome 7 of *Candida albicans*. *Genetics* 149, 1739–1752.
- Chindamporn, A., Nakagawa, Y., Homma, M., Chibana, H., Doi, M., Tanaka, K., 1995. Analysis of the chromosomal localization of the repetitive sequences (RPSs) in *Candida albicans*. *Microbiology* 141, 469–476. <https://doi.org/10.1099/13500872-141-2-469>
- Chindamporn, A., Nakagawa, Y., Mizuguchi, I., Chibana, H., Doi, M., Tanaka, K., 1998. Repetitive sequences (RPSs) in the chromosomes of *Candida albicans* are sandwiched between two novel stretches, HOK and RB2, common to each chromosome. *Microbiology* 144, 849–857. <https://doi.org/10.1099/00221287-144-4-849>
- Chowdhary, A., Hagen, F., Sharma, C., Al-Hatmi, A.M.S., Giuffrè, L., Giosa, D., Fan, S., Badali, H., Felice, M.R., de Hoog, S., Meis, J.F., Romeo, O., 2017. Whole Genome-Based Amplified Fragment Length Polymorphism Analysis Reveals Genetic Diversity in *Candida africana*. *Front. Microbiol.* 8. <https://doi.org/10.3389/fmicb.2017.00556>
- Chu, W.S., Magee, B.B., Magee, P.T., 1993. Construction of an *Sfi*I macrorestriction map of the *Candida albicans* genome. *J. Bacteriol.* 175, 6637–6651.
- Cleveland, A.A., Harrison, L.H., Farley, M.M., Hollick, R., Stein, B., Chiller, T.M., Lockhart, S.R., Park, B.J., 2015. Declining Incidence of Candidemia and the Shifting Epidemiology of *Candida* Resistance in Two US Metropolitan Areas, 2008–2013: Results from Population-Based Surveillance. *PLOS ONE* 10, e0120452. <https://doi.org/10.1371/journal.pone.0120452>
- Coste, A., Turner, V., Ischer, F., Morschhäuser, J., Forche, A., Selmecki, A., Berman, J., Bille, J., Sanglard, D., 2006. A mutation in *Tac1p*, a transcription factor regulating *CDR1* and *CDR2*, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in *Candida albicans*. *Genetics* 172, 2139–2156. <https://doi.org/10.1534/genetics.105.054767>

- Coste, A.T., Crittin, J., Bauser, C., Rohde, B., Sanglard, D., 2009. Functional Analysis of cis- and trans-Acting Elements of the *Candida albicans* CDR2 Promoter with a Novel Promoter Reporter System. *Eukaryot. Cell* 8, 1250–1267. <https://doi.org/10.1128/EC.00069-09>
- Cottier, F., Pavelka, N., 2012. Complexity and dynamics of host-fungal interactions. *Immunol. Res.* 53, 127–135. <https://doi.org/10.1007/s12026-012-8265-y>
- d’Enfert, C., Bounoux, M.-E., Feri, A., Legrand, M., Loll-Krippelber, R., Marton, T., Maufrais, C., Ropars, J., Sertour, N., Sitterlé, E., 2017. Genome Diversity and Dynamics in *Candida albicans*, in: Prasad, R. (Ed.), *Candida Albicans: Cellular and Molecular Biology*. Springer International Publishing, Cham, pp. 205–232. https://doi.org/10.1007/978-3-319-50409-4_11
- Dalle, F., Franco, N., Lopez, J., Vagner, O., Caillot, D., Chavanet, P., Cuisenier, B., Aho, S., Lizard, S., Bonnin, A., 2000. Comparative genotyping of *Candida albicans* bloodstream and nonbloodstream isolates at a polymorphic microsatellite locus. *J. Clin. Microbiol.* 38, 4554–4559.
- Darwin, C., 1859. *L’Origine des espèces*.
- de Hoon, M.J.L., Imoto, S., Nolan, J., Miyano, S., 2004. Open source clustering software. *Bioinforma. Oxf. Engl.* 20, 1453–1454. <https://doi.org/10.1093/bioinformatics/bth078>
- Deresinski, S.C., Stevens, D.A., 2003. Caspofungin. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 36, 1445–1457. <https://doi.org/10.1086/375080>
- Deva, R., Ciccoli, R., Kock, L., Nigam, S., 2001. Involvement of aspirin-sensitive oxylipins in vulvovaginal candidiasis. *FEMS Microbiol. Lett.* 198, 37–43.
- Deva, R., Ciccoli, R., Schewe, T., Kock, J.L., Nigam, S., 2000. Arachidonic acid stimulates cell growth and forms a novel oxygenated metabolite in *Candida albicans*. *Biochim. Biophys. Acta* 1486, 299–311.
- Dichtl, K., Samantaray, S., Wagener, J., 2016. Cell wall integrity signalling in human pathogenic fungi: CWI signalling in pathogenic fungi. *Cell. Microbiol.* 18, 1228–1238. <https://doi.org/10.1111/cmi.12612>
- Diogo, D., Bouchier, C., d’Enfert, C., Bounoux, M.-E., 2009. Loss of heterozygosity in commensal isolates of the asexual diploid yeast *Candida albicans*. *Fungal Genet. Biol.* 46, 159–168. <https://doi.org/10.1016/j.fgb.2008.11.005>
- Dodgson, A.R., Dodgson, K.J., Pujol, C., Pfaller, M.A., Soll, D.R., 2004. Clade-specific flucytosine resistance is due to a single nucleotide change in the *FUR1* gene of *Candida albicans*. *Antimicrob. Agents Chemother.* 48, 2223–2227. <https://doi.org/10.1128/AAC.48.6.2223-2227.2004>
- Dumitru, R., Navarathna, D.H.M.L.P., Semighini, C.P., Elowsky, C.G., Dumitru, R.V., Dignard, D., Whiteway, M., Atkin, A.L., Nickerson, K.W., 2007. In vivo and in vitro anaerobic mating in *Candida albicans*. *Eukaryot. Cell* 6, 465–472. <https://doi.org/10.1128/EC.00316-06>
- Dunkel, N., Blaß, J., Rogers, P.D., Morschhäuser, J., 2008. Mutations in the multi-drug resistance regulator *MRR1*, followed by loss of heterozygosity, are the main cause of *MDR1* overexpression in fluconazole-resistant *Candida albicans* strains. *Mol. Microbiol.* 69, 827–840. <https://doi.org/10.1111/j.1365-2958.2008.06309.x>
- Edmond, M.B., Wallace, S.E., McClish, D.K., Pfaller, M.A., Jones, R.N., Wenzel, R.P., 1999. Nosocomial Bloodstream Infections in United States Hospitals: A Three-Year Analysis. *Clin. Infect. Dis.* 29, 239–244. <https://doi.org/10.1086/520192>
- Eggimann, P., Garbino, J., Pittet, D., 2003. Epidemiology of *Candida* species infections in critically ill non-immunosuppressed patients. *Lancet Infect. Dis.* 3, 685–702.
- Evans, M.V., Turton, H.E., Grant, C.M., Dawes, I.W., 1998. Toxicity of linoleic acid hydroperoxide to *Saccharomyces cerevisiae*: involvement of a respiration-related process for maximal sensitivity and adaptive response. *J. Bacteriol.* 180, 483–490.

- Flanagan, P.R., Fletcher, J., Boyle, H., Sulea, R., Moran, G.P., Sullivan, D.J., 2018. Expansion of the TLO gene family enhances the virulence of *Candida* species. *PLOS ONE* 13, e0200852. <https://doi.org/10.1371/journal.pone.0200852>
- Flowers, S.A., Barker, K.S., Berkow, E.L., Toner, G., Chadwick, S.G., Gygax, S.E., Morschhäuser, J., Rogers, P.D., 2012. Gain-of-function mutations in UPC2 are a frequent cause of ERG11 upregulation in azole-resistant clinical isolates of *Candida albicans*. *Eukaryot. Cell* 11, 1289–1299. <https://doi.org/10.1128/EC.00215-12>
- Forche, A., 2014. Large-Scale Chromosomal Changes and Associated Fitness Consequences in Pathogenic Fungi. *Curr. Fungal Infect. Rep.* 8, 163–170. <https://doi.org/10.1007/s12281-014-0181-2>
- Forche, A., Abbey, D., Pisithkul, T., Weinzierl, M.A., Ringstrom, T., Bruck, D., Petersen, K., Berman, J., 2011. Stress alters rates and types of loss of heterozygosity in *Candida albicans*. *mBio* 2. <https://doi.org/10.1128/mBio.00129-11>
- Forche, A., Alby, K., Schaefer, D., Johnson, A.D., Berman, J., Bennett, R.J., 2008. The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. *PLoS Biol.* 6, e110. <https://doi.org/10.1371/journal.pbio.0060110>
- Forche, A., Cromie, G., Gerstein, A.C., Solis, N.V., Pisithkul, T., Srifa, W., Jeffery, E., Abbey, D., Filler, S.G., Dudley, A.M., Berman, J., 2018. Rapid Phenotypic and Genotypic Diversification After Exposure to the Oral Host Niche in *Candida albicans*. *Genetics* 209, 725–741. <https://doi.org/10.1534/genetics.118.301019>
- Forche, A., Magee, P.T., Selmecki, A., Berman, J., May, G., 2009. Evolution in *Candida albicans* populations during a single passage through a mouse host. *Genetics* 182, 799–811. <https://doi.org/10.1534/genetics.109.103325>
- Forche, A., May, G., Magee, P.T., 2005. Demonstration of loss of heterozygosity by single-nucleotide polymorphism microarray analysis and alterations in strain morphology in *Candida albicans* strains during infection. *Eukaryot. Cell* 4, 156–165. <https://doi.org/10.1128/EC.4.1.156-165.2005>
- Ford, C.B., Funt, J.M., Abbey, D., Issi, L., Guiducci, C., Martinez, D.A., Delorey, T., Li, B.Y., White, T.C., Cuomo, C., Rao, R.P., Berman, J., Thompson, D.A., Regev, A., 2015. The evolution of drug resistance in clinical isolates of *Candida albicans*. *eLife* 4, e00662. <https://doi.org/10.7554/eLife.00662>
- Freire-Benítez, V., Price, R.J., Tarrant, D., Berman, J., Buscaino, A., 2016. *Candida albicans* repetitive elements display epigenetic diversity and plasticity. *Sci. Rep.* 6. <https://doi.org/10.1038/srep22989>
- Gabaldón, T., Naranjo-Ortíz, M.A., Marcet-Houben, M., 2016. Evolutionary genomics of yeast pathogens in the Saccharomycotina. *FEMS Yeast Res.* 16, fow064. <https://doi.org/10.1093/femsyr/fow064>
- Gerstein, A.C., Berman, J., 2015. Shift and adapt: the costs and benefits of karyotype variations. *Curr. Opin. Microbiol.* 26, 130–136. <https://doi.org/10.1016/j.mib.2015.06.010>
- Ghannoum, M.A., Jurevic, R.J., Mukherjee, P.K., Cui, F., Sikaroodi, M., Naqvi, A., Gillevet, P.M., 2010. Characterization of the oral fungal microbiome (mycobiome) in healthy individuals. *PLoS Pathog.* 6, e1000713. <https://doi.org/10.1371/journal.ppat.1000713>
- Gillum, A.M., Tsay, E.Y., Kirsch, D.R., 1984. Isolation of the *Candida albicans* gene for orotidine-5'-phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF* mutations. *Mol. Gen. Genet.* MGG 198, 179–182.
- Gomes, A.C., Miranda, I., Silva, R.M., Moura, G.R., Thomas, B., Akoulitchev, A., Santos, M.A., 2007. A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. *Genome Biol.* 8, R206. <https://doi.org/10.1186/gb-2007-8-10-r206>
- Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351. <https://doi.org/10.1038/nrg.2016.49>
- Grandcolas, P., 2009. Chapitre 4 : « Adaptation », in: *Les Mondes Darwiniens*. p. pg. 88.

- Hampe, I.A.I., Friedman, J., Edgerton, M., Morschhäuser, J., 2017. An acquired mechanism of antifungal drug resistance simultaneously enables *Candida albicans* to escape from intrinsic host defenses. *PLOS Pathog.* 13, e1006655. <https://doi.org/10.1371/journal.ppat.1006655>
- Hanaoka, N., Takano, Y., Shibuya, K., Fugo, H., Uehara, Y., Niimi, M., 2008. Identification of the putative protein phosphatase gene *PTC1* as a virulence-related gene using a silkworm model of *Candida albicans* infection. *Eukaryot. Cell* 7, 1640–1648. <https://doi.org/10.1128/EC.00129-08>
- Haran, J., Boyle, H., Hokamp, K., Yeomans, T., Liu, Z., Church, M., Fleming, A.B., Anderson, M.Z., Berman, J., Myers, L.C., Sullivan, D.J., Moran, G.P., 2014. Telomeric ORFs (TLOs) in *Candida* spp. Encode Mediator Subunits That Regulate Distinct Virulence Traits. *PLoS Genet.* 10, e1004658. <https://doi.org/10.1371/journal.pgen.1004658>
- Hellborg, L., Woolfit, M., Arthursson-Hellborg, M., Piškur, J., 2008. Complex evolution of the *DAL5* transporter family. *BMC Genomics* 9, 164. <https://doi.org/10.1186/1471-2164-9-164>
- Hickman, M.A., Paulson, C., Dudley, A., Berman, J., 2015. Parasexual Ploidy Reduction Drives Population Heterogeneity Through Random and Transient Aneuploidy in *Candida albicans*. *Genetics* 200, 781–794. <https://doi.org/10.1534/genetics.115.178020>
- Hickman, M.A., Zeng, G., Forche, A., Hirakawa, M.P., Abbey, D., Harrison, B.D., Wang, Y.-M., Su, C., Bennett, R.J., Wang, Y., Berman, J., 2013. The “obligate diploid” *Candida albicans* forms mating-competent haploids. *Nature* 494, 55–59. <https://doi.org/10.1038/nature11865>
- Hilton, C., Markie, D., Corner, B., Rikkerink, E., Poulter, R., 1985. Heat shock induces chromosome loss in the yeast *Candida albicans*. *Mol. Gen. Genet.* MGG 200, 162–168.
- Hirakawa, M.P., Chyou, D.E., Huang, D., Slan, A.R., Bennett, R.J., 2017. Parasex Generates Phenotypic Diversity de Novo and Impacts Drug Resistance and Virulence in *Candida albicans*. *Genetics* 207, 1195–1211. <https://doi.org/10.1534/genetics.117.300295>
- Hirakawa, M.P., Martinez, D.A., Sakthikumar, S., Anderson, M.Z., Berlin, A., Gujja, S., Zeng, Q., Zisson, E., Wang, J.M., Greenberg, J.M., Berman, J., Bennett, R.J., Cuomo, C.A., 2015. Genetic and phenotypic intra-species variation in *Candida albicans*. *Genome Res.* 25, 413–425. <https://doi.org/10.1101/gr.174623.114>
- Homann, O.R., Dea, J., Noble, S.M., Johnson, A.D., 2009. A phenotypic profile of the *Candida albicans* regulatory network. *PLoS Genet.* 5, e1000783. <https://doi.org/10.1371/journal.pgen.1000783>
- Hood-DeGrenier, J.K., 2011. Identification of phosphatase 2A-like Sit4-mediated signalling and ubiquitin-dependent protein sorting as modulators of caffeine sensitivity in *S. cerevisiae*. *Yeast Chichester Engl.* 28, 189–204. <https://doi.org/10.1002/yea.1830>
- Hube, B., 2004. From commensal to pathogen: stage- and tissue-specific gene expression of *Candida albicans*. *Curr. Opin. Microbiol.* 7, 336–341. <https://doi.org/10.1016/j.mib.2004.06.003>
- Hull, C.M., Raisner, R.M., Johnson, A.D., 2000. Evidence for mating of the “asexual” yeast *Candida albicans* in a mammalian host. *Science* 289, 307–310.
- Hunter, P.R., 1995. Discrimination of strains of *Candida albicans* isolated from deep and superficial sites by resistotyping. *Mycoses* 38, 37–40.
- Iwaguchi, S.-I., Sato, M., Magee, B.B., Magee, P.T., Makimura, K., Suzuki, T., 2001. Extensive chromosome translocation in a clinical isolate showing the distinctive carbohydrate assimilation profile from a candidiasis patient. *Yeast* 18, 1035–1046. <https://doi.org/10.1002/yea.748>
- Jackson, A.P., Gamble, J.A., Yeomans, T., Moran, G.P., Saunders, D., Harris, D., Aslett, M., Barrell, J.F., Butler, G., Citiulo, F., Coleman, D.C., de Groot, P.W.J., Goodwin, T.J., Quail, M.A., McQuillan, J., Munro, C.A., Pain, A., Poulter, R.T., Rajandream, M.-A., Renauld, H., Spiering, M.J., Tivey, A., Gow, N.A.R., Barrell, B., Sullivan, D.J., Berriman, M., 2009. Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Res.* 19, 2231–2244. <https://doi.org/10.1101/gr.097501.109>

- Jacobsen, M.D., Duncan, A.D., Bain, J., Johnson, E.M., Naglik, J.R., Shaw, D.J., Gow, N.A.R., Odds, F.C., 2008. Mixed *Candida albicans* strain populations in colonized and infected mucosal tissues. *FEMS Yeast Res.* 8, 1334–1338. <https://doi.org/10.1111/j.1567-1364.2008.00438.x>
- Janbon, G., Sherman, F., Rustchenko, E., 1998. Monosomy of a specific chromosome determines L-sorbose utilization: a novel regulatory mechanism in *Candida albicans*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5150–5155.
- Jones, T., Federspiel, N.A., Chibana, H., Dungan, J., Kalman, S., Magee, B.B., Newport, G., Thorstenson, Y.R., Agabian, N., Magee, P.T., Davis, R.W., Scherer, S., 2004. The diploid genome sequence of *Candida albicans*. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7329–7334. <https://doi.org/10.1073/pnas.0401648101>
- Kabir, M.A., Hussain, M.A., Ahmad, Z., 2012. *Candida albicans*: A Model Organism for Studying Fungal Pathogens. *ISRN Microbiol.* 2012, 1–15. <https://doi.org/10.5402/2012/538694>
- Kakade, P., Sadhale, P., Sanyal, K., Nagaraja, V., 2016. ZCF32, a fungus specific Zn(II)₂ Cys₆ transcription factor, is a repressor of the biofilm development in the human pathogen *Candida albicans*. *Sci. Rep.* 6, 31124. <https://doi.org/10.1038/srep31124>
- Kenny, E.F., Herzig, A., Krüger, R., Muth, A., Mondal, S., Thompson, P.R., Brinkmann, V., Bernuth, H. von, Zychlinsky, A., 2017. Diverse stimuli engage different neutrophil extracellular trap pathways. *eLife* 6. <https://doi.org/10.7554/eLife.24437>
- Khandelwal, N.K., Chauhan, N., Sarkar, P., Esquivel, B.D., Coccetti, P., Singh, A., Coste, A.T., Gupta, M., Sanglard, D., White, T.C., Chauvel, M., d’Enfert, C., Chattopadhyay, A., Gaur, N.A., Mondal, A.K., Prasad, R., 2018. Azole resistance in a *Candida albicans* mutant lacking the ABC transporter *CDR6/ROA1* depends on TOR signaling. *J. Biol. Chem.* 293, 412–432. <https://doi.org/10.1074/jbc.M117.807032>
- Kim, S.H., Clark, S.T., Surendra, A., Copeland, J.K., Wang, P.W., Ammar, R., Collins, C., Tullis, D.E., Nislow, C., Hwang, D.M., Guttman, D.S., Cowen, L.E., 2015. Global Analysis of the Fungal Microbiome in Cystic Fibrosis Patients Reveals Loss of Function of the Transcriptional Repressor Nrg1 as a Mechanism of Pathogen Adaptation. *PLoS Pathog.* 11, e1005308. <https://doi.org/10.1371/journal.ppat.1005308>
- Kirkpatrick, C.H., 2001. Chronic mucocutaneous candidiasis. *Pediatr. Infect. Dis. J.* 20, 197–206.
- Klotz, S.A., Gaur, N.K., De Armond, R., Sheppard, D., Khardori, N., Edwards, J.E., Lipke, P.N., El-Azizi, M., 2007. *Candida albicans* Als proteins mediate aggregation with bacteria and yeasts. *Med. Mycol.* 45, 363–370. <https://doi.org/10.1080/13693780701299333>
- Kock, J.L.F., Sebolai, O.M., Pohl, C.H., van Wyk, P.W.J., Lodolo, E.J., 2007. Oxylipin studies expose aspirin as antifungal. *FEMS Yeast Res.* 7, 1207–1217. <https://doi.org/10.1111/j.1567-1364.2007.00273.x>
- Kritikos, G., Banzhaf, M., Herrera-Dominguez, L., Koumoutsi, A., Wartel, M., Zietek, M., Typas, A., 2017. A tool named Iris for versatile high-throughput phenotyping in microorganisms. *Nat. Microbiol.* 2. <https://doi.org/10.1038/nmicrobiol.2017.14>
- Kullberg, B.J., Arendrup, M.C., 2015. Invasive Candidiasis. *N. Engl. J. Med.* 373, 1445–1456. <https://doi.org/10.1056/NEJMra1315399>
- Kumamoto, C.A., 2008. Niche-specific gene expression during *C. albicans* infection. *Curr. Opin. Microbiol.* 11, 325–330. <https://doi.org/10.1016/j.mib.2008.05.008>
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Kuranda, K., Leberre, V., Sokol, S., Palamarczyk, G., Francois, J., 2006. Investigating the caffeine effects in the yeast *Saccharomyces cerevisiae* brings new insights into the connection between TOR, PKC and Ras/cAMP signalling pathways. *Mol. Microbiol.* 61, 1147–1166. <https://doi.org/10.1111/j.1365-2958.2006.05300.x>
- Lackner, M., Tscherner, M., Schaller, M., Kuchler, K., Mair, C., Sartori, B., Istel, F., Arendrup, M.C., Lass-Flörl, C., 2014. Positions and Numbers of *FKS* Mutations in *Candida albicans* Selectively Influence *In Vitro* and *In Vivo*

- Susceptibilities to Echinocandin Treatment. *Antimicrob. Agents Chemother.* 58, 3626–3635. <https://doi.org/10.1128/AAC.00123-14>
- Lang, G.I., Murray, A.W., 2008. Estimating the Per-Base-Pair Mutation Rate in the Yeast *Saccharomyces cerevisiae*. *Genetics* 178, 67–82. <https://doi.org/10.1534/genetics.107.071506>
- Lephart, P.R., Chibana, H., Magee, P.T., 2005. Effect of the major repeat sequence on chromosome loss in *Candida albicans*. *Eukaryot. Cell* 4, 733–741. <https://doi.org/10.1128/EC.4.4.733-741.2005>
- Levitin, A., Whiteway, M., 2007. The effect of prostaglandin E2 on transcriptional responses of *Candida albicans*. *Microbiol. Res.* 162, 201–210. <https://doi.org/10.1016/j.micres.2007.02.001>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, J., Vinh, D.C., Casanova, J.-L., Puel, A., 2017. Inborn errors of immunity underlying fungal diseases in otherwise healthy individuals. *Curr. Opin. Microbiol.* 40, 46–57. <https://doi.org/10.1016/j.mib.2017.10.016>
- Limjindaporn, T., Khalaf, R.A., Fonzi, W.A., 2003. Nitrogen metabolism and virulence of *Candida albicans* require the GATA-type transcriptional activator encoded by GAT1. *Mol. Microbiol.* 50, 993–1004.
- Ling, Y., Cypowyj, S., Aytakin, C., Galicchio, M., Camcioglu, Y., Nepesov, S., Ikinciogullari, A., Dogu, F., Belkadi, A., Levy, R., Migaud, M., Boisson, B., Bolze, A., Itan, Y., Goudin, N., Cottineau, J., Picard, C., Abel, L., Bustamante, J., Casanova, J.-L., Puel, A., 2015. Inherited IL-17RC deficiency in patients with chronic mucocutaneous candidiasis. *J. Exp. Med.* 212, 619–631. <https://doi.org/10.1084/jem.20141065>
- Liu, L., Okada, S., Kong, X.-F., Kreins, A.Y., Cypowyj, S., Abhyankar, A., Toubiana, J., Itan, Y., Audry, M., Nitschke, P., Masson, C., Toth, B., Flatot, J., Migaud, M., Chrabieh, M., Kochetkov, T., Bolze, A., Borghesi, A., Toulon, A., Hiller, J., Eyerich, S., Eyerich, K., Gulácsy, V., Chernyshova, L., Chernyshov, V., Bondarenko, A., María Cortés Grimaldo, R., Blancas-Galicia, L., Madrigal Beas, I.M., Roesler, J., Magdorf, K., Engelhard, D., Thumerelle, C., Burgel, P.-R., Hoernes, M., Drexel, B., Seger, R., Kusuma, T., Jansson, A.F., Sawalle-Belohradsky, J., Belohradsky, B., Jouanguy, E., Bustamante, J., Bué, M., Karin, N., Wildbaum, G., Bodemer, C., Lortholary, O., Fischer, A., Blanche, S., Al-Muhsen, S., Reichenbach, J., Kobayashi, M., Rosales, F.E., Lozano, C.T., Kilic, S.S., Oleastro, M., Etzioni, A., Traidl-Hoffmann, C., Renner, E.D., Abel, L., Picard, C., Maródi, L., Boisson-Dupuis, S., Puel, A., Casanova, J.-L., 2011. Gain-of-function human *STAT1* mutations impair IL-17 immunity and underlie chronic mucocutaneous candidiasis. *J. Exp. Med.* 208, 1635–1648. <https://doi.org/10.1084/jem.20110958>
- Löffler, J., Kelly, S.L., Hebart, H., Schumacher, U., Lass-Flörl, C., Einsele, H., 1997. Molecular analysis of *cyp51* from fluconazole-resistant *Candida albicans* strains. *FEMS Microbiol. Lett.* 151, 263–268.
- Lott, T.J., Fundyga, R.E., Kuykendall, R.J., Arnold, J., 2005. The human commensal yeast, *Candida albicans*, has an ancient origin. *Fungal Genet. Biol.* 42, 444–451. <https://doi.org/10.1016/j.fgb.2005.01.012>
- Lu, Y., Su, C., Mao, X., Raniga, P.P., Liu, H., Chen, J., 2008. Efg1-mediated recruitment of NuA4 to promoters is required for hypha-specific Swi/Snf binding and activation in *Candida albicans*. *Mol. Biol. Cell* 19, 4260–4272. <https://doi.org/10.1091/mbc.e08-02-0173>
- MacCallum, D.M., Castillo, L., Nather, K., Munro, C.A., Brown, A.J.P., Gow, N.A.R., Odds, F.C., 2009. Property differences among the four major *Candida albicans* strain clades. *Eukaryot. Cell* 8, 373–387. <https://doi.org/10.1128/EC.00387-08>
- Magee, B.B., Magee, P.T., 2000. Induction of mating in *Candida albicans* by construction of MTL α and MTL β strains. *Science* 289, 310–313.
- Magwene, P.M., Kayıkçı, Ö., Granek, J.A., Reininga, J.M., Scholl, Z., Murray, D., 2011. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 108, 1987–1992. <https://doi.org/10.1073/pnas.1012544108>

- Massey, S.E., Moura, G., Beltrão, P., Almeida, R., Garey, J.R., Tuite, M.F., Santos, M.A.S., 2003. Comparative evolutionary genomics unveils the molecular mechanism of reassignment of the CTG codon in *Candida* spp. *Genome Res.* 13, 544–557. <https://doi.org/10.1101/gr.811003>
- McCullough, M.J., Clemons, K.V., Stevens, D.A., 1999. Molecular epidemiology of the global and temporal diversity of *Candida albicans*. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 29, 1220–1225. <https://doi.org/10.1086/313455>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McManus, B.A., Coleman, D.C., 2014. Molecular epidemiology, phylogeny and evolution of *Candida albicans*. *Infect. Genet. Evol.* 21, 166–178. <https://doi.org/10.1016/j.meegid.2013.11.008>
- McManus, B.A., McGovern, E., Moran, G.P., Healy, C.M., Nunn, J., Fleming, P., Costigan, C., Sullivan, D.J., Coleman, D.C., 2011. Microbiological Screening of Irish Patients with Autoimmune Polyendocrinopathy-Candidiasis-Ectodermal Dystrophy Reveals Persistence of *Candida albicans* Strains, Gradual Reduction in Susceptibility to Azoles, and Incidences of Clinical Signs of Oral Candidiasis without Culture Evidence. *J. Clin. Microbiol.* 49, 1879–1889. <https://doi.org/10.1128/JCM.00026-11>
- Metzker, M.L., 2010. Sequencing technologies — the next generation. *Nat. Rev. Genet.* 11, 31–46. <https://doi.org/10.1038/nrg2626>
- Miranda, I., Rocha, R., Santos, M.C., Mateus, D.D., Moura, G.R., Carreto, L., Santos, M.A.S., 2007. A genetic code alteration is a phenotype diversity generator in the human pathogen *Candida albicans*. *PLoS One* 2, e996. <https://doi.org/10.1371/journal.pone.0000996>
- Miranda, I., Silva-Dias, A., Rocha, R., Teixeira-Santos, R., Coelho, C., Gonçalves, T., Santos, M.A.S., Pina-Vaz, C., Solis, N.V., Filler, S.G., Rodrigues, A.G., 2013. *Candida albicans* CUG mistranslation is a mechanism to create cell surface variation. *mBio* 4. <https://doi.org/10.1128/mBio.00285-13>
- Moorhouse, A.J., Rennison, C., Raza, M., Lilic, D., Gow, N.A.R., 2016. Clonal Strain Persistence of *Candida albicans* Isolates from Chronic Mucocutaneous Candidiasis Patients. *PLOS ONE* 11, e0145888. <https://doi.org/10.1371/journal.pone.0145888>
- Mount, H.O., Revie, N.M., Todd, R.T., Anstett, K., Collins, C., Costanzo, M., Boone, C., Robbins, N., Selmecki, A., Cowen, L.E., 2018. Global analysis of genetic circuitry and adaptive mechanisms enabling resistance to the azole antifungal drugs. *PLoS Genet.* 14, e1007319. <https://doi.org/10.1371/journal.pgen.1007319>
- Moyes, D.L., Runglall, M., Murciano, C., Shen, C., Nayar, D., Thavaraj, S., Kohli, A., Islam, A., Mora-Montes, H., Challacombe, S.J., Naglik, J.R., 2010. A Biphasic Innate Immune MAPK Response Discriminates between the Yeast and Hyphal Forms of *Candida albicans* in Epithelial Cells. *Cell Host Microbe* 8, 225–235. <https://doi.org/10.1016/j.chom.2010.08.002>
- Moyes, D.L., Wilson, D., Richardson, J.P., Mogavero, S., Tang, S.X., Wernecke, J., Höfs, S., Gratacap, R.L., Robbins, J., Runglall, M., Murciano, C., Blagojevic, M., Thavaraj, S., Förster, T.M., Hebecker, B., Kasper, L., Vizcay, G., Iancu, S.I., Kichik, N., Häder, A., Kurzai, O., Luo, T., Krüger, T., Kniemeyer, O., Cota, E., Bader, O., Wheeler, R.T., Gutschmann, T., Hube, B., Naglik, J.R., 2016. Candidalysin is a fungal peptide toxin critical for mucosal infection. *Nature* 532, 64–68. <https://doi.org/10.1038/nature17625>
- Muzzey, D., Schwartz, K., Weissman, J.S., Sherlock, G., 2013. Assembly of a phased diploid *Candida albicans* genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. *Genome Biol.* 14, R97. <https://doi.org/10.1186/gb-2013-14-9-r97>
- Muzzey, D., Sherlock, G., Weissman, J.S., 2014. Extensive and coordinated control of allele-specific expression by both transcription and translation in *Candida albicans*. *Genome Res.* 24, 963–973. <https://doi.org/10.1101/gr.166322.113>

- Naglik, J.R., König, A., Hube, B., Gaffen, S.L., 2017. *Candida albicans*-epithelial interactions and induction of mucosal innate immunity. *Curr. Opin. Microbiol.* 40, 104–112. <https://doi.org/10.1016/j.mib.2017.10.030>
- Naglik, J.R., Richardson, J.P., Moyes, D.L., 2014. *Candida albicans* pathogenicity and epithelial immunity. *PLoS Pathog.* 10, e1004257. <https://doi.org/10.1371/journal.ppat.1004257>
- Noble, S.M., Johnson, A.D., 2007. Genetics of *Candida albicans*, a diploid human fungal pathogen. *Annu. Rev. Genet.* 41, 193–211. <https://doi.org/10.1146/annurev.genet.41.042007.170146>
- O'Connor, M., Peifer, M., Bender, W., 1989. Construction of large DNA segments in *Escherichia coli*. *Science* 244, 1307–1312. <https://doi.org/10.1126/science.2660262>
- Odds, F.C., 2010. Molecular phylogenetics and epidemiology of *Candida albicans*. *Future Microbiol.* 5, 67–79. <https://doi.org/10.2217/fmb.09.113>
- Odds, F.C., Bougnoux, M.-E., Shaw, D.J., Bain, J.M., Davidson, A.D., Diogo, D., Jacobsen, M.D., Lecomte, M., Li, S.-Y., Tavanti, A., Maiden, M.C.J., Gow, N.A.R., d'Enfert, C., 2007. Molecular phylogenetics of *Candida albicans*. *Eukaryot. Cell* 6, 1041–1052. <https://doi.org/10.1128/EC.00041-07>
- Odds, F.C., Davidson, A.D., Jacobsen, M.D., Tavanti, A., Whyte, J.A., Kibbler, C.C., Ellis, D.H., Maiden, M.C.J., Shaw, D.J., Gow, N. a. R., 2006. *Candida albicans* strain maintenance, replacement, and microvariation demonstrated by multilocus sequence typing. *J. Clin. Microbiol.* 44, 3647–3658. <https://doi.org/10.1128/JCM.00934-06>
- Odds, F.C., Jacobsen, M.D., 2008. Multilocus sequence typing of pathogenic *Candida* species. *Eukaryot. Cell* 7, 1075–1084. <https://doi.org/10.1128/EC.00062-08>
- Okada, S., Puel, A., Casanova, J.-L., Kobayashi, M., 2016. Chronic mucocutaneous candidiasis disease associated with inborn errors of IL-17 immunity. *Clin. Transl. Immunol.* 5, e114. <https://doi.org/10.1038/cti.2016.71>
- Pande, K., Chen, C., Noble, S.M., 2013. Passage through the mammalian gut triggers a phenotypic switch that promotes *Candida albicans* commensalism. *Nat. Genet.* 45, 1088–1091. <https://doi.org/10.1038/ng.2710>
- Pappas, P.G., Kauffman, C.A., Andes, D.R., Clancy, C.J., Marr, K.A., Ostrosky-Zeichner, L., Reboli, A.C., Schuster, M.G., Vazquez, J.A., Walsh, T.J., Zaoutis, T.E., Sobel, J.D., 2016. Clinical Practice Guideline for the Management of Candidiasis: 2016 Update by the Infectious Diseases Society of America. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 62, e1-50. <https://doi.org/10.1093/cid/civ933>
- Pavelka, N., Rancati, G., Zhu, J., Bradford, W.D., Saraf, A., Florens, L., Sanderson, B.W., Hattem, G.L., Li, R., 2010. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468, 321–325. <https://doi.org/10.1038/nature09529>
- Pfaller, M.A., Diekema, D.J., 2012. Progress in antifungal susceptibility testing of *Candida* spp. by use of Clinical and Laboratory Standards Institute broth microdilution methods, 2010 to 2012. *J. Clin. Microbiol.* 50, 2846–2856. <https://doi.org/10.1128/JCM.00937-12>
- Pfaller, M.A., Diekema, D.J., 2007. Epidemiology of Invasive Candidiasis: a Persistent Public Health Problem. *Clin. Microbiol. Rev.* 20, 133–163. <https://doi.org/10.1128/CMR.00029-06>
- Pfaller, M.A., Diekema, D.J., 2004. Rare and emerging opportunistic fungal pathogens: concern for resistance beyond *Candida albicans* and *Aspergillus fumigatus*. *J. Clin. Microbiol.* 42, 4419–4431. <https://doi.org/10.1128/JCM.42.10.4419-4431.2004>
- Pfaller, M.A., Diekema, D.J., Gibbs, D.L., Newell, V.A., Ellis, D., Tullio, V., Rodloff, A., Fu, W., Ling, T.A., Global Antifungal Surveillance Group, 2010. Results from the ARTEMIS DISK Global Antifungal Surveillance Study, 1997 to 2007: a 10.5-year analysis of susceptibilities of *Candida* Species to fluconazole and voriconazole as determined by CLSI standardized disk diffusion. *J. Clin. Microbiol.* 48, 1366–1377. <https://doi.org/10.1128/JCM.02117-09>

- Pfaller, M.A., Messer, S.A., Woosley, L.N., Jones, R.N., Castanheira, M., 2013. Echinocandin and triazole antifungal susceptibility profiles for clinical opportunistic yeast and mold isolates collected from 2010 to 2011: application of new CLSI clinical breakpoints and epidemiological cutoff values for characterization of geographic and temporal trends of antifungal resistance. *J. Clin. Microbiol.* 51, 2571–2581. <https://doi.org/10.1128/JCM.00308-13>
- Pfaller, M.A., Rhomberg, P.R., Messer, S.A., Jones, R.N., Castanheira, M., 2015. Isavuconazole, micafungin, and 8 comparator antifungal agents' susceptibility profiles for common and uncommon opportunistic fungi collected in 2013: temporal analysis of antifungal drug resistance using CLSI species-specific clinical breakpoints and proposed epidemiological cutoff values. *Diagn. Microbiol. Infect. Dis.* 82, 303–313. <https://doi.org/10.1016/j.diagmicrobio.2015.04.008>
- Puel, A., Cypowyj, S., Bustamante, J., Wright, J.F., Liu, L., Lim, H.K., Migaud, M., Israel, L., Chrabieh, M., Audry, M., Gumbleton, M., Toulon, A., Bodemer, C., El-Baghdadi, J., Whitters, M., Paradis, T., Brooks, J., Collins, M., Wolfman, N.M., Al-Muhsen, S., Galicchio, M., Abel, L., Picard, C., Casanova, J.-L., 2011. Chronic mucocutaneous candidiasis in humans with inborn errors of interleukin-17 immunity. *Science* 332, 65–68. <https://doi.org/10.1126/science.1200439>
- Puel, A., Cypowyj, S., Maródi, L., Abel, L., Picard, C., Casanova, J.-L., 2012. Inborn errors of human IL-17 immunity underlie chronic mucocutaneous candidiasis: *Curr. Opin. Allergy Clin. Immunol.* 12, 616–622. <https://doi.org/10.1097/ACI.0b013e328358cc0b>
- Pujol, C., Joly, S., Lockhart, S.R., Noel, S., Tibayrenc, M., Soll, D.R., 1997. Parity among the randomly amplified polymorphic DNA method, multilocus enzyme electrophoresis, and Southern blot hybridization with the moderately repetitive DNA probe Ca3 for fingerprinting *Candida albicans*. *J. Clin. Microbiol.* 35, 2348–2358.
- Pujol, C., Pfaller, M.A., Soll, D.R., 2004. Flucytosine resistance is restricted to a single genetic clade of *Candida albicans*. *Antimicrob. Agents Chemother.* 48, 262–266.
- Qi, Q.G., Hu, T., Zhou, X.D., 2005. Frequency, species and molecular characterization of oral *Candida* in hosts of different age in China. *J. Oral Pathol. Med. Off. Publ. Int. Assoc. Oral Pathol. Am. Acad. Oral Pathol.* 34, 352–356. <https://doi.org/10.1111/j.1600-0714.2005.00326.x>
- Richardson, J.P., Moyes, D.L., 2015. Adaptive immune responses to *Candida albicans* infection. *Virulence* 6, 327–337. <https://doi.org/10.1080/21505594.2015.1004977>
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., Müller, M., 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12, 77. <https://doi.org/10.1186/1471-2105-12-77>
- Rodriguez, C., Soulié, C., Marcelin, A.-G., Calvez, V., Descamps, D., Charpentier, C., Flandre, P., Recordon-Pinson, P., Bellecave, P., Pawlotsky, J.-M., Masquelier, B., the ANRS AC11 Study Group, 2015. HIV-1 Coreceptor Usage Assessment by Ultra-Deep Pyrosequencing and Response to Maraviroc. *PLOS ONE* 10, e0127816. <https://doi.org/10.1371/journal.pone.0127816>
- Ropars, J., Maufrays, C., Diogo, D., Marcet-Houben, M., Perin, A., Sertour, N., Mosca, K., Permal, E., Laval, G., Bouchier, C., Ma, L., Schwartz, K., Voelz, K., May, R.C., Poulain, J., Battail, C., Wincker, P., Borman, A.M., Chowdhary, A., Fan, S., Kim, S.H., Le Pape, P., Romeo, O., Shin, J.H., Gabaldon, T., Sherlock, G., Bournoux, M.-E., d'Enfert, C., 2018. Gene flow contributes to diversification of the major fungal pathogen *Candida albicans*. *Nat. Commun.* 9, 2253. <https://doi.org/10.1038/s41467-018-04787-4>
- Rustchenko, E.P., Howard, D.H., Sherman, F., 1994. Chromosomal alterations of *Candida albicans* are associated with the gain and loss of assimilating functions. *J. Bacteriol.* 176, 3231–3241.
- Sachidanandam, R., Weissman, D., Schmidt, S.C., Kakol, J.M., Stein, L.D., Marth, G., Sherry, S., Mullikin, J.C., Mortimore, B.J., Willey, D.L., Hunt, S.E., Cole, C.G., Coggill, P.C., Rice, C.M., Ning, Z., Rogers, J., Bentley, D.R., Kwok, P.-Y., Mardis, E.R., Yeh, R.T., Schultz, B., Cook, L., Davenport, R., Dante, M., Fulton, L., Hillier, L., Waterston, R.H., McPherson, J.D., Gilman, B., Schaffner, S., Van Etten, W.J., Reich, D., Higgins, J., Daly, M.J.,

- Blumenstiel, B., Baldwin, J., Stange-Thomann, N., Zody, M.C., Linton, L., Lander, E.S., Altshuler, D., 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933. <https://doi.org/10.1038/35057149>
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- Sanglard, D., Bille, J., 2002. Action of and resistance to antifungal agents. In: *Candida and candidiasis*. Calderone RA, p. 370.
- Sanglard, D., Ischer, F., Koymans, L., Bille, J., 1998. Amino acid substitutions in the cytochrome P-450 lanosterol 14 α -demethylase (CYP51A1) from azole-resistant *Candida albicans* clinical isolates contribute to resistance to azole antifungal agents. *Antimicrob. Agents Chemother.* 42, 241–253.
- Sanglard, D., Ischer, F., Marchetti, O., Entenza, J., Bille, J., 2003. Calcineurin A of *Candida albicans*: involvement in antifungal tolerance, cell morphogenesis and virulence. *Mol. Microbiol.* 48, 959–976.
- Scherer, S., Magee, P.T., 1990. Genetics of *Candida albicans*. *Microbiol. Rev.* 54, 226–241.
- Schönherr, F.A., Sparber, F., Kirchner, F.R., Guiducci, E., Trautwein-Weidner, K., Gladiator, A., Sertour, N., Hetzel, U., Le, G.T.T., Pavelka, N., d'Enfert, C., Bougnoux, M.-E., Corti, C.F., LeibundGut-Landmann, S., 2017. The intraspecies diversity of *C. albicans* triggers qualitatively and temporally distinct host responses that determine the balance between commensalism and pathogenicity. *Mucosal Immunol.* 10, 1335–1350. <https://doi.org/10.1038/mi.2017.2>
- Selmecki, A., Forche, A., Berman, J., 2010. Genomic plasticity of the human fungal pathogen *Candida albicans*. *Eukaryot. Cell* 9, 991–1008. <https://doi.org/10.1128/EC.00060-10>
- Selmecki, A., Forche, A., Berman, J., 2006. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science* 313, 367–370. <https://doi.org/10.1126/science.1128242>
- Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A., Berman, J., 2008. An isochromosome confers drug resistance in vivo by amplification of two genes, *ERG11* and *TAC1*. *Mol. Microbiol.* 68, 624–641. <https://doi.org/10.1111/j.1365-2958.2008.06176.x>
- Selmecki, A.M., Dulmage, K., Cowen, L.E., Anderson, J.B., Berman, J., 2009. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet.* 5, e1000705. <https://doi.org/10.1371/journal.pgen.1000705>
- Shekhar-Guturja, T., Gunaherath, G.M.K.B., Wijeratne, E.M.K., Lambert, J.-P., Averette, A.F., Lee, S.C., Kim, T., Bahn, Y.-S., Tripodi, F., Ammar, R., Döhl, K., Niewola-Staszewska, K., Schmitt, L., Loewith, R.J., Roth, F.P., Sanglard, D., Andes, D., Nislow, C., Coccetti, P., Gingras, A.-C., Heitman, J., Gunatilaka, A.A.L., Cowen, L.E., 2016. Dual action antifungal small molecule modulates multidrug efflux and TOR signaling. *Nat. Chem. Biol.* 12, 867–875. <https://doi.org/10.1038/nchembio.2165>
- Shin, J.H., Bougnoux, M.-E., d'Enfert, C., Kim, S.H., Moon, C.-J., Joo, M.Y., Lee, K., Kim, M.-N., Lee, H.S., Shin, M.G., Suh, S.P., Ryang, D.W., 2011. Genetic diversity among Korean *Candida albicans* bloodstream isolates: assessment by multilocus sequence typing and restriction endonuclease analysis of genomic DNA by use of *Bss*HII. *J. Clin. Microbiol.* 49, 2572–2577. <https://doi.org/10.1128/JCM.02153-10>
- Siegel, J.J., Amon, A., 2012. New Insights into the Troubles of Aneuploidy. *Annu. Rev. Cell Dev. Biol.* 28, 189–214. <https://doi.org/10.1146/annurev-cellbio-101011-155807>
- Siikala, E., Rautemaa, R., Richardson, M., Saxen, H., Bowyer, P., Sanglard, D., 2010. Persistent *Candida albicans* colonization and molecular mechanisms of azole resistance in autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) patients. *J. Antimicrob. Chemother.* 65, 2505–2513. <https://doi.org/10.1093/jac/dkq354>

- Singh, R.P., Prasad, H.K., Sinha, I., Agarwal, N., Natarajan, K., 2011. Cap2-HAP complex is a critical transcriptional regulator that has dual but contrasting roles in regulation of iron homeostasis in *Candida albicans*. *J. Biol. Chem.* 286, 25154–25170. <https://doi.org/10.1074/jbc.M111.233569>
- Skrzypek, M.S., Binkley, J., Binkley, G., Miyasato, S.R., Simison, M., Sherlock, G., 2017. The *Candida* Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res.* 45, D592–D596. <https://doi.org/10.1093/nar/gkw924>
- Sudbery, P.E., 2011. Growth of *Candida albicans* hyphae. *Nat. Rev. Microbiol.* 9, 737–748. <https://doi.org/10.1038/nrmicro2636>
- Suh, S.-O., Blackwell, M., Kurtzman, C.P., Lachance, M.-A., 2006. Phylogenetics of Saccharomycetales, the ascomycete yeasts. *Mycologia* 98, 1006–1017.
- Sullivan, D.J., Berman, J., Myers, L.C., Moran, G.P., 2015. Telomeric ORFs in *Candida albicans*: does mediator tail wag the yeast? *PLoS Pathog.* 11, e1004614. <https://doi.org/10.1371/journal.ppat.1004614>
- Swynghedauw, B., 2012. Introduction: S’adapter, s’adapter vite, par tous les moyens, in: L’Homme Peut-Il s’adapter à Lui-Même ? p. pg. 18-20.
- Tait, E., Simon, M.C., King, S., Brown, A.J., Gow, N.A., Shaw, D.J., 1997. A *Candida albicans* genome project: cosmid contigs, physical mapping, and gene isolation. *Fungal Genet. Biol.* FG B 21, 308–314. <https://doi.org/10.1006/fgbi.1997.0983>
- Takagi, Y., Fukano, H., Shimozato, K., Tanaka, R., Horii, T., Kawamoto, F., Kanbe, T., 2013. Genotypes of *Candida albicans* isolated from healthy individuals and their distribution in patients with oral candidiasis. *J. Infect. Chemother. Off. J. Jpn. Soc. Chemother.* 19, 1072–1079. <https://doi.org/10.1007/s10156-013-0626-5>
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101, 11030–11035. <https://doi.org/10.1073/pnas.0404206101>
- Tavanti, A., Davidson, A.D., Fordyce, M.J., Gow, N.A.R., Maiden, M.C.J., Odds, F.C., 2005. Population structure and properties of *Candida albicans*, as determined by multilocus sequence typing. *J. Clin. Microbiol.* 43, 5601–5613. <https://doi.org/10.1128/JCM.43.11.5601-5613.2005>
- Tibshirani, R., 1996. « Regression shrinkage and selection via the lasso » 267–288.
- Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., Amon, A., 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* 317, 916–924. <https://doi.org/10.1126/science.1142210>
- Toubiana, J., Okada, S., Hiller, J., Oleastro, M., Lagos Gomez, M., Aldave Becerra, J.C., Ouachée-Chardin, M., Fouyssac, F., Girisha, K.M., Etzioni, A., Van Montfrans, J., Camcioglu, Y., Kerns, L.A., Belohradsky, B., Blanche, S., Bousfiha, A., Rodriguez-Gallego, C., Meyts, I., Kisand, K., Reichenbach, J., Renner, E.D., Rosenzweig, S., Grimbacher, B., van de Veerdonk, F.L., Traidl-Hoffmann, C., Picard, C., Marodi, L., Morio, T., Kobayashi, M., Lilic, D., Milner, J.D., Holland, S., Casanova, J.-L., Puel, A., International STAT1 Gain-of-Function Study Group, 2016. Heterozygous STAT1 gain-of-function mutations underlie an unexpectedly broad clinical phenotype. *Blood* 127, 3154–3164. <https://doi.org/10.1182/blood-2015-11-679902>
- Trautwein-Weidner, K., Gladiator, A., Nur, S., Diethelm, P., LeibundGut-Landmann, S., 2015. IL-17-mediated antifungal defense in the oral mucosa is independent of neutrophils. *Mucosal Immunol.* 8, 221–231. <https://doi.org/10.1038/mi.2014.57>
- van de Veerdonk, F.L., Plantinga, T.S., Hoischen, A., Smeekens, S.P., Joosten, L.A.B., Gilissen, C., Arts, P., Rosentul, D.C., Carmichael, A.J., Smits-van der Graaf, C.A.A., Kullberg, B.J., van der Meer, J.W.M., Lilic, D., Veltman, J.A., Netea, M.G., 2011. STAT1 mutations in autosomal dominant chronic mucocutaneous candidiasis. *N. Engl. J. Med.* 365, 54–61. <https://doi.org/10.1056/NEJMoa1100102>
- van het Hoog, M., Rast, T.J., Martchenko, M., Grindle, S., Dignard, D., Hogues, H., Cuomo, C., Berriman, M., Scherer, S., Magee, B., Whiteway, M., Chibana, H., Nantel, A., Magee, P., 2007. Assembly of the *Candida*

- albicans genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol.* 8, R52. <https://doi.org/10.1186/gb-2007-8-4-r52>
- Vandeputte, P., Ischer, F., Sanglard, D., Coste, A.T., 2011. In Vivo Systematic Analysis of *Candida albicans* Zn2-Cys6 Transcription Factors Mutants for Mice Organ Colonization. *PLoS ONE* 6, e26962. <https://doi.org/10.1371/journal.pone.0026962>
- Viaud, M.C., Balhadère, P.V., Talbot, N.J., 2002. A Magnaporthe grisea cyclophilin acts as a virulence determinant during plant infection. *Plant Cell* 14, 917–930.
- Wagih, O., Usaj, M., Baryshnikova, A., VanderSluis, B., Kuzmin, E., Costanzo, M., Myers, C.L., Andrews, B.J., Boone, C.M., Parts, L., 2013. SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. *Nucleic Acids Res.* 41, W591–W596. <https://doi.org/10.1093/nar/gkt400>
- Wang, H., Kong, F., Sorrell, T.C., Wang, B., McNicholas, P., Pantarat, N., Ellis, D., Xiao, M., Widmer, F., Chen, S.C., 2009. Rapid detection of ERG11 gene mutations in clinical *Candida albicans* isolates with reduced susceptibility to fluconazole by rolling circle amplification and DNA sequencing. *BMC Microbiol.* 9, 167. <https://doi.org/10.1186/1471-2180-9-167>
- Wartenberg, A., Linde, J., Martin, R., Schreiner, M., Horn, F., Jacobsen, I.D., Jenull, S., Wolf, T., Kuchler, K., Guthke, R., Kurzai, O., Forche, A., d'Enfert, C., Brunke, S., Hube, B., 2014. Microevolution of *Candida albicans* in macrophages restores filamentation in a nonfilamentous mutant. *PLoS Genet.* 10, e1004824. <https://doi.org/10.1371/journal.pgen.1004824>
- Wells, R.S., Higgs, J.M., Macdonald, A., Valdimarsson, H., Holt, P.J., 1972. Familial chronic muco-cutaneous candidiasis. *J. Med. Genet.* 9, 302–310.
- Wertheimer, N.B., Stone, N., Berman, J., 2016. Ploidy dynamics and evolvability in fungi. *Philos. Trans. R. Soc. B Biol. Sci.* 371, 20150461. <https://doi.org/10.1098/rstb.2015.0461>
- White, T.C., 1997. The presence of an R467K amino acid substitution and loss of allelic variation correlate with an azole-resistant lanosterol 14 α demethylase in *Candida albicans*. *Antimicrob. Agents Chemother.* 41, 1488–1494.
- Wu, Y., Gao, N., Li, C., Gao, J., Ying, C., 2017. A newly identified amino acid substitution T123I in the 14 α -demethylase (Erg11p) of *Candida albicans* confers azole resistance. *FEMS Yeast Res.* 17. <https://doi.org/10.1093/femsyr/fox012>
- Wysokar, A., Tibbetts, K., McCown, M., Homer, N., Fennell, T., n.d. Picard: A set of tools for working with next generation sequencing data in BAM format.
- Xu, J., Boyd, C.M., Livingston, E., Meyer, W., Madden, J.F., Mitchell, T.G., 1999. Species and genotypic diversities and similarities of pathogenic yeasts colonizing women. *J. Clin. Microbiol.* 37, 3835–3843.
- Zhang, L., Yan, L., Jiang, J., Wang, Y., Jiang, Y., Yan, T., Cao, Y., 2014. The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast *Candida albicans*. *Virulence* 5, 655–664. <https://doi.org/10.4161/viru.32180>
- Zhang, N., 2003. Sixty Alleles of the ALS7 Open Reading Frame in *Candida albicans*: ALS7 Is a Hypermutable Contingency Locus. *Genome Res.* 13, 2005–2017. <https://doi.org/10.1101/gr.1024903>
- Zhang, N., Magee, B.B., Magee, P.T., Holland, B.R., Rodrigues, E., Holmes, A.R., Cannon, R.D., Schmid, J., 2015. Selective Advantages of a Parasexual Cycle for the Yeast *Candida albicans*. *Genetics* 200, 1117–1132. <https://doi.org/10.1534/genetics.115.177170>

ANNEXES

1 **Within-Host Genomic Diversity of *Candida albicans* in Healthy Carriers:**

2 **Implication for Microevolution Studies**

3
4 Emilie Sitterlé^{1,2,3}, Corinne Maufrais⁴, Natacha Sertour¹, Matthieu Palayret⁵, Christophe
5 d'Enfert¹, Marie-Elisabeth Bougnoux^{1,3*}

6
7 ¹Fungal Biology and Pathogenicity Unit, Department of Mycology, Institut Pasteur, INRA,
8 Paris, France

9 ²Université Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, France

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

12 ⁴Center for Bioinformatics, BioStatistics and Integrative Biology (C3BI), USR 3756 IP
13 CNRS, Institut Pasteur, Paris, France

14 ⁵ Department of Biology, Ecole Polytechnique, Palaiseau, France.

15
16
17
18
19
20 *Corresponding author:

21 Marie-Elisabeth Bougnoux, MD, PhD,

22 Institut Pasteur, Unité Biologie et Pathogénicité Fongiques, Département Mycologie, 25 rue
23 du Docteur Roux, F-75015 Paris, France; Phone: +33 (1) 40 61 31 26; E-mail:
24 bougnoux@pasteur.fr

25

26 **Abstract**

27 Genomic variations in *Candida albicans*, a major fungal pathogen of humans, have been
28 observed upon exposure of this yeast to different stresses and experimental infections,
29 possibly contributing to subsequent adaptation to these stress conditions. Yet, little is known
30 about the extent of genomic diversity that is associated with commensalism, the predominant
31 lifestyle of *C. albicans* in humans. In this study, we investigated the genetic diversity of *C.*
32 *albicans* oral isolates recovered from healthy individuals, using multilocus sequencing typing
33 (MLST) and whole genome sequencing. While MLST revealed occasional differences
34 between isolates collected from a single individual, genomic sequencing showed that they
35 differed by numerous single nucleotide polymorphisms, mostly resulting from short-range
36 loss-of-heterozygosity events. These differences were shown to have occurred upon human
37 carriage of *C. albicans* rather than subsequent *in vitro* manipulation of the isolates. Thus, *C.*
38 *albicans* intra-sample diversity appears common in healthy individuals, higher than that
39 observed using MLST. We propose that diversifying lineages coexist in a single human
40 individual, and this diversity can enable rapid adaptation under stress exposure. These results
41 are crucial for the interpretation of longitudinal studies evaluating the evolution of the *C.*
42 *albicans* genome.

43

44 **Introduction**

45 All living organisms must adapt in order to thrive within their natural niches. They must also
46 be able to rapidly adopt strategies to adapt and survive under stressful conditions. Many
47 fungal pathogens have evolved a highly plastic genome, thereby enabling the generation of
48 genomic diversity¹. Because *C. albicans* is one of the leading fungal pathogens of humans,
49 genomic analysis and the mechanisms that allow this yeast to persist in humans have attracted
50 interest. *C. albicans* is the most common fungal commensal of humans and is considered a
51 facultative component of the normal human digestive microbiota²⁻⁴. This yeast is also a major
52 opportunistic pathogen responsible for both superficial and disseminated infections in
53 immunocompromised patients^{5,6}. In these patients, infections frequently originate from an
54 endogenous source, mainly the digestive tract, which represents the major reservoir of this
55 yeast^{7,8}. Therefore, the human digestive tract is probably the most relevant niche to
56 investigate the genomic diversification in this species.

57 *C. albicans* is predominantly diploid and demonstrates a significant degree of genetic
58 diversity across isolates, notably variations in the distribution of heterozygous polymorphisms
59 along the genome⁹⁻¹³. At the population level, molecular typing has revealed that *C. albicans*
60 strains belong to five major and thirteen minor genetic clusters^{14,15}. Some of these clusters
61 exhibit geographical enrichment or phenotypic specificities¹⁴⁻²⁴. Yet, no correlation between
62 cluster assignment and the ability of strains to cause different forms of infection has been
63 established¹⁴.

64 In addition to heterozygosity, the genome of *C. albicans* displays a high level of plasticity²⁵.
65 While *C. albicans* is an asexual organism with a predominantly clonal mode of
66 reproduction²⁶, it can also employ a parasexual cycle^{10,26-29}. This cycle allows *C. albicans* to
67 alternate between diploid and tetraploid states independently of meiosis, and is frequently
68 accompanied by the generation of aneuploidies and mitotic recombination (and consequently

69 loss-of-heterozygosity [LOH] events) between chromosome homologs^{26,27,29}. LOH events are
70 also observed during clonal propagation of diploid isolates, and the mechanisms that underlie
71 these events have been well-studied^{25,30}. Interestingly, it has been shown that environmental
72 modifications such as oxidative stress, high temperature, ultraviolet light, or exposure to
73 antifungal agents increase the rate of LOH events or induce ploidy variations in the *C.*
74 *albicans* genome^{26,31,32}. Genomic rearrangements have also been detected *in vivo*, for instance
75 during a passage in mice, which is not the natural host of *C. albicans*^{33–35}. In humans, a study
76 conducted by Ford et al. (2015) analysing the whole genome of 43 *C. albicans* isolates has
77 revealed that these events were also common during the course of an oral infection³⁶. Under
78 these different stressful conditions, both aneuploidy and LOH events can arise quickly,
79 enabling *C. albicans* to survive and possibly adapt to changing environments¹. However, little
80 is known about the occurrence and diversity of such genetic variations within the natural host
81 when *C. albicans* is a commensal. Studies using molecular typing have shown that *C.*
82 *albicans* strains can persist in healthy individuals for many years^{37,38} and evolve through
83 minor genetic variations^{37,39–45}. To broaden our understanding on the genomic diversity within
84 the healthy host, we addressed the question of the genome-wide genetic heterogeneity
85 between several *C. albicans* isolates obtained from a healthy individual. Our study indicated
86 that genetically distinct and yet closely related isolates co-exist in a healthy individual. This
87 intrinsic within-host genomic diversity should be taken into consideration when evaluating
88 the genomic evolution of *C. albicans* in longitudinal studies, as these studies often
89 characterise a single isolate at any time point.

90

91 **Results**

92 **Oral *C. albicans* carriage in healthy individuals**

93 We screened 56 undergraduate students to evaluate the prevalence of oral *Candida* carriage in
94 healthy individuals. Ten of the 56 students (17.9%) were carriers of *Candida* spp., 8
95 harboured only *C. albicans*, and 2 harboured both *C. albicans* and *C. glabrata*. The
96 prevalence of oral *C. albicans* carriage in this healthy population was 17.9%.

97 Up to 8 *C. albicans* isolates (colonies) from each carrier were analysed using MLST. As
98 shown in **Supplementary Table S1**, 10 diploid sequence types (DSTs) belonging to 5 genetic
99 clusters were identified among the 49 isolates analysed from the 10 carriers. Among the 7
100 students for which several isolates were typed, only one (student G) had isolates that
101 displayed different DSTs. Indeed, isolate G1 differed from isolates G2–G8 at one locus
102 (*ZWF1*). Sequence comparisons of the *ZWF1* locus of these isolates showed that the
103 variations between G1 and G2–G8 resulted from a LOH event involving 5 heterozygous
104 positions (**Supplementary Table S1**). Overall, our MLST analysis, while suggesting limited
105 within-host genetic diversity, indicated that *C. albicans* isolates collected from a single carrier
106 in the oral niche could show some genetic diversity, consistent with previous studies^{37,39–44}.

107

108 **Oral samples harbour genetically diverse *C. albicans* isolates**

109 Based on the above observation, we explored the extent of the genetic variations across
110 isolates collected from a single carrier. Isolates A1, A2, and A3 from carrier A (all having
111 DST 66); D1, D2, and D3 from carrier D (all having DST 1765); and G1 (DST 1768), G2,
112 and G3 (DST 1769) from carrier G were deep-sequenced using Illumina technology (average
113 sequencing depth, 155×, range, 110–229×; **Fig. 1A** and **Supplementary Table S2**). None of
114 the sequenced isolates showed aneuploid chromosomes (data not shown).

115 Single nucleotide polymorphisms (SNPs) were identified by mapping the sequencing reads to
116 haplotype A of the *C. albicans* SC5314 reference genome^{46,47}, and then inter-isolate
117 comparisons were conducted (**Table 1**). Isolates from carrier A showed an average of 3830
118 differences. Isolates from carrier D showed an average of 757 differences. Isolates from
119 carrier G showed an average of 1404 differences. Notably, a large majority ($95.7 \pm 3.1\%$)
120 (s.d.) of the pair-wise differences was accounted by heterozygous vs homozygous genotypes.
121 Interestingly, a similar number of differences were observed between the two isolates of
122 identical DST obtained from carrier G (G2 vs G3, 1452 differences) than when these isolates
123 were compared to isolate G1 that had a different DST (G1 vs G2, 1446 differences; and G3 vs
124 G1, 1314 differences) (**Table 1**).

125 Loss-of-heterozygosity events encompassing two or more heterozygous sites have been
126 shown to account for a large part of the genetic differences that occur between isolates from a
127 given genetic cluster^{12,48}. LOH events result from the loss of chromosomes (*e.g.*
128 chromosomes 6 and 7 in the three isolates from carrier G; **Fig. 2: panel a**); mitotic
129 recombination (MR) or break-induced replication (BIR) that then extend from one internal
130 chromosomal location to the chromosomal end (*e.g.* at the right arm of chromosome 1 of all
131 isolates from carrier A; **Fig. 2: panel a**); and gene conversion events that are internal to a
132 chromosome. Importantly, we observed that most differences between isolates from the same
133 carrier resulted from LOH events (**Table 1**). The number of LOH events observed between
134 isolates varied across carriers (254, 106, and 145 on average for samples A, D, and G,
135 respectively). We qualified the LOH events by their minimal size MinS, the distance
136 between the two heterozygous SNPs located within the LOH and closest to its flanks
137 (**Fig. 3A** and Materials and Methods). Results presented in **Figure 3B** and **Figure 4** showed
138 that 95% of the LOH events had MinS values < 3 kb. Larger events were also observed (**Fig.**
139 **2** and **Fig. 3B**). For instance, isolate A3 differed from isolates A1 and A2 by a 454,851 bp

140 MR/BIR event on the left arm of chromosome 2, and this accounted for 2271 differences
141 between these strains (**Fig. 2, panel a**). Isolate G2 differed from isolates G1 and G3 by a 3538
142 bp MR/BIR event on right arm of chromosome 2, and this accounted for 31 differences
143 between these strains (**Fig. 2: panel a**). Additional large LOH events were detected in large,
144 almost homozygous regions shared by the three isolates of a carrier, and thus did not account
145 for major differences between these strains (**Fig. 5**).

146 The 5-prime ends of all LOH were mapped on the 8 chromosomes to appreciate the density of
147 these events. As shown in **Figure 2, panel b**, these LOH events were randomly distributed
148 across heterozygous regions and affected all chromosomes. Most LOH did not extend to the
149 chromosomal ends, suggesting that they might correspond to gene conversion events. In
150 addition, the number of LOH differed between samples. Isolates from carrier A differed by
151 more LOH than isolates from carriers D or G. To illustrate the complexity of these events, a
152 complete description of the coding regions impacted by LOH events on chromosome 1 of
153 isolates G1–G3 is presented in **Supplementary Table S3**. We observed alternating
154 homozygous and heterozygous regions between the 3 genomes. In order to ensure that these
155 LOH events were not artifacts, we performed a Sanger sequencing of 19 events observed on
156 Chr1 of isolates G1-G3. Among the 19 events tested, 18 (95%) were confirmed by Sanger
157 sequencing.

158

159 **Within-host *C. albicans* genomic diversity was significantly higher than that of *in vitro***
160 **grown *C. albicans***

161 We then evaluated whether the genetic variations observed across carrier isolates could result
162 from the propagation steps performed *in vitro* (growth on ChromAgar plates and in liquid
163 YPD for -80°C storage and genomic DNA preparation) and/or our analytical pipeline. To this
164 aim, we plated cells from three *C. albicans* strains from our collection (hereafter referred to as

165 strains X [DST 2281], Y [DST 1047], and Z [DST 1400]) on rich medium, selected three
166 colonies (clonal isolates) per strain (namely isolates X1-3, Y1-3, and Z1-3), prepared genomic
167 DNA from these isolates, and subjected them to deep-sequencing using the Illumina
168 technology (average sequencing depth, 102×; range, 58–145× [**Supplementary Table S2**]).
169 SNPs were identified as described above. Strikingly, clonal isolates obtained from any given
170 strain differed by an average of 400 differences (range, 352–441), with a large majority (97.9
171 ± 1.6%) (s.d.) being accounted by heterozygous vs homozygous genotypes (**Table 1**). No
172 large-scale genetic changes, such as aneuploidy or long-range LOH, were observed between
173 the clonal isolates. Rather, all pair-wise comparisons revealed short-range LOH (average
174 number, 70; range, 60–81) whose MinS distribution was displaced towards smaller sizes
175 compared to that observed for isolates obtained from carriers (**Fig. 4 and Supplementary**
176 **Fig. S1**). Similar to those observed in isolates from healthy carriers, these LOH events were
177 randomly distributed across heterozygous regions and affected all chromosomes
178 (**Supplementary Fig. S2: panels b**).

179 The genomic variability between the carrier isolates (A1-3, D1-3, and G1-3) or clonal isolates
180 (X1-3, Y1-3, and Z1-3) was quantified using the numbers of pair-wise SNPs and pair-wise
181 LOH events. For each criterion, the diversity of the carrier isolates was significantly higher
182 than the diversity of clonal isolates ($p=0.007$ for SNPs, and $p= 0.0013$ for LOH events; one
183 way ANOVA test) (**Fig. 6A and B**). By focusing on SNPs, we studied the pair-wise
184 variability (mutation frequency) across different types of genomic features that have been
185 defined in the *C. albicans* genome, namely intergenic regions, ORFs, repeat regions, long-
186 terminal repeats (LTRs), and retrotransposons. A two-way ANOVA showed a significant
187 difference between carrier isolates and clonal isolates and between genomic features. Indeed,
188 repeat regions were significantly more mutated than other features of whatever group that was
189 considered (carrier or clonal isolates; $p < 0.01$; post-hoc test: Tukey honestly significant

190 difference [HSD]). Furthermore, carrier isolates were significantly more variable in intergenic
191 regions than clonal isolates ($p < 0.01$, post-hoc test: Tukey HSD) (**Fig. 6C**).

192 Because LOH events in clonal isolates were significantly smaller than those observed in
193 carrier isolates, we evaluated whether they could in part result from miscalling of
194 polymorphic positions. Toward this aim, we inspected the allelic ratio (ABHet = number of
195 reads for non-reference allele/total number of reads) of all polymorphic positions involved in
196 LOH events, with the expectation that they should have a Gaussian distribution centred on a
197 ABHet value of 0.5 for a diploid genome. The results presented in **Figure 7A** showed that
198 ABHet values obtained for carrier isolates had a bimodal distribution with the majority
199 achieving a Gaussian distribution centred on a ABHet value of 0.5 and a minority achieving a
200 Gaussian distribution centred on a ABHet value of 0.85. The second population might
201 correspond to polymorphisms located in repeat regions and is partly excluded in our
202 analytical pipeline that only considers polymorphisms with ABHet values >0.15 and <0.85 as
203 heterozygous. Yet, because the two curves overlap, some polymorphisms in the second
204 population are not excluded and may cause artefacts in the identification of LOH. Notably,
205 ABHet values obtained for clonal isolates had a Gaussian distribution centred on a value
206 around 0.85 (**Fig. 7B**). It may ensue that most LOH events identified in pair-wise comparison
207 of clonal isolates are artefactual, corresponding partly to repeat regions (**Fig. 6C**) and
208 explaining their smaller size compared to those observed in pair-wise comparisons of carrier
209 isolates (**Fig. 4**). Thus, the number of LOH events in clonal isolates might be overestimated
210 (as well the number of LOH events in carrier isolates that also show a population of LOH
211 events similar in size to those in the clonal isolates [**Fig. 4**]), thus reinforcing our conclusion
212 that the level of genomic diversity observed between the *C. albicans* isolates obtained from
213 oral carriers was significantly higher than that observed between clonal isolates obtained *in*

214 *vitro*. Taken together, our results showed that a single oral sample exhibits a genetically
215 heterogeneous *C. albicans* cell population.

216

217 **Discussion**

218 This work represents, to our knowledge, the first genome-wide analysis of the within-host
219 diversity of *C. albicans* in healthy individuals. We first explored the level of oral *Candida*
220 carriage in a group of healthy young individuals. As expected, the carriage of *C. albicans* was
221 common in this population with a prevalence of 17.9%, while other *Candida* species were
222 rarely found, emphasizing the predominant role of *C. albicans* in the normal oral microbiota
223 of humans^{4,39}. We then studied the genomic diversity of *C. albicans* within each healthy
224 carrier. To this aim, we analysed several isolates from oral samples and determined the
225 intrinsic level of genomic diversity in the context of commensalism. Overall, our results
226 showed that a single oral sample exhibited a genetically heterogeneous *C. albicans*
227 population, with isolates differing from each other by multiple short-range LOH events
228 corresponding to gene conversion tracts. MLST is insufficient to pinpoint this heterogeneity,
229 except in rare cases where a short-range LOH or SNP affected one of the MLST loci, as was
230 the case for only one isolate in our study.

231 Whole genome sequencing offers opportunities to extend the analysis of *C. albicans* diversity
232 at the genomic level^{49,50}. Unlike longitudinal studies evaluating the evolution of the *C.*
233 *albicans* genome in patients using only one colony per sample at different time points^{50,51}, our
234 analysis characterised several isolates from the same oral sample collected at a single time
235 point. For each sample, we compared the whole genome of 3 different isolates, focusing on
236 pair-wise SNPs and LOH events. Indeed, the occurrence of these events has been correlated
237 with the response of *C. albicans* to various stresses both *in vitro* and *in vivo*^{30,31,33,52-54}.

238 However, little is known about the occurrence and the role of these events during
239 commensalism in humans⁵⁵.

240 As mentioned above, while a low level of genotypic diversity was found by MLST, the
241 genome-wide analysis identified numerous genetic variants, mainly short-range (gene
242 conversion) LOH events, between carrier isolates, indicating a genetically heterogeneous *C.*
243 *albicans* population in oral samples of healthy individuals. To confirm that these events were
244 not due to sequencing errors, to our genome analysis pipeline, or to clonal reproduction *in*
245 *vitro*, we quantified and compared the extent of variations observed between carrier isolates
246 and those observed between clonal isolates. Our results showed that the level of genomic
247 diversity observed between the *C. albicans* isolates obtained from oral carriers was
248 significantly higher than that observed between clonal isolates propagated *in vitro*, confirming
249 that oral samples harbour a genetically heterogeneous *C. albicans* population. The diversity
250 across carrier isolates may reflect the evolutionary consequences of life-long commensal
251 carriage, initiated through mother-to-child transmission, and the recurrent exposure of
252 individual isolates to genotoxic stress, such as upon interaction with immune cells or
253 environmental changes. Differences in the extent of divergence between isolates from a single
254 carrier (*e.g.* isolates from carrier A were more divergent between each other than isolates
255 from carrier B or G) may reflect a longer divergence time or differences in terms of genomic
256 plasticity. Importantly, genomic differences across carrier isolates were mostly accounted by
257 short-range LOH events. No example of aneuploidy was observed across the nine investigated
258 carrier isolates, and only a few long-range BIR/MC events were identified. SNPs between the
259 carrier isolates were significantly enriched in intergenic regions, while a trend for enrichment
260 in ORFs was also observed (**Fig. 6C**). This was in contrast with repeat regions, LTRs, and
261 repeat regions that showed similar mutation frequencies between carrier and clonal isolates.
262 Hence, SNPs and LOH events distinguish isolates within a carrier, because their enrichment

263 in intergenic regions and ORFs may have functional consequences. Of note, we observed a
264 high frequency of SNPs in repeat regions for both carrier and clonal isolates, which may
265 reflect the limitations of SNP analyses within structurally complex regions⁵⁶.
266 Our results suggest that *C. albicans* diversity in the healthy host might be higher than could
267 have been anticipated from studies that have used MLST. This probably reflects the fact that
268 most epidemiological studies have typed one isolate per individual and therefore did not
269 address intra-host variability. More importantly, it reflects the higher discriminatory power of
270 whole genome sequencing over MLST. We suppose that diversifying lineages coexist in each
271 human individual, and that this diversity could enable rapid adaptation under stress exposure.
272 Indeed, the human digestive tract is the main reservoir for infection from which the most
273 suitable variant could emerge^{7,8}. It is known that *C. albicans* genomic variations, ranging
274 from SNPs to large scale genetic changes, can facilitate adaptation to environmental changes
275 and improve the persistence of the fungus in various host niches⁵⁷. The main example is the
276 link to genomic variations in *C. albicans* isolates that were resistant to antifungal drugs^{52,58}.
277 However, other studies have shown the implication of genetic variations on *C. albicans*
278 pathogenicity by modifying the expression of virulence factors and the interaction with the
279 host^{9,59-63}.
280 The natural diversity observed during commensalism might explain the remarkable ability of
281 *C. albicans* to adapt to stress conditions within the host as well as the disparate response
282 observed during infections. Our results provide crucial information for future longitudinal
283 studies aimed at evaluating genomic evolution in healthy individuals or patients. Indeed, these
284 studies offer a unique view at the *in vivo* evolution of *C. albicans*⁵⁰. For instance, work by
285 Ford et al. (2015) has revealed the occurrence of numerous genomic changes over time in *C.*
286 *albicans* strains isolated from patients throughout the course of a fluconazole treatment⁵⁰.
287 However, in this study, only one isolate was analysed at each time point, and it cannot be

288 excluded that some of the variations observed between isolates may have preexisted due to
289 intrinsic diversity of *C. albicans* in the healthy host. Hence, future longitudinal studies should,
290 whenever possible, assess multiple isolates at each time point in order to disentangle pre-
291 existing variations from those acquired in the course of the study.

292

293 **Material and methods**

294 **Study cohort and sample collection**

295 The studied population comprised of 56 20-to-22-year-old healthy volunteers from a French
296 military school. The students were living across 2 different dormitories, one for members of
297 the swim team and the other for members of the volleyball team. Both groups shared the same
298 hygiene facilities and canteen, and none of them has taken antibiotics within 6 months. They
299 also attended common classes and military preparation. The general information had been
300 given collectively during classes by one of the investigators and the students willing to
301 participate agreed to have an oral swabbing.

302 Swabs were extemporaneously plated on a chromogenic plate BBL™ CHROMagar™
303 *Candida* (BioMérieux, Marcy-l'Étoile, France) and incubated for 5 days at 37°C. Colonies
304 from each plate were enumerated. *C. albicans* detection was based on growth colour on
305 CHROMagar™ plate typically represented in green. Up to 8 green single colonies
306 (presumptive *C. albicans* strains) were picked and sub-cultured onto separate media for
307 species identification by mass spectrometry-type Maldi-TOF and frozen in 30% glycerol at -
308 80°C. For the next analysis, the yeast cells were grown at 30°C in liquid or solid media in
309 YPD (1% yeast extract, 2% peptone, 2% dextrose). Solid media were obtained by adding 2%
310 of agar.

311

312

313 **Multilocus Sequence Typing**

314 *C. albicans* isolates obtained from primo culture of oral swabbing of each healthy carrier were
315 typed using MLST as described previously⁶⁴. Briefly, alleles of the seven housekeeping genes
316 (*AAT1a*, *ACCI*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b*) were amplified and sequenced for
317 each isolate. Sequencing was performed on both strands using an ABI Prism 3130xl genetic
318 analyzer (Applied Biosystems, Foster City, CA, USA). Allele and diploid sequence type
319 (DST) assignments were determined using the *C. albicans* multilocus sequence typing
320 database (<http://pubmlst.org/calbicans>).

321

322 **Illumina sequencing**

323 The genomic DNA was extracted from subcultures using the QIAamp® DNA Mini Kit
324 (Qiagen, Courtaboeuf, France), according to the manufacturer's instruction, with an additional
325 mechanical lysis step (FastPrep; MP Biomedicals, Illkirch, France) following the addition of
326 chemical lysis buffer. Genomic DNA was processed to prepare libraries for Illumina
327 sequencing. Libraries were prepared using the NEXTflex™ PCR-Free DNA Sequencing kit
328 (Illumina®) according to the manufacturer's recommendations. HiSeq2000 or HiSeq2500
329 platforms was used to generate 101 bp paired-ends reads.

330

331 **Genomic variants analysis**

332 Sequences were mapped to the genome of *C. albicans* reference strain SC5314, assembly 22
333 (version A22-s06-m01-r01), available from CGD⁶⁵ using BWA v0.7.7 with default
334 parameters⁶⁶. The next processing was performed with the Genome analysis toolkit v3.1⁶⁷. To
335 minimize false-positive SNP calls near insertion/deletion events, poorly aligning regions were
336 identified and realigned using the GATK RealignerTargetCreator and IndelRealigner
337 modules. The variant calling was performed with a GATK HaplotypeCaller. Poor quality

338 SNPs were filtered using the GATK VariantFiltration module, with best practices
339 recommended annotation for hard filters (QD < 2.0, MQ < 40.0, FS > 60.0, HaplotypeScore
340 >13.0, MQRankSum< -12.5, ReadPosRankSum< -8.0). Then we used several dedicated
341 scripts to select SNPs identified at positions with a minimal sequencing depth of 18×. We
342 evaluated the allelic balance for heterozygous calls (ABHet) and homozygous calls (ABHom)
343 with AlleleBalance annotation GATK module. Heterozygous SNPs were defined as positions
344 where 15% or more of the calls showed one allele, and 85% or less of the calls showed a
345 second allele. Homozygous SNPs were defined as positions where more than 98% of the calls
346 differed from the reference genome⁶⁸. The output files gather high quality SNPs used to
347 compare each pair of genomes from isolates or clonal colonies and a complete LOH analysis
348 report (number of event, minimal and maximal size). LOH events were defined by at least 2
349 successive losses of heterozygous SNP positions (transition from a heterozygous to
350 homozygous position between two genome comparisons). The minimal LOH event size
351 (MinS) was defined by the distance between the two heterozygous positions located within
352 the LOH event and closest to its flanks. The maximal LOH event size (MaxS) was defined by
353 the number of base pairs between the heterozygous positions found before and after the LOH
354 event observed on the genome (**Fig. 3A**). At the end of the process, the dedicated script
355 generated figures summarizing all of the data, such as sequencing depth for the identification
356 of aneuploidies and heterozygosity density maps (performed for 1 kb sliding windows)
357 plotted across the 8 chromosomes as described previously^{69,70}.

358

359 **Sanger sequencing**

360 A selection of ORFs impacted by a LOH event across chromosome 1 of isolates from
361 individual G (G1, G2, and G3) were confirmed by Sanger sequencing. PCR was performed in
362 an Eppendorf Mastercycler ep gradient thermal cycler. The PCR mixture contained 1 µl of the

363 extracted genomic DNA, 2 µl of the 10× PCR buffer; 2 µl MgCl₂ (50 mM); 1 µl of a mix of
364 deoxynucleoside triphosphates (dNTP) (5 mM); 0.5 µl (each) primer (10 µM); 0.2 µl of Taq
365 polymerase (Invitrogen), and water to a final volume of 20 µl. Primers used in this study are
366 listed in **Supplementary Table S4**. The following conditions were used: initial denaturation
367 at 94°C for 3 min; 30 cycles with denaturation at 94°C for 40 s, annealing at 54°C for 40 s,
368 and extension at 72°C for 1 min/kb; and a final extension time at 72°C for 10 min. The PCR
369 products were verified by electrophoresis on a 1% agarose gel. The PCR products were then
370 sequenced at Eurofins sequencing facility using ABI 3730XL sequencing machines (Applied
371 Biosystems). Sequence analysis and SNP detection were performed using the SeqScape™
372 v3.0 software.

373

374 **Statistical analysis**

375 Statistical significance was determined using one way ANOVA (**Fig. 6A and B**), and two-
376 way ANOVA with repeated measures and post-hoc Tukey HSD tests (**Fig. 6C**). A $p < 0.01$
377 was considered significant and denoted by a double asterisk.

378

379 **Ethics Statement**

380 The study was made in accordance with the Declaration of Helsinki and the best
381 National Recommendations at the moment of the sampling (Ecole Polytechnique France
382 20112008).

383

384 **Data availability**

385 All code and relevant datasets generated during and/or analysed during the current
386 study are available from the corresponding author upon request. Raw reads have been

387 deposited at the NCBI Sequence Read Archive under BioProject ID PRJNA487353
388 [<https://www.ncbi.nlm.nih.gov/bioproject/487353>].

389

390 **References**

- 391 1. Forche, A. Large-Scale Chromosomal Changes and Associated Fitness Consequences
392 in Pathogenic Fungi. *Curr. Fungal Infect. Rep.* **8**, 163–170 (2014).
- 393 2. Kleinegger, C. L., Lockhart, S. R., Vargas, K. & Soll, D. R. Frequency, intensity, species,
394 and strains of oral *Candida* vary as a function of host age. *J. Clin. Microbiol.* **34**, 2246–
395 2254 (1996).
- 396 3. Hoffmann, C. *et al.* Archaea and fungi of the human gut microbiome: correlations
397 with diet and bacterial residents. *PLoS One* **8**, e66019 (2013).
- 398 4. Ghannoum, M. A. *et al.* Characterization of the oral fungal microbiome (mycobiome)
399 in healthy individuals. *PLoS Pathog.* **6**, e1000713 (2010).
- 400 5. Gow, N. A. R. A developmental program for *Candida* commensalism. *Nat. Genet.* **45**,
401 967–968 (2013).
- 402 6. Naglik, J. R., König, A., Hube, B. & Gaffen, S. L. *Candida albicans*-epithelial interactions
403 and induction of mucosal innate immunity. *Curr. Opin. Microbiol.* **40**, 104–112
404 (2017).
- 405 7. Kullberg, B. J. & Arendrup, M. C. Invasive Candidiasis. *N. Engl. J. Med.* **373**, 1445–1456
406 (2015).
- 407 8. Poulain, D. *Candida albicans*, plasticity and pathogenesis. *Crit. Rev. Microbiol.* **41**,
408 208–217 (2015).
- 409 9. Hirakawa, M. P. *et al.* Genetic and phenotypic intra-species variation in *Candida*
410 *albicans*. *Genome Res.* **25**, 413–425 (2015).
- 411 10. Hirakawa, M. P., Chyou, D. E., Huang, D., Slan, A. R. & Bennett, R. J. Parasex Generates
412 Phenotypic Diversity de Novo and Impacts Drug Resistance and Virulence in *Candida*
413 *albicans*. *Genetics* **207**, 1195–1211 (2017).

- 414 11. Hickman, M. A. *et al.* The 'obligate diploid' *Candida albicans* forms mating-competent
415 haploids. *Nature* **494**, 55–59 (2013).
- 416 12. Ropars, J. *et al.* Gene flow contributes to diversification of the major fungal pathogen
417 *Candida albicans*. *Nat. Commun.* **9**, 2253 (2018).
- 418 13. Butler, G. *et al.* Evolution of pathogenicity and sexual reproduction in eight *Candida*
419 genomes. *Nature* **459**, 657–662 (2009).
- 420 14. Odds, F. C. *et al.* Molecular phylogenetics of *Candida albicans*. *Eukaryot. Cell* **6**, 1041–
421 1052 (2007).
- 422 15. Shin, J. H. *et al.* Genetic diversity among Korean *Candida albicans* bloodstream
423 isolates: assessment by multilocus sequence typing and restriction endonuclease
424 analysis of genomic DNA by use of BssHII. *J. Clin. Microbiol.* **49**, 2572–2577 (2011).
- 425 16. Tavanti, A. *et al.* Population structure and properties of *Candida albicans*, as
426 determined by multilocus sequence typing. *J. Clin. Microbiol.* **43**, 5601–5613 (2005).
- 427 17. Dodgson, A. R., Dodgson, K. J., Pujol, C., Pfaller, M. A. & Soll, D. R. Clade-specific
428 flucytosine resistance is due to a single nucleotide change in the *FUR1* gene of
429 *Candida albicans*. *Antimicrob. Agents Chemother.* **48**, 2223–2227 (2004).
- 430 18. MacCallum, D. M. *et al.* Property differences among the four major *Candida albicans*
431 strain clades. *Eukaryot. Cell* **8**, 373–387 (2009).
- 432 19. Blignaut, E., Molepo, J., Pujol, C., Soll, D. R. & Pfaller, M. A. Clade-related amphotericin
433 B resistance among South African *Candida albicans* isolates. *Diagn. Microbiol. Infect.*
434 *Dis.* **53**, 29–31 (2005).
- 435 20. Chowdhary, A. *et al.* Whole Genome-Based Amplified Fragment Length
436 Polymorphism Analysis Reveals Genetic Diversity in *Candida africana*. *Front.*
437 *Microbiol.* **8**, (2017).

- 438 21. Calderón-Noreña, D. M. *et al.* A single nucleotide polymorphism uncovers a novel
439 function for the transcription factor Ace2 during *Candida albicans* hyphal
440 development. *PLoS Genet.* **11**, e1005152 (2015).
- 441 22. McCullough, M. J., Clemons, K. V. & Stevens, D. A. Molecular epidemiology of the
442 global and temporal diversity of *Candida albicans*. *Clin. Infect. Dis. Off. Publ. Infect.*
443 *Dis. Soc. Am.* **29**, 1220–1225 (1999).
- 444 23. Takakura, S. *et al.* Comparison of *Candida albicans* strain types among isolates from
445 three countries. *Int. J. Med. Microbiol. IJMM* **298**, 663–668 (2008).
- 446 24. McManus, B. A. *et al.* Enrichment of Multilocus Sequence Typing Clade 1 with Oral
447 *Candida albicans* Isolates in Patients with Untreated Periodontitis. *J. Clin. Microbiol.*
448 **50**, 3335–3344 (2012).
- 449 25. Bennett, R. J., Forche, A. & Berman, J. Rapid mechanisms for generating genome
450 diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring*
451 *Harb. Perspect. Med.* **4**, (2014).
- 452 26. Forche, A. *et al.* The parasexual cycle in *Candida albicans* provides an alternative
453 pathway to meiosis for the formation of recombinant strains. *PLoS Biol.* **6**, e110
454 (2008).
- 455 27. Bennett, R. J. & Johnson, A. D. Completion of a parasexual cycle in *Candida albicans*
456 by induced chromosome loss in tetraploid strains. *EMBO J.* **22**, 2505–2515 (2003).
- 457 28. Bennett, R. J. The parasexual lifestyle of *Candida albicans*. *Curr. Opin. Microbiol.* **28**,
458 10–17 (2015).
- 459 29. Hickman, M. A., Paulson, C., Dudley, A. & Berman, J. Parasexual Ploidy Reduction
460 Drives Population Heterogeneity Through Random and Transient Aneuploidy in
461 *Candida albicans*. *Genetics* **200**, 781–794 (2015).

- 462 30. Berman, J. Ploidy plasticity: a rapid and reversible strategy for adaptation to stress.
463 *FEMS Yeast Res.* **16**, (2016).
- 464 31. Forche, A. *et al.* Stress alters rates and types of loss of heterozygosity in *Candida*
465 *albicans*. *mBio* **2**, (2011).
- 466 32. Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B. & Berman, J. Acquisition of
467 aneuploidy provides increased fitness during the evolution of antifungal drug
468 resistance. *PLoS Genet.* **5**, e1000705 (2009).
- 469 33. Forche, A., Magee, P. T., Selmecki, A., Berman, J. & May, G. Evolution in *Candida*
470 *albicans* populations during a single passage through a mouse host. *Genetics* **182**,
471 799–811 (2009).
- 472 34. Chen, X., Magee, B. B., Dawson, D., Magee, P. T. & Kumamoto, C. A. Chromosome 1
473 trisomy compromises the virulence of *Candida albicans*. *Mol. Microbiol.* **51**, 551–565
474 (2004).
- 475 35. Forche, A. *et al.* Rapid Phenotypic and Genotypic Diversification After Exposure to
476 the Oral Host Niche in *Candida albicans*. *Genetics* **209**, 725–741 (2018).
- 477 36. Ford, C. B. *et al.* The evolution of drug resistance in clinical isolates of *Candida*
478 *albicans*. *eLife* **4**, e00662 (2015).
- 479 37. Odds, F. C. *et al.* *Candida albicans* strain maintenance, replacement, and
480 microvariation demonstrated by multilocus sequence typing. *J. Clin. Microbiol.* **44**,
481 3647–3658 (2006).
- 482 38. Angebault, C. *et al.* *Candida albicans* is not always the preferential yeast colonizing
483 humans: a study in Wayampi Amerindians. *J. Infect. Dis.* **208**, 1705–1716 (2013).
- 484 39. Jacobsen, M. D. *et al.* Mixed *Candida albicans* strain populations in colonized and
485 infected mucosal tissues. *FEMS Yeast Res.* **8**, 1334–1338 (2008).

- 486 40. Bounoux, M.-E. *et al.* Multilocus sequence typing reveals intrafamilial transmission
487 and microevolutions of *Candida albicans* isolates from the human digestive tract. *J.*
488 *Clin. Microbiol.* **44**, 1810–1820 (2006).
- 489 41. Diogo, D., Bouchier, C., d’Enfert, C. & Bounoux, M.-E. Loss of heterozygosity in
490 commensal isolates of the asexual diploid yeast *Candida albicans*. *Fungal Genet. Biol.*
491 *FG B* **46**, 159–168 (2009).
- 492 42. Xu, J. *et al.* Species and genotypic diversities and similarities of pathogenic yeasts
493 colonizing women. *J. Clin. Microbiol.* **37**, 3835–3843 (1999).
- 494 43. Takagi, Y. *et al.* Genotypes of *Candida albicans* isolated from healthy individuals and
495 their distribution in patients with oral candidiasis. *J. Infect. Chemother. Off. J. Jpn. Soc.*
496 *Chemother.* **19**, 1072–1079 (2013).
- 497 44. Zhang, N. *et al.* Multi-Locus Next-Generation Sequence Typing of DNA Extracted
498 From Pooled Colonies Detects Multiple Unrelated *Candida albicans* Strains in a
499 Significant Proportion of Patient Samples. *Front. Microbiol.* **9**, (2018).
- 500 45. Choo, K. H., Lee, H. J., Knight, N. J., Holmes, A. R. & Cannon, R. D. Multilocus sequence
501 typing (MLST) analysis of *Candida albicans* isolates colonizing acrylic dentures
502 before and after denture replacement. *Med. Mycol.* myw128 (2016).
503 doi:10.1093/mmy/myw128
- 504 46. Jones, T. *et al.* The diploid genome sequence of *Candida albicans*. *Proc. Natl. Acad. Sci.*
505 *U. S. A.* **101**, 7329–7334 (2004).
- 506 47. van het Hoog, M. *et al.* Assembly of the *Candida albicans* genome into sixteen
507 supercontigs aligned on the eight chromosomes. *Genome Biol.* **8**, R52 (2007).
- 508 48. Hirakawa, M. P. *et al.* Genetic and phenotypic intra-species variation in *Candida*
509 *albicans*. *Genome Res.* **25**, 413–425 (2015).

- 510 49. Dettman, J. R. *et al.* Evolutionary insight from whole-genome sequencing of
511 experimentally evolved microbes. *Mol. Ecol.* **21**, 2058–2077 (2012).
- 512 50. Ford, C. B. *et al.* The evolution of drug resistance in clinical isolates of *Candida*
513 *albicans*. *eLife* **4**, e00662 (2015).
- 514 51. Coste, A. *et al.* Genotypic evolution of azole resistance mechanisms in sequential
515 *Candida albicans* isolates. *Eukaryot. Cell* **6**, 1889–1904 (2007).
- 516 52. Selmecki, A., Forche, A. & Berman, J. Aneuploidy and isochromosome formation in
517 drug-resistant *Candida albicans*. *Science* **313**, 367–370 (2006).
- 518 53. Rustchenko, E. P., Howard, D. H. & Sherman, F. Variation in assimilating functions
519 occurs in spontaneous *Candida albicans* mutants having chromosomal alterations.
520 *Microbiol. Read. Engl.* **143 (Pt 5)**, 1765–1778 (1997).
- 521 54. Selmecki, A., Forche, A. & Berman, J. Genomic plasticity of the human fungal
522 pathogen *Candida albicans*. *Eukaryot. Cell* **9**, 991–1008 (2010).
- 523 55. Pande, K., Chen, C. & Noble, S. M. Passage through the mammalian gut triggers a
524 phenotypic switch that promotes *Candida albicans* commensalism. *Nat. Genet.* **45**,
525 1088–1091 (2013).
- 526 56. Zojer, M. *et al.* Variant profiling of evolving prokaryotic populations. *PeerJ* **5**, e2997
527 (2017).
- 528 57. Braunsdorf, C. & LeibundGut-Landmann, S. Modulation of the Fungal-Host
529 Interaction by the Intra-Species Diversity of *C. albicans*. *Pathog. Basel Switz.* **7**,
530 (2018).
- 531 58. White, T., Bruns, T., Lee, S. & Taylor, J. Amplification and direct sequencing of fungal
532 ribosomal RNA genes for phylogenetics. in *PCR Protocols: A Guide to Methods and*
533 *Applications* 315 – 322 (1990).

- 534 59. Schönherr, F. A. *et al.* The intraspecies diversity of *C. albicans* triggers qualitatively
535 and temporally distinct host responses that determine the balance between
536 commensalism and pathogenicity. *Mucosal Immunol.* **10**, 1335–1350 (2017).
- 537 60. Zhou, Z., Jordens, Z., Zhang, S., Zhang, N. & Schmid, J. Highly mutable tandem DNA
538 repeats generate a cell wall protein variant more frequent in disease-causing
539 *Candida albicans* isolates than in commensal isolates. *PLOS ONE* **12**, e0180246
540 (2017).
- 541 61. Moyes, D. L. *et al.* Candidalysin is a fungal peptide toxin critical for mucosal infection.
542 *Nature* **532**, 64–68 (2016).
- 543 62. Hube, B. From commensal to pathogen: stage- and tissue-specific gene expression of
544 *Candida albicans*. *Curr. Opin. Microbiol.* **7**, 336–341 (2004).
- 545 63. Schmid, J. *et al.* Increased Mortality in Young Candidemia Patients Associated with
546 Presence of a *Candida albicans* General-Purpose Genotype. *J. Clin. Microbiol.* **49**,
547 3250–3256 (2011).
- 548 64. Bougnoux, M.-E. *et al.* Collaborative Consensus for Optimized Multilocus Sequence
549 Typing of *Candida albicans*. *J. Clin. Microbiol.* **41**, 5265–5266 (2003).
- 550 65. Binkley, J. *et al.* The *Candida* Genome Database: the new homology information page
551 highlights protein similarity and phylogeny. *Nucleic Acids Res.* **42**, D711-716 (2014).
- 552 66. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
553 transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).
- 554 67. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for
555 analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303
556 (2010).

- 557 68. Feri, A. *et al.* Analysis of Repair Mechanisms following an Induced Double-Strand
558 Break Uncovers Recessive Deleterious Alleles in the *Candida albicans* Diploid
559 Genome. *mBio* **7**, (2016).
- 560 69. Loll-Krippleber, R. *et al.* A study of the DNA damage checkpoint in *Candida albicans*:
561 uncoupling of the functions of Rad53 in DNA repair, cell cycle regulation and
562 genotoxic stress-induced polarized growth. *Mol. Microbiol.* **91**, 452–471 (2014).
- 563 70. Abbey, D. A. *et al.* YMAP: a pipeline for visualization of copy number variation and
564 loss of heterozygosity in eukaryotic pathogens. *Genome Med.* **6**, (2014).
- 565 71. Wikivisual. [https://www.wikihow.com/Collect-DNA#/Image:Collect-DNA-Step-10-](https://www.wikihow.com/Collect-DNA#/Image:Collect-DNA-Step-10-Version-2.jpg)
566 [Version-2.jpg](https://www.wikihow.com/Collect-DNA#/Image:Collect-DNA-Step-10-Version-2.jpg).

567

568

569 **Acknowledgements**

570 E.S. was the recipient of a Ph.D. grant from DIM Malinf -Région Ile-de-France (DIM Malinf
571 – Région Ile-de-France 2014 to M-E. B.).

572 This work has been supported by the French Government’s Investissement d’Avenir program
573 (Laboratoire d’Excellence Integrative Biology of Emerging Infectious Diseases, ANR-10-
574 LABX-62-IBEID to C.d’E). We thank Anne-Flore Bonvalot, Camille Phan and Charles
575 Pinard for their precious help in this work.

576

577 **Author Contributions**

578 E.S., M.P., C.E., and M-E.B. designed and planned the study. E.S., C.M., N.S., M.P., and M-
579 E.B. supplied isolates and whole-genome sequencing. E.S., M.P., and N.S. performed
580 laboratory experiments. E.S., C.M., N.S., M.P., C.E., and M-E.B. analysed the data. E.S.,
581 C.E., and M-E.B. drafted the manuscript. All authors reviewed and contributed to the final
582 manuscript. All authors read and approved the final manuscript.

583

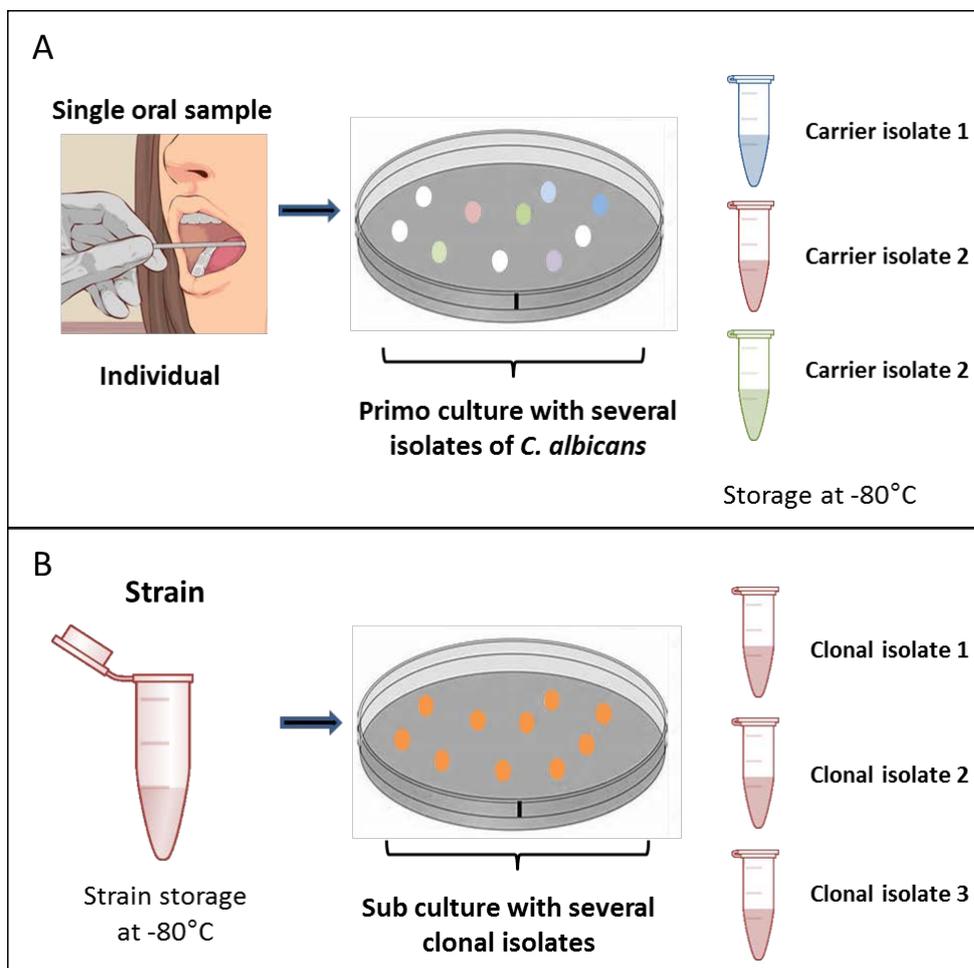
584 **Additional Information**

585 **Competing interests.** The authors declare no competing interests.

586

587 **Table and Figures**

588 **Figure 1: Protocol of the *C. albicans* genomic diversity analysis** **A. Genomic diversity**
589 **between “carrier isolates”.** In this context, 3 different “carrier isolates” were selected on the
590 “primo culture” of the single oral swabbing of the individual. The whole genome of the 3
591 isolated colonies was analysed in order to determine the population genomic diversity within
592 single oral sample. In total, 3 different single oral samples were selected from three
593 independent individuals (A, D and G). **B. Genomic variability between “clonal isolates.”**
594 In this context, 3 different clonal isolates were selected on the “sub culture” of the different
595 strains. The whole genome of the 3 clonal isolates was analysed in order to determine the
596 basal genomic variability between clones. In total, 3 different independent strains from our
597 collection of *C. albicans* clinical strains were analysed (X, Y and Z). (Part of the illustration⁷¹
598 is copyright-free).



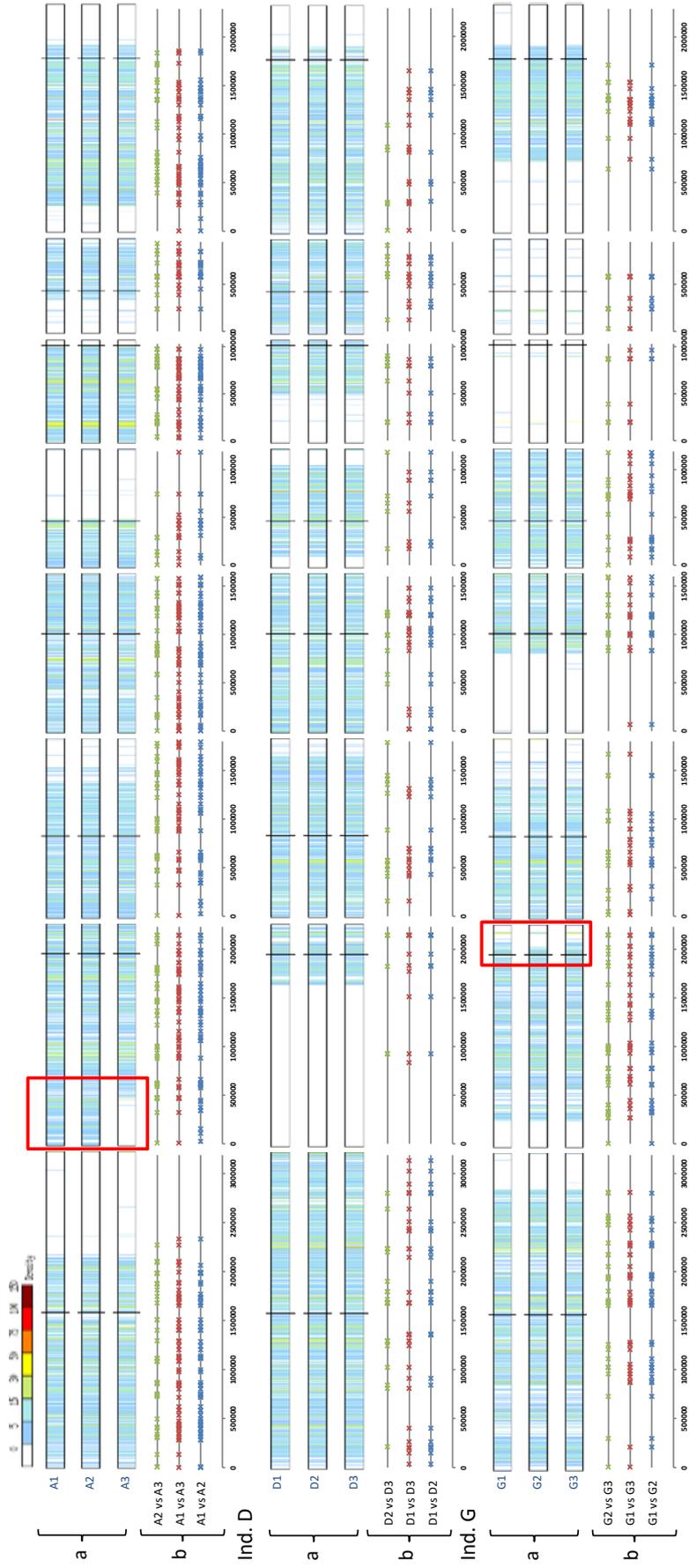
599

600 **Table 1. Characteristics of the SNPs and LOH event detected from pair-wise genome**
601 **comparison from carrier isolates (individual samples A, D and G) and clonal isolates**
602 **(strains X, Y and Z).**

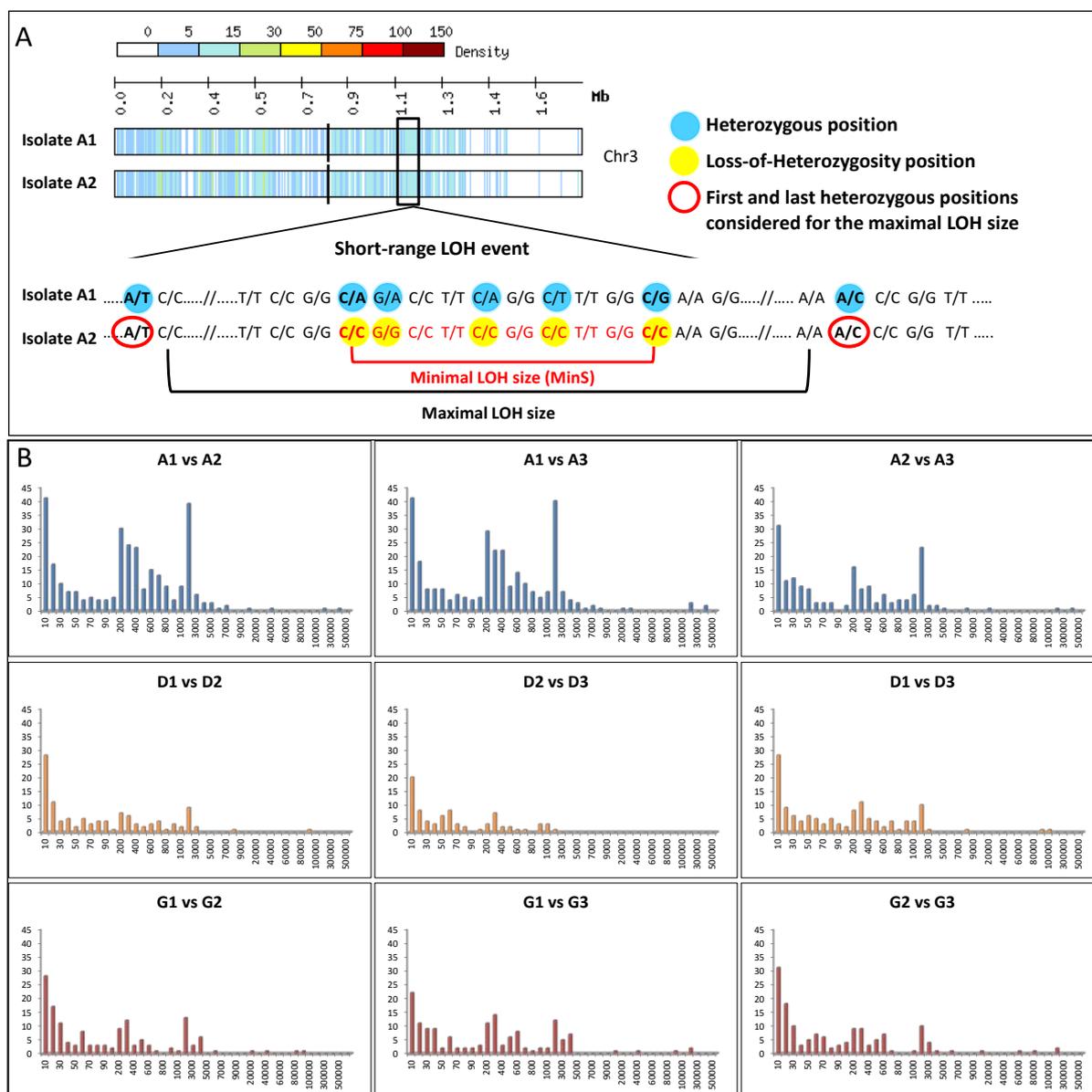
Material	Genome comparisons	SNP characteristics					LOH characteristics						
		Total SNPs	Heterozygous SNPs	Homozygous SNPs	Intergenic SNPs	Coding regions SNPs	Non synonymous SNPs	Number of LOH	Number of SNPs involved in LOH	Median of the maximal LOH size (MaxS) (bp)	Median of the minimal LOH size (MinS) (bp)	Range of the minimal LOH size (MinS) (bp)	Range of the maximal LOH size (MaxS) (bp)
Single oral sample from													
Carrier isolate													
Individual A	A1 vs A2	2 528	2 517	11	1 144	1 384	531	291	2225				
	A2 vs A3	3 843	3 825	18	1 643	2 200	883	173	3667	1184	21- 458 739	217	2-454 851
	A3 vs A1	5 119	5 082	38	2 247	2 872	1 114	297	4815				
	D1 vs D2	767	735	32	351	416	155	111	634				
	D2 vs D3	549	506	43	255	294	106	78	426				
	D3 vs D1	955	887	68	488	467	186	128	774	738	10-339 442	57	2-91 287
	G1 vs G2	1 446	1 361	85	694	752	282	145	1188				
	G2 vs G3	1 452	1 343	109	559	893	344	145	1186	947	11-359 209	80	2-198 176
	G3 vs G1	1 314	1 254	60	573	741	268	146	1117				
Strain													
Clonal isolate													
X	X1 vs X2	395	381	14	186	229	87	66	280				
	X2 vs X3	435	418	17	172	263	93	68	320	354	7-96 382	19	2-2 161
	X3 vs X1	441	422	19	177	264	112	73	300				
Y	Y1 vs Y2	394	390	4	136	258	92	72	276				
	Y2 vs Y3	393	385	8	111	282	111	66	281	334	17-10 099	20	2-2 722
	Y3 vs Y1	393	383	10	129	264	112	81	303				
Z	Z1 vs Z2	352	350	2	66	286	117	60	263				
	Z2 vs Z3	440	440	0	95	345	133	77	273	262	6-78 344	18	2-2 155
	Z3 vs Z1	360	358	2	103	257	118	64	351				

604 **Figure 2. Representation of LOH events between genomes from the different carrier**
605 **isolates selected from individual (Ind.) A, D and G. Panels a: Detection of large LOH**
606 **event by chromosome.** For each genome from isolates, heterozygous SNPs density was
607 mapped on the 8 chromosomes (1 Kb sliding windows). Homozygous regions are indicated in
608 light or white colour. Appearance of large LOH event (MR/BIR) between genomes is
609 indicated by a red square. **Panels b: Density of LOH events by chromosome.** For each pair
610 of genome comparisons the starting location of all LOH events was mapped on the 8
611 chromosomes.

Individual A

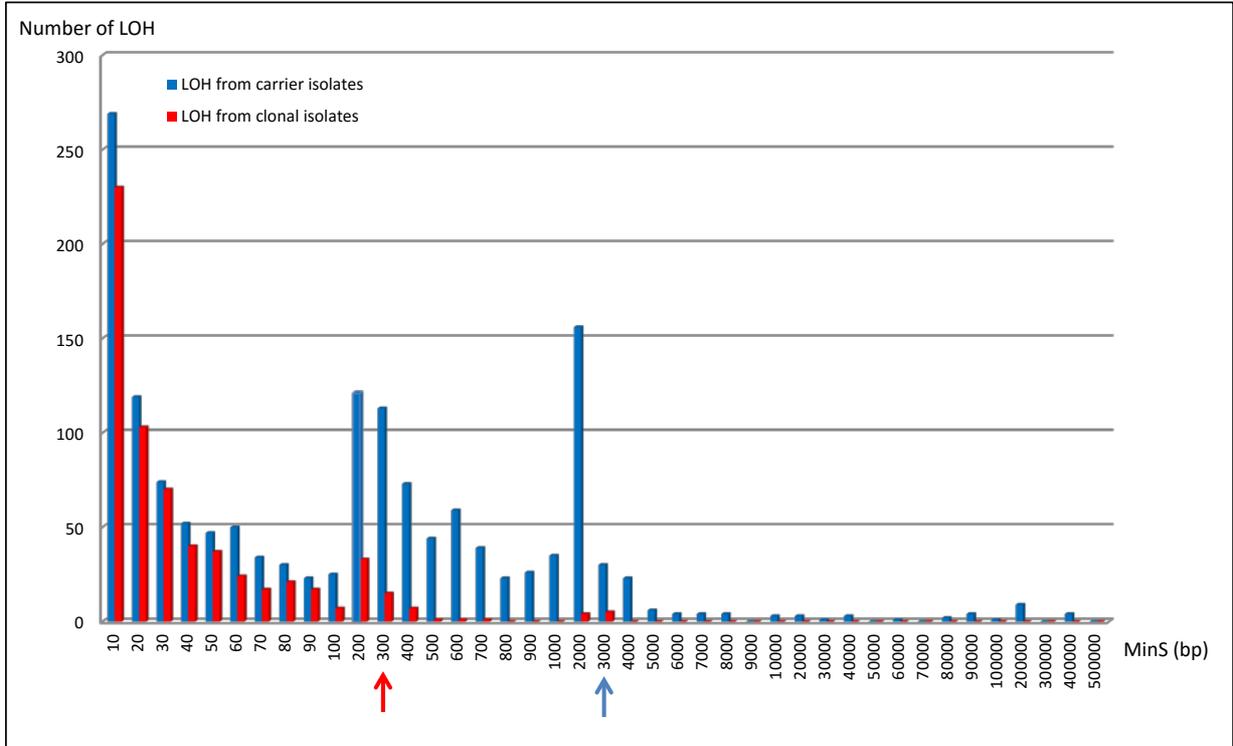


613 **Figure 3. LOH characterisation A. Definition of LOH event and size determination.**
 614 Example of one short-range LOH event observed on chromosome 3 between the genome
 615 from carrier isolates A1 and A2. The first part represents the heatmap of the heterozygous
 616 SNPs density for the chromosome 3 of the 2 genomes (1 Kb sliding windows). The second
 617 part represents the associated diploid genome sequences. In this example the LOH MinS is 10
 618 bps. **B. Distribution of MinS LOH event from the 9 pair-wise carrier isolates**
 619 **comparisons from individual sample A, D and G.** The x-axis corresponds to the classes of
 620 LOH size (MinS in bp). The y-axis corresponds to the number of events observed by MinS
 621 classes.



622

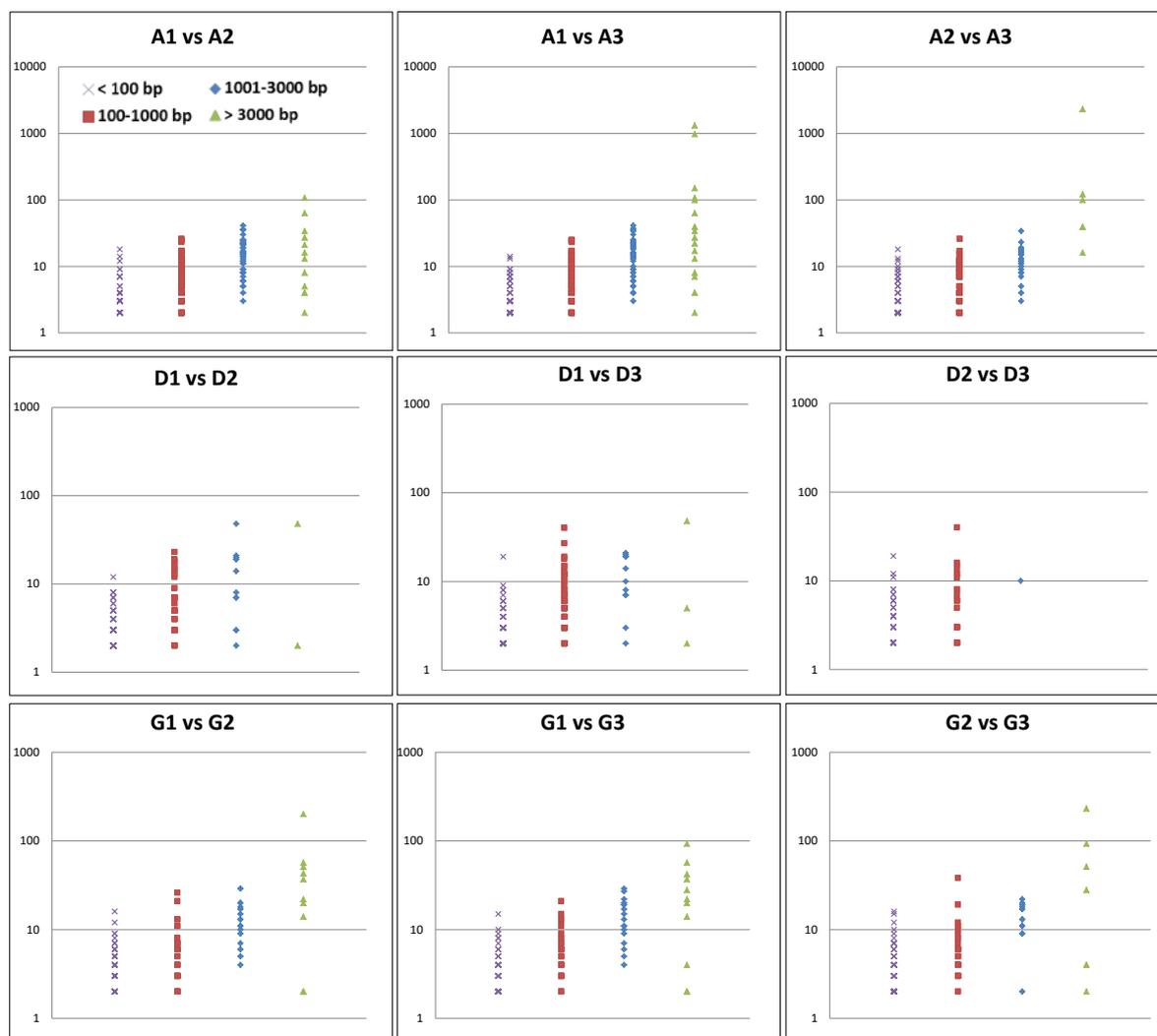
623 **Figure 4. Comparison of MinS LOH event distribution between carrier isolates and**
624 **Clonal isolates.** The arrows indicate the MinS of LOH that include 95% of the total number
625 of LOH event detected, 3000 bp for carrier isolates (blue) and 300 bp for clonal isolates (red)



626

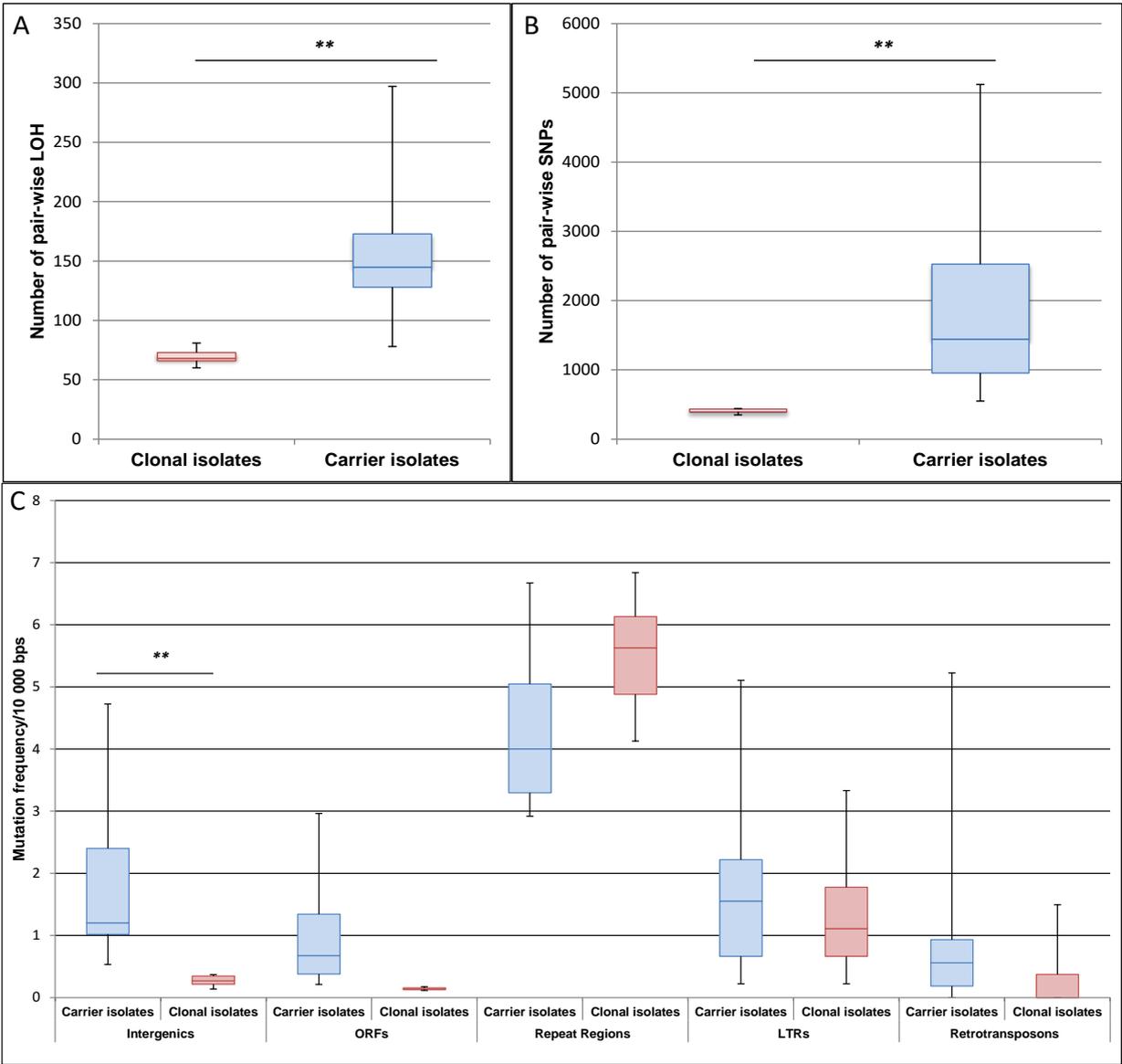
627

628 **Figure 5. Distribution of the LOH event by MinS classes and the number of SNPs**
 629 **differences involved by event for the 9 pair-wise carrier isolates comparisons from**
 630 **individual sample A, D and G.** The x-axis corresponds to the classes of LOH size (MinS in
 631 bp). The y-axis is in log scale and corresponds to the number of SNPs differences involved by
 632 LOH event.



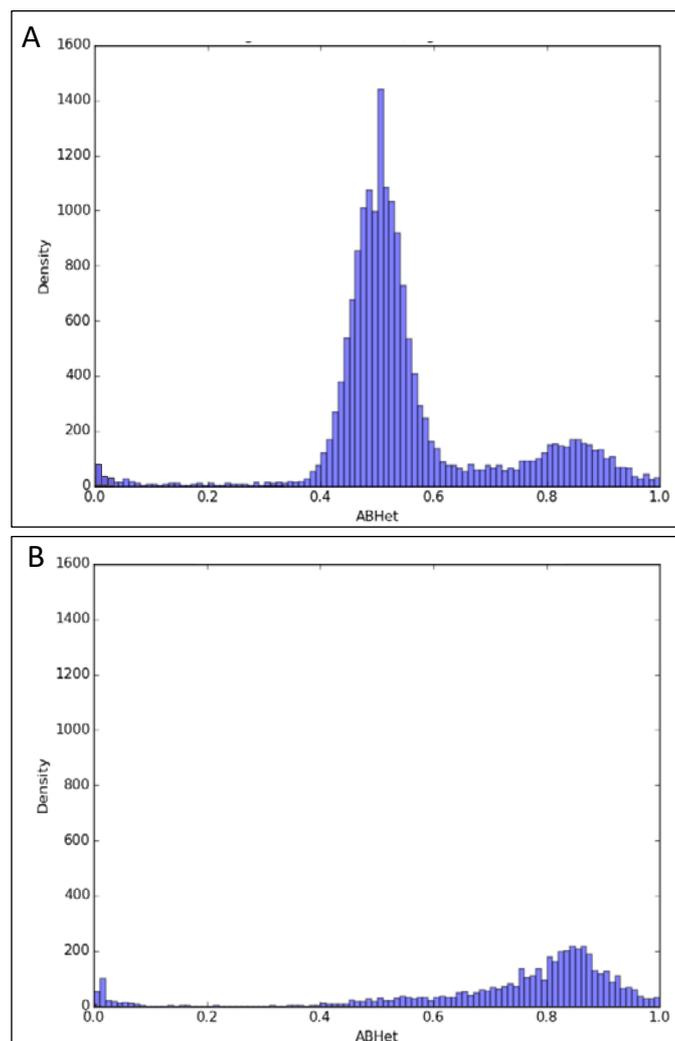
633

634 **Figure 6. Statistical analysis of the genomic diversity between carrier and clonal isolates.**
635 **A. Comparison of the number of SNPs between genomes from clonal and carrier**
636 **isolates.** The number of SNPs detected between the genomes from carrier isolates was
637 significantly higher than thus observed between the genome from clonal isolates (One way
638 ANOVA test, $p=0.007$) **B. Comparison of the number of LOH events observed from**
639 **genomes of clonal and carrier isolates.** The number of LOH events observed between the
640 genomes from carrier isolates was significantly higher than thus observed between the
641 genome from clonal isolates (One way ANOVA test, $p= 0.0013$). **C. Mutation frequency**
642 **comparisons for carrier and clonal isolates by genomic regions.** The mutation frequencies
643 were represented for 10.000 bps of the different genomic regions. Repeat regions were
644 significantly more mutated than other features whatever group considered (carrier or clonal
645 isolates) ($p < 0.01^{**}$; Post-hoc test: Tukey HSD). Carrier isolates were significantly more
646 variable in intergenic regions than clonal isolates ($p < 0.01^{**}$, Post-hoc test: Tukey HSD).



647

648 **Figure 7. Comparison of the allelic ratio distribution of all polymorphic positions**
649 **involved in LOH event. A. Comparison from carrier isolates. B. Comparison from clonal**
650 **isolates.** The allelic ratio ($ABHet = \text{number of reads for non-reference allele} / \text{total number of}$
651 reads) was determined with GATK tools. For each heterozygous SNP differences involved in
652 LOH events between pair of genomes the ABHet ratio at the position of the 3 isolates was
653 plot on the histogram. For carrier isolates (A), we observed a bimodal distribution with the
654 majority achieving a Gaussian distribution centered on a ABHet value of 0.5, which is the
655 expected value for diploid genome, and a minority achieving a Gaussian distribution centred
656 on a ABHet value of 0.85, which is the limit cut-off for the heterozygous SNPs definition in
657 our pipeline. This last distribution (around 0.85) was the only one found for clonal isolates
658 (B).



659

660 **Supplementary Tables S1, S2, S3 and S4 and Figures S1 and S2**

661

662 **Within-Host Genomic Diversity of *Candida albicans* in Healthy Carriers:**

663 **Implication for Microevolution Studies**

664

665 Emilie Sitterlé^{1,2,3}, Corinne Maufrais⁴, Natacha Sertour¹, Matthieu Palayret⁵, Christophe

666 d'Enfert¹, Marie-Elisabeth Bougnoux^{1,3*}

667

668 **Table S1: Principal characteristic of the *C. albicans* isolates obtained from the oral**
 669 **primo culture of the 10 healthy carriers.**

Individual	Age	Sex	Isolate		MLST								DST	Clade
			Number of <i>C. albicans</i> isolate in primo culture	Name of isolate	ACC1	AAT1	ADP1	MPIb	SAY1	VPS13	ZWF1b			
A	21	M	8	A1-8	3	2	5	2	2	6	5		66	1
B	21	M	2	B1-2	3	2	5	2	2	24	5		417	1
C	20	M	1	C1	14	43	6	4	7	3	8		1771	4
D	22	M	16	D1-8	4	36	4	94	4	26	4		1765	2
E	21	M	8	E1-8	3	8	6	4	7	3	35		1766	4
F	20	M	19	F1-8	3	3	10	3	26	39	95		1767	9
G	21	M	10	G1	7	13	15	60	34	37	12	ATTTCATTA	1768	3
				G2-8	7	13	15	60	34	37	15	MYTYCCAYTA	1769	3
H	22	M	4	H1-4	5	2	5	2	2	6	5		69	1
I	22	M	1	I1	5	2	5	2	2	6	5		69	1
J	20	F	1	J1	7	4	25	41	160	118	6		1770	3

670

671

672

Table S2. MLST and sequencing summary data from carrier or clonal isolates.

		MLST		Illumina	
		Clade	DST	Reads	Sequencing depth (mean)
Single oral sample from	Carrier isolate				
Individual A	A1	1	66	8 709 970	123
	A4	1	66	12 498 872	176
	A8	1	66	7 973 129	112
Individual D	D2	2	1765	12 430 900	172
	D4	2	1765	10 394 399	144
	D6	2	1765	10 326 576	144
Individual G	G1	3	1768	11 762 371	187
	G2	3	1769	8 684 792	110
	G3	3	1769	14 439 056	229
Strain	Clonal isolate				
X	X1	4	2281	8 954 225	98
	X2	4	2281	8 048 372	92
	X3	4	2281	7 396 486	86
Y	Y1	1	1047	4 600 098	58
	Y2	1	1047	10 436 389	130
	Y3	1	1047	8 913 126	110
Z	Z1	1	1400	11 408 364	145
	Z2	1	1400	7 740 394	97
	Z3	1	1400	8 232 147	103

673

674

675 **Table S3: Cartography of coding regions impacted by LOH events on chromosome 1**
676 **for the 3 genomes analysed for the individual G.** Grey boxes indicate heterozygous regions
677 and white boxes homozygous regions.

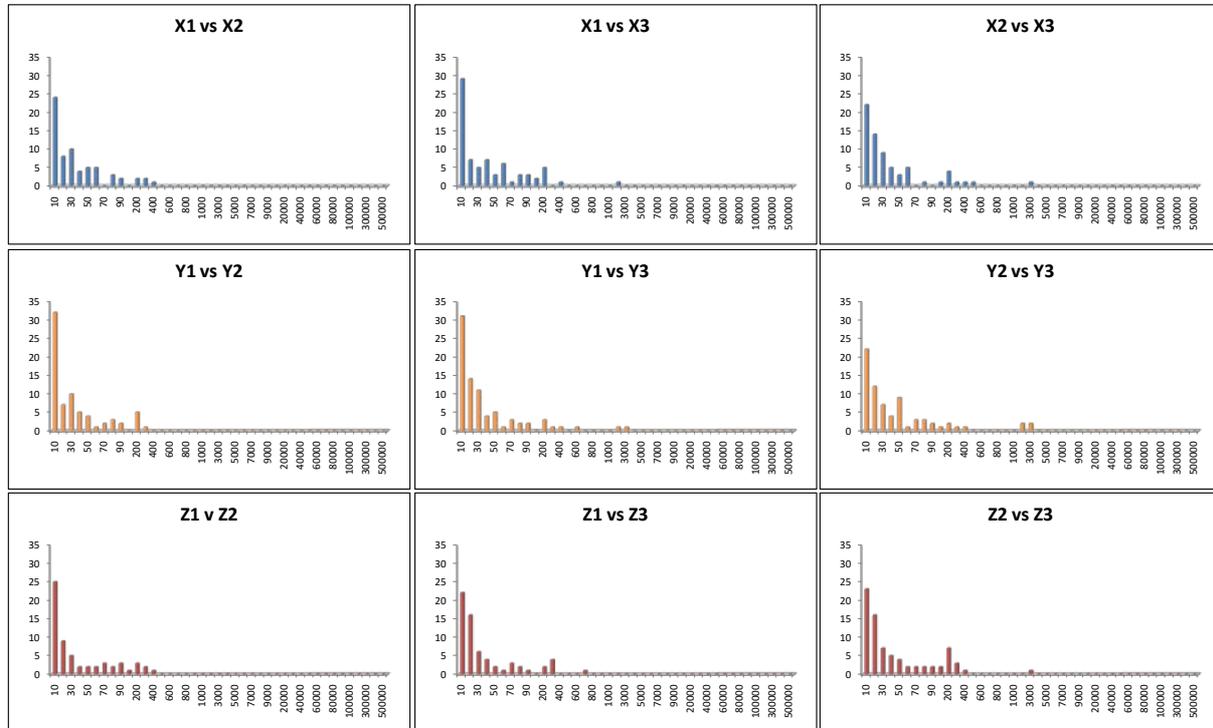
Chromosome 1								
LOH position start	LOH position end	LOH size (MinS in bp)	ORF name	G1	G2	G3	Confirmation by sanger sequencing	Number of LOH positions analyzed on the fragment
207292	209765	2473	C1_01050C_A MBF1 C1_01070C_A				Yes	8
296151	297134	983	C1_01500W_A				not tested	
723 588	728071	4483	C1_03450C_A C1_03460C_A				yes	3
867415	867976	561	IFD6				not tested	
925378	926429	1051	TUS1				Yes	6
955818	957058	1240	C1_04630C_A				Yes	4
982992	984515	1523	LYS21				not tested	
1021162	1021631	469	C1_04920C_A				not tested	
1053374	1055986	2612	ARG2 PSF2 RRP6				NO	9
1109047	1110523	1476	C1_05280W_A				Yes	4
1203093	1204760	1667	C1_05750C_A PGA26				Yes	6
1255766	1255989	223	C1_05970W_A				not tested	
1256172	1256412	240	C1_05970W_A				not tested	
1278773	1280984	2211	MSI3				Yes	5
1658913	1659854	941	C1_07650W_A				not tested	
1680887	1681713	826	C1_07730W_A				not tested	
1684233	1685604	1371	C1_07730W_A					
1685983	1686115	132	C1_07730W_A					
1689714	1690340	626	C1_07730W_A					
1703162	1706234	3072	C1_07810C_A C1_07820W_A				Yes	7
1736705	1737306	601	C1_07980C_A				not tested	
1791007	1791449	442	CAT8				Yes	5
1796848	1800880	4032	MLT1				not tested	
1925226	1926447	1221	AFG1 C1_08860C_A				Yes	8
1931633	1934096	2463	HEM14				Yes	5
1942700	1949589	6889	PFK26				not tested	
1951516	1956207	4691	PFK26 C1_08970W_A ZWF1				Yes	4
1958916	1960411	1495	KEX2 C1_09000W_A				not tested	
1972754	1973040	286	C1_09060C_A				not tested	
2170989	2172468	1479	VPS41				not tested	3
2174039	2174769	730	C1_09940W_A C1_09950C_A				Yes	10
2259086	2262213	3127	C1_10230C_A				Yes	4
2272828	2273967	1139	GCA1				not tested	
2425525	2426803	1278	RTG3 C1_11000C_A				not tested	
2475466	2477757	2291	C1_11270W_A C1_11280W_A				Yes	4
2477830	2481511	3681	C1_11290W_A PSP1				Yes	4
2513020	2513962	942	SAM2				not tested	
2515127	2519573	4446	C1_11460W_A DPP1				Yes	5
2542170	2544832	2662	PLD1				Yes	7

678

Table S4: Primers used to confirm LOH events.

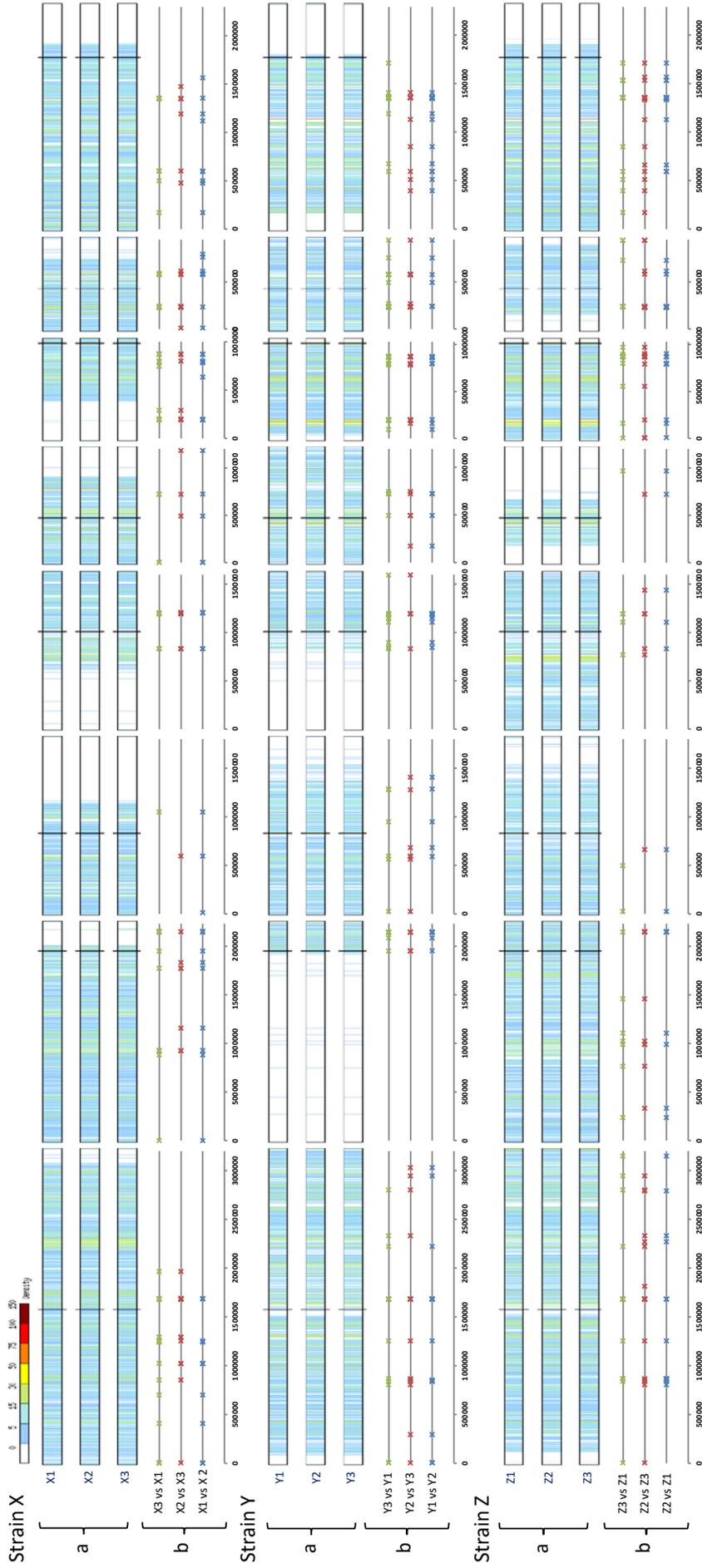
Primer name	Sequence
Ca22chr1A_208432-209275_F	AACTGCCAACACCAAGTC
Ca22chr1A_208432-209275_R	CGGTAGAGCAACAACAAT
Ca22chr1A_727151-727560_F	TGTTGGCGGGAAAACATT
Ca22chr1A_727151-727560_R	CCAGCATGTGCCTCTTCA
Ca22chr1A_925531-926239_F	TGCTTTGGACCTTTCGGC
Ca22chr1A_925531-926239_R	AAAAAACTACCGAACTCC
Ca22chr1A_955982-956386_F	CAAATCCTTCTCTTTTG
Ca22chr1A_955982-956386_R	TATTATCGTTGGTATTGG
Ca22chr1A_1019822-1020221_F	GATACCAAATCATGTGAA
Ca22chr1A_1019822-1020221_R	TGATTCAAGTTATTCTGA
Ca22chr1A_1054078-1054870_F	TTGTCACGTATCTATTCT
Ca22chr1A_1054078-1054870_R	TAGACTACTACTATCTAG
Ca22chr1A_1109780-1110453_F	CTCTGGAGTAGGGTCAAT
Ca22chr1A_1109780-1110453_R	CTCAATTGATTACCAAC
Ca22chr1A_1203706-1204259_F	CCCTCATTATTGATTAGT
Ca22chr1A_1203706-1204259_R	TACCAAGCAGTCTTTCCC
Ca22chr1A_1280077-1280540_F	ACATCAACAAAGGCAACT
Ca22chr1A_1280077-1280540_R	GGTGAAAGTGCTAAGAAT
Ca22chr1A_1705477-1706195_F	CTGGTGTTATGAGTGGAT
Ca22chr1A_1705477-1706195_R	TTAAAAATCTTGGGAAGG
Ca22chr1A_1791150-1791539_F	TTGACGACGAGGTTGATG
Ca22chr1A_1791150-1791539_R	AACATTCATAAGAAAGCG
Ca22chr1A_1926025-1926508_F	AACTTGCATTGTGAGTTT
Ca22chr1A_1926025-1926508_R	CAGAGTTTTTGAATAAACTG
Ca22chr1A_1931660-1932134_F	TTCTTCAGGGGAAATCAT
Ca22chr1A_1931660-1932134_R	TAAGGTTTCCATTCCATC
Ca22chr1A_ZWF1_F	GTTTCATTTGATCCTGAAGC
Ca22chr1A_ZWF1_R	GCCATTGATAAGTACCTGGAT
Ca22chr1A_2171026-2171696_F	CACACACACCCACACACA
Ca22chr1A_2171026-2171696_R	TACTGGTAGCGTTTTTAA
Ca22chr1A_2174288-2174676_F	ATCCAATAATAATCGTC
Ca22chr1A_2174288-2174676_R	AATTGAACCAGATGAACC
Ca22chr1A_2259122-2259818_F	AACATACATATACAGATC
Ca22chr1A_2259122-2259818_R	GTA CTCTTTGGCAAAGGG
Ca22chr1A_2321844-2322256_F	GTTGAAATATCTTTCACC
Ca22chr1A_2321844-2322256_R	CGAAAATATTTTTTCAGC
Ca22chr1A_2476789-2477561_F	CAATATAACATATGGTATTCC
Ca22chr1A_2476789-2477561_R	TGAAAGAGAATTGACCCT
Ca22chr1A_2478121-2478546_F	TCATTATGGATCATTGTC
Ca22chr1A_2478121-2478546_R	TCCATTAGCACAATGTAT
Ca22chr1A_2517400-2518253_F	CTACCAAAAACAAATAATCTC
Ca22chr1A_2517400-2518253_R	CTATTCCACCAAATCCTC
Ca22chr1A_2542450-2542980_F	TCGTGACGACTCATCACA
Ca22chr1A_2542450-2542980_R	ATTGCAGGAGCACCAGTG

681 **Figure S1. Distribution of MinS LOH events from the 9 pair-wise clonal isolates**
 682 **comparisons from strains X, Y and Z. The x-axis corresponds to the class of LOH size**
 683 **(MinS in bp). The y-axis corresponds to the number of events observed by MinS classes.**



684

685 **Figure S2. Representation of LOH events between genomes from the clonal isolates**
686 **selected from the 3 strains (X, Y and Z). Panels a: Detection of large LOH event by**
687 **chromosome.** For each genome from clonal isolates, heterozygous SNPs density was mapped
688 on the 8 chromosomes (1 Kb sliding windows). Homozygous regions are indicated in light or
689 white colour. No appearance of large LOH event was observed between genome comparisons
690 from clonal isolates by strain. **Panels b: Density of LOH events by chromosome.** For each
691 pair of genome comparison the starting location of all LOH events was mapped on the 8
692 chromosomes.



1 **A Useful Tool to Highlight New Resistance Mutations in Genes Involved in**
2 ***Candida albicans* Resistance to Azoles and Echinocandines**

3
4 *Emilie Sitterlé*^{1,2,3}, *Alix Coste*⁴, *Thomas Obadia*^{5,6}, *Corinne Maufrais*⁵, *Murielle Chauvel*¹,
5 *Natacha Sertour*¹, *Anne Puel*^{7,8}, *Dominique Sanglard*⁴, *Christophe d'Enfert*¹, *Marie-*
6 *Elisabeth Bougnoux*^{1,3}

7
8 ¹Fungal Biology and Pathogenicity Unit, Department of Mycology, Institut Pasteur, INRA,
9 Paris, France

10 ²Université Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, France

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

13 ⁴Institut de Microbiology, University of Lausanne and University Hospital Center, Lausanne,
14 Switzerland

15 ⁵Institut Pasteur - Bioinformatics and Biostatistics Hub - C3BI, USR 3756 IP CNRS, Paris,
16 France

17 ⁶Malaria Parasites & Hosts Unit, Department of Parasites & Insect Vectors, Institut Pasteur,
18 Paris, France

19 ⁷Laboratory of Human Genetics of Infectious Diseases, Necker Branch INSERM U1163
20 Paris, France

21 ⁸Paris Descartes University, Imagine Institute Paris, France

22

23

24

25

1 *Corresponding author:

2 Marie-Elisabeth Bournoux, MD, PhD

3 Institut Pasteur, Unité Biologie et Pathogénicité Fongiques, Département Mycologie, 25 rue

4 du Docteur Roux, F-75015 Paris, France ; Phone : +33 (1) 40 61 31 26 ; E-mail :

5 bournoux@pasteur.fr

6

1 **Abstract**

2 **Background:** Genome of *Candida albicans* is highly variable inducing a high level of gene
3 polymorphism. Point mutations in genes involved in resistance to antifungal may either
4 confer phenotypic resistance or be natural polymorphisms.

5 **Objectives:** Establish a comprehensive repertoire of natural polymorphisms in genes involved
6 in azoles and echinocandins resistance to help identify resistance mutations in clinical
7 isolates.

8 **Methods:** We took advantage of genome sequencing data from 151 unrelated *C. albicans*
9 strains susceptible to antifungal agents. We analyzed the sequences of genes involved in
10 resistance to azoles (*ERG11*, *TAC1*, *MRR1*, *UPC2*) and echinocandins (*FKSI*) and compared
11 them to the reference sequences of SC5314 strain. We tested then the capacity of the
12 repertoire to highlight resistance mutations in 10 resistant isolates. Selected putative new
13 resistant mutations were analyzed by transcript level measurements of *CDR2* and *ERG11* with
14 RT-qPCR and by mutagenesis experiments or mutational prediction effect.

15 **Results:** The repertoire included 126 amino-acid substitutions. The final rate of new
16 mutations appearance per strain was less than one. Using this repertoire, we identified from
17 the 10 resistant strains, 22 additional substitutions. Among these, 10 had already been
18 associated to resistance. The remaining 12 substitutions affected Tac1p (n=6/7), Erg11p
19 (n=4/11) and Upc2p (n=2/3). Eight were tested by mutagenesis (6 in Tac1p and 2 in Upc2p).
20 Four substitutions in Tac1p (H263Y, A790V, H839Y and P971S) conferred a high fold
21 change increase of the fluconazole MIC. Additionally, 2 substitutions in Erg11p (Y64H,
22 P236S) were predicted as deleterious.

23 **Conclusions:** This repertoire is a powerful tool helping to rapidly unveil mutations linked to
24 antifungal resistance.

25

1 **Introduction**

2 *Candida albicans* is the most frequent human fungal pathogen, causing a variety of skin and
3 soft tissue infections in healthy people. In patients with deficient immune systems, it can
4 behave as an aggressive pathogen responsible for disseminated diseases with mortality rate is
5 as high as 40%, even when patients receive antifungal therapy (Kullberg and Arendrup, 2015;
6 Noble and Johnson, 2007). Azoles and echinocandins are the 2 main classes of antifungal
7 agents used to treat candidiasis (Cornely et al., 2012; Pappas et al., 2016). However drug
8 resistance is emerging in patients with infection leading to the risk of treatment failure (Al-
9 Dorzi et al., 2018; Fisher et al., 2018; Lackner et al., 2014; Pfaller et al., 2013; Whaley et al.,
10 2017). Azole resistance can occur through various mechanisms such as alterations of *ERG11*
11 (encoding drug target lanosterol 14 α -demethylase) by overexpression of multidrug transporter
12 gene *CDR1*, *CDR2*, or *MDR1* or by overexpression of the *ERG11* (Asai et al., 1999; Coste et
13 al., 2004; Oliver et al., 2007). The expression of these genes is controlled by transcription
14 factors (TFs) *TAC1*, *MRR1* and *UPC2* involved respectively in the control of efflux pumps
15 *CDR1/2*, *MDR1* and the azole target *ERG11*. Several gain-of-function (GOF) mutations were
16 described in hyperactive alleles of these TFs, resulting in the overexpression of target genes
17 (Lohberger et al., 2014). In addition, echinocandin resistance can occur through point
18 mutations within hot spot (HS) regions of *FKS1* gene encoding for the β -(1,3)-D-glucan
19 synthase complex (Lackner et al., 2014; Morio et al., 2017).

20 For the two classes of antifungal agents the genetic basis of these alterations results mainly
21 from point mutations (Coste et al., 2006; Dunkel et al., 2008; MacPherson et al., 2005;
22 Marichal et al., 1999). However azole resistance can also arise from large-scale genetic
23 changes through genome rearrangements, including loss of heterozygosity (LOH) events
24 (Coste et al., 2006; Dunkel and Morschhäuser, 2011) and copy number variation (CNV) of
25 relevant genes by segmental or whole chromosomes aneuploidies (Coste et al., 2006;

1 Selmecki et al., 2010, 2008, 2009). The generation of such variants is facilitated by the high
2 genomic plasticity of this yeast, especially under stress conditions (Bennett et al., 2014). In
3 addition to the occurrence of these events, the diploid genome of *C. albicans* displays a high
4 level of natural polymorphisms (Ropars et al., 2018). Analysis of the genome-wide variation
5 of 182 *C. albicans* isolates, conducted by Ropars *et al.*, revealed they contain on average 1
6 heterozygous SNP every 204 bp (Ropars et al., 2018). Furthermore, when strains are
7 compared to SC5314, they displayed homozygous variant every 235 to 1404 bp depending on
8 the genetic cluster from which they originate (Hirakawa et al., 2015). In this context, point
9 mutations found in genes involved in resistance mechanisms of *C. albicans* to antifungal
10 drugs may either confer phenotypic resistance or be natural polymorphisms yielding their
11 interpretation difficult.

12 In this study, we took advantage of whole genome sequencing data from a large panel of
13 unrelated epidemiologically susceptible *C. albicans* clinical strains to establish a
14 comprehensive repertoire of the non-synonymous natural polymorphisms in main genes
15 involved in resistance to azoles (*ERG11*, *TAC1*, *MRR1* and *UPC2*) and echinocandins
16 (*FKSI*). This repertoire was subsequently used to screen resistant clinical isolates for
17 mutations linked to antifungal resistance. It confirmed its efficiency for detecting mutations
18 involved in phenotypic resistance.

19

1 **Results**

2 **Constitution of comprehensive natural polymorphism repertoire in gene involved** 3 **in resistance to azoles and echinocandines**

4 **Characteristics of the susceptible *C. albicans* clinical isolates**

5 A total of 151 epidemiologically unrelated clinical strains representative to the genetic
6 diversity of *C. albicans* were used to establish the comprehensive repertoire of natural
7 polymorphisms. Characteristics of strains, MLST and MICs to fluconazole and caspofungin
8 are summarized in the supplementary **Table S1**. The MICs distributions of the 151 strains are
9 shown in **Figure 1**. The fluconazole MICs ranged from 0.023 to 0.75 $\mu\text{g/ml}$ and caspofungin
10 from 0.008 to 0.38 $\mu\text{g/ml}$ (**Table S1**).

11 **Gene analysis and constitution of the repertoire of amino-acid substitutions**

12 We took advantage of whole genome sequences of the 151 susceptible *C. albicans* strains to
13 extensively analyze the polymorphism of genes involved in resistance to azoles and
14 echinocandines. Sequencing data characteristics for the 151 susceptible *C. albicans* strains as
15 well as SC5314 are summarized in the supplementary **Table S1**. For each non synonymous
16 polymorphism position, compared to SC5314 reference strain, the genotype obtained for the
17 different clinical strains are provide in **Table S3 and resumed in Table 1**. Overall, 126
18 distinct mutations conferring amino-acid changes were identified from the 151 strains and
19 constitute the repertoire of natural polymorphisms (**Table 1**). These substitutions were
20 distributed across all 5 genes and are shown in **Figure 2**. In all, there were 38, 20, 15, 33 and
21 20 mutations respectively in *Tac1p*, *Erg11p*, *Upc2p*, *Mrr1p* and *Fks1p*.

22 Among these substitutions, 41% (52/126) were found both at the heterozygous and
23 homozygous state while 54% (69/126) were found only in heterozygous state and 4% (5/126)
24 in homozygous state. Homozygous and heterozygous substitution frequencies were
25 determined for each position and are shown in **Table 1**. A low frequency of the substitution

1 was observed in 38% of cases for the heterozygous mutations (48/126) and 12% (15/126) for
2 the homozygous mutations (**Table 1**). Of note, a frequency of 0.65% corresponds to one
3 strain.

4 **Assessment of the natural genetic polymorphism diversity**

5 The polymorphism diversity representativity of the set of 151 susceptible strains is illustrated
6 in the **Figure 3** by the mean number of new mutations that could be found for each new strain
7 explored. The plateau trend achieved after ~20 strains, with a rate of new mutations less than
8 1 per new strain, confirmed that natural genetic diversity is well captured by the described
9 repertoire.

10

11 **Capacity evaluation of the repertoire to detect resistance mutations**

12 **Characteristics of the resistant *C. albicans* clinical isolates**

13 A set of 10 strains exhibited *in vitro* resistance to caspofungin and/or fluconazole where used
14 to test the capacity of the repertoire to detect resistance mutations. The characteristics of these
15 strains and results of *in vitro* susceptibility testing are detailed in **Table 2** and in
16 supplementary **Table S2**. All strains were resistant to fluconazole with 7 harboring a high-
17 level resistance (MIC > 256 µg/ml) and one was also intermediate to caspofungin, according
18 to the CLSI breakpoints (M27-S4. Wayne, 2012).

19 **Polymorphisms analysis compared to the repertoire of natural amino acid** 20 **substitutions**

21 The whole genome sequences of the 10 resistant strains allowed us to obtain an average
22 sequencing depth of 123x. The sequencing data are shown in the supplementary **Table S2**.

23 The analysis of the 5 genes involved in resistance allowed the identification of 45 non-
24 synonymous mutations compared to the sequences of SC5314 reference strain (**Table S3**).

25 Using the repertoire described above (**Table 1**), 23 substitutions were identified as natural

1 polymorphisms. The remaining 22 amino-acid changes affecting Erg11p (n=11), Tac1p (n=7),
2 Upc2p (n=3) and Fks1p (n=1), are listed in the supplementary **Table S4**. In order to assess
3 whether these mutations correspond only to the natural polymorphisms that would be
4 observed by analyzing 10 additional susceptible strains, the number of new mutations arising
5 in the 10 resistant strains was compared to the distribution of the number of new mutations
6 arising in 10 susceptible strains. The introduction of the 10 resistant strains caused an increase
7 in cumulative mutations at the end of the curve compared to 10 susceptible strains (**Figure**
8 **4A**). This discrepancy is illustrated by **Figure 4B**, showing that the number of new mutations
9 arising in the 10 resistant strains was significantly different to the distribution of the number
10 of new mutations arising in 10 susceptible strains. This suggests the appearance of additional
11 mutations to those expected to be due to natural polymorphisms (**Figure 4B**; $p < 0.0001$).

12 Among the 22 amino-acid substitutions selected using the repertoire 10 have already been
13 associated to azoles or echinocandins resistance (**Table S4**). One in Tac1p (A736T) (Coste et
14 al., 2006), 7 in Erg11p (Y132H, R157K, S405F, G448E, G450E, R467I) (Chau et al., 2004;
15 Favre et al., 1999; Goldman et al., 2004; Kakeya et al., 2000; Lamb et al., 2000; Lee et al.,
16 2004; Löffler et al., 1997; Marichal et al., 1999; Perea et al., 2001; Sanglard et al., 1998;
17 Sanglard and Bille, 2002; Wang et al., 2009; White, 1997; Xu et al., 2008) and the recently
18 described T123I (Wu et al., 2017), 1 in Upc2p (G648D^h) (Flowers et al., 2012), and 1 in
19 Fks1p (R647G) (Lackner et al., 2014). The distribution by patient of these known mutations
20 as well as the 12 remaining mutations selected by our tool, are summarized in the **Table 2**.
21 Each of the 10 fluconazole-resistant strains harbored at least one known mutation that can
22 explain phenotypic resistance. However, the high level of resistance of some strains
23 (CEC4504 and CEC4503) cannot be explained by the unique presence of one heterozygous or
24 homozygous substitution in the Erg11p, suggesting the presence of additional mutations
25 involved in resistance.

1 This hypothesis was supported by a new comparison of the mutation accumulation curves of
2 the two sets of strains. We removed the 10 mutations already known to be involved in
3 resistance so that differences between susceptible and resistant strains is restricted only to the
4 newly-identified mutations (n=12). Again, the introduction of the 10 resistant strains causes
5 an increase in cumulative mutations (**Figures 4C**). The number of new mutations arising in
6 the 10 resistant strains was still significantly different to the distribution of the number of new
7 mutations arising in 10 susceptible strains (**Figure 4D**; $p = 0.0014$). We predicted that at least
8 4 out of the 12 new mutations were distinct from the occurrence of natural polymorphisms.
9 To confirm that some of these mutations were responsible for additional resistance phenotype,
10 quantification of genes expression involved in azoles resistance, mutagenesis experiments or
11 mutation effect prediction were performed.

12

13 **Analysis of putative new resistance mutations**

14 **Overexpression level measurement of *CDR2* and *ERG11***

15 The strains harboring putative new resistant mutations in transcription factors *TAC1* and
16 *UPC2* were analyzed by transcript level measurements by quantitative RT-PCR of *CDR2* and
17 *ERG11* respectively. The results of changes in gene expression (n fold) normalized to *ACT1*
18 expression and relative to that of control strain (SC5314 reference strain for *ERG11* and
19 CEC718 resistant strain harboring the known A736T mutation for *CDR2*) are resumed in
20 **Table 3**. Three strains carrying new amino-acid changes in Upc2p showed an overexpression
21 level of *ERG11* (7.5, 4.5 and 5.9 fold increase on average for CEC718 (G657D), CEC4521
22 (D654N^h) and CEC4522 (D654N) strains respectively). We also observed an overexpression
23 level of *CDR2* for the 4 strains carrying the combination of the 6 putative new mutations in
24 *TAC1* (1.7, 12.1, 2.7 and 3.9 on average fold increase respectively for CEC723, CEC4504,
25 CEC4515 and CEC4527 strains compared to the CEC718 strain). They carried A790V or

1 different combinations of the H263Y, H839Y, C858Y, N823K and P971S amino-acid
2 changes (**Table 3**).

3 ***TAC1* and *UPC2* mutagenesis experiments**

4 Of the 12 putative new mutations described above, the 6 mutations in *TAC1* and the 2
5 mutations observed in *UPC2* have been tested by mutagenesis experiments to confirm their
6 role in resistance phenotypes. No effects in term of fluconazole MIC changes were observed
7 concerning the *UPC2* mutants compared to the *C. albicans* strain transformed with the *UPC2*
8 wild-type allele (**Figure 5B**). Introduction of 4 of the 6 mutations found in *TAC1* allele result
9 in high increase of the fluconazole MIC compared to the *C. albicans* strain transformed with
10 the *TAC1* wild-type allele (H263Y, A790V, H839Y, and P971S conferred a 8, 7.8, 85 and 5.3
11 fold changes increase respectively) (**Figure 5A**).

12 **Mutation effect prediction of the Erg11p**

13 We then performed mutational prediction effect by using prediction software to evaluate
14 effect of mutations in Erg11p using reference protein 5v5z ((Keniya et al., 2017).

15 The impact of the 3 new amino-acid changes found in Erg11p (P30S^h, Y64H, P236S) was
16 evaluated and compared with the results of the mutations already known to be involved in
17 resistance. PoPMuSiCv3.1, HoTMuSiCv0.5 software were used to predict $\Delta\Delta G$, ΔT_m and
18 solvent accessibility of amino-acids and SNPMuSiC to predict the overall effect of mutations.

19 The D9V^h substitutions could not be evaluated due to its absence in the crystallized protein.

20 The P30S^h substitution was predicted with a neutral impact by SNPMuSiC and with a
21 putative stabilizing effect on the protein ($\Delta\Delta G = -0.27$ and $\Delta T_m = 0.35$) with Hot and
22 PoPMuSiC. These results were not in favor of a mutation involved in resistance. However,
23 the 2 other mutations (Y64H and P236S) showed a low level of amino-acid accessibility <
24 20% (5.85 and 17.94) and an assumed destabilizing effect on protein ($\Delta\Delta G = 2.03$ and 1.41;
25 $\Delta T_m = -2.48$ and -1.6 respectively for Y64H, P236S). These results were corroborated by

1 SNPMuSic software that predicted a deleterious effect for both mutations.
2 For comparison, the mean values obtained for the substitutions already known to be involved
3 in resistance and found in this study (T123I, Y132H, R157K, S405F, G448E, G450E, R467I)
4 were $\Delta\Delta G=1.0$ and $\Delta T_m = -1.3$ with an average amino acid accessibility level of 14.37%.

5

6 **Discussion**

7 Antifungal drug resistance in *Candida* sp. is a growing problem in patients with infections
8 that increase the risk of treatment failure (Al-Dorzi et al., 2018; Lackner et al., 2014; Pfaller et
9 al., 2013; Whaley et al., 2017). However, despite the high number of genomic data obtained
10 in recent years, concerning the mechanisms of resistance and the many efforts to elucidate the
11 links between phenotype and genotype, an accurate and interpretative classification of the
12 effects of mutations involved in the phenotypic resistance remains challenging. Non-
13 synonymous mutations that modify both codon and the associated amino-acid may alter the
14 biophysical properties of the encoded protein, such as its thermodynamic stability or
15 solubility, or its functional properties by affecting the active site or binding to ligands and
16 proteins partner (Ancien et al., 2018). Nevertheless, the resistance to antifungals drugs is
17 complex due to the heterogeneity of the molecular mechanisms that may underlie the
18 phenotype (Ancien et al., 2018). Additionally, the genome of *C. albicans* displays a high level
19 of gene polymorphisms. Point mutations in genes involved in resistance to antifungal drugs
20 may either confer phenotypic resistance or be natural polymorphisms yielding their
21 interpretation more difficult (Morio et al., 2010).

22 To facilitate the first step of mutation evaluation we have established, in this study, a
23 comprehensive repertoire of natural polymorphism observed on the 5 main genes involved in
24 azoles (*ERG11*, *TAC1*, *UPC2* and *MRR1*) and echinocandins (*FKSI*) resistance. The
25 originality of this work was the use of a large panel of susceptible strains of *C. albicans*

1 (n=151) selected to be representative of the diversity of *C. albicans*. These strains were
2 distributed in 15 different clades and have been isolated in different infectious contexts
3 (invasive or superficial) or during commensalism but also in many geographical areas. In
4 addition, the strains have a distribution of MICs to fluconazole and caspofungin comparable
5 to that observed for the establishment of Ecoff. We therefore assume that this collection is
6 representative of the diversity of *C. albicans* and have used it to constitute comprehensive
7 natural polymorphism repertoire. We were only interested in the non-synonymous mutations
8 that we expressed in amino-acid changes. This repertoire contains a total of 126 distinct
9 amino-acid substitutions over the 5 genes, enabling it to be representative, in these genes, of
10 the genetic diversity of *C. albicans*. Thanks to our collection, the probability of finding a new
11 mutation of natural polymorphism, when analyzing a new strain, is less than 1 (0.33 on
12 average).

13 In order to test the capacity of our tool to highlight resistance mutation, we tested it for the
14 analysis of 10 phenotypically resistant strains. Using the repertoire as a guide, 23 mutations
15 could be ruled out from the analysis as identified as natural polymorphisms. Concerning the
16 22 mutations selected by our tool, 10 were already been described as involved in resistance
17 and 12 were news. Although the two strains harboring Upc2p substitutions showed an
18 increase in *ERG11* expression by RT-qPCR, the mutagenesis experiment did not confirm the
19 involvement of these mutations in the observed phenotype, indicating that the overexpression
20 of *ERG11* was UPC2-independent. These strains also harbored additional resistance
21 mechanisms which likely participated in their resistant phenotype. These finding have already
22 been observed in some strains of *C. albicans* (Berkow et al., 2015; Heilmann et al., 2010;
23 Jensen et al., 2015).

24 However, as predicted by our tool, 4 out of selected mutations found in Tac1p (H263Y,
25 A790V, H839Y, and P971S) have been confirmed by biological evidence to increase

1 fluconazole MIC, confirming the major interest of this tool to highlight new resistant
2 mutations. Additionally, 2 out of them in Erg11p (Y64H, P236S) were predicted with high
3 confidence as deleterious effect on the protein structure. Interestingly, these positions have
4 been already predicted to be vital amino acid for various interactions in the active site of
5 *C.albicans* Erg11p (Debnath and Addya, 2014; Hargrove et al., 2017; Sagatova et al., 2015),
6 suggesting the likely role of these substitutions in increasing azole resistance in our strains. It
7 is evident that significant geometric changes in the Erg11p active site domain will occur if the
8 amino acids responsible for key molecular interactions (with its substrate or inhibitors) are
9 substituted (Debnath and Addya, 2014). However, the specific contributions of individual
10 *ERG11* mutations to azole resistance in *C. albicans* have not been widely explored and very
11 few mutations were demonstrated to be directly responsible for the resistance (Flowers et al.,
12 2015). The impact of a non-synonymous mutation on azole resistance can be demonstrated by
13 heterologous expression of the *ERG11* allele in different model, such as *Saccharomyces*
14 *cerevisiae* (Lamb et al., 2000; Sanglard et al., 1998) or, recently described, in genetically
15 engineered strain of *Candida lusitanae*, conferring numerous advantages compared to the
16 first one (Accoceberry et al., 2017). This system could be used to definitively confirm the
17 involvement of these mutations in azole resistance.

18 To conclude, our study highlights the high diversity and frequency of amino acid substitutions
19 in the different genes involved in antifungal resistance. This genetic polymorphism renders
20 difficult to interpret mutations, which is considerably facilitated by the availability of our
21 repertoire, thus helping to rapidly unveil mutations potentially linked to antifungal resistance.

22

1 **Material and methods**

2 **Strains collection**

3 Two sets of *C. albicans* clinical strains were used in this study and are listed in supplementary
4 **Table S1 and S2**. The first one consists of 151 epidemiologically-unrelated strains
5 susceptible to antifungal agents selected to represent the natural population diversity of *C.*
6 *albicans* (Ropars et al., 2018). These strains constitute a collection of isolates collected from
7 various geographical origins (France, Niger, Morocco, Brazil, Belgium, Senegal, United
8 Kingdom, Guiana and South Korea). The majority of them were collected in the course of
9 infections (20% invasive 29/151; 37% superficial 56/151) while 29% of them (44/151) have
10 been isolated in the course of commensalism and 14 % (22/151) were not classified. They
11 were distributed among 128 different DSTs representing 15 different clades.

12 The second set of strains consists of 10 isolates resistant *in vitro* to caspofungine and /or
13 fluconazole. The *C. albicans* laboratory reference strain SC5314 was also used in this study
14 (Gillum et al., 1984).

15

16 **Culture media**

17 *C. albicans* isolates were obtained by samples cultured on chromogenic plate CHROMagar™
18 *Candida* (BioMérieux, Marcy-l'Étoile, France) for 5 days at 37°C. Colony morphology,
19 growth color and Mass spectrometry type MALDI-TOF was used for species identification.
20 The isolates were then stored in glycerol 30% and frozen at -80°C. For further analysis,
21 isolates were sub cultured from frozen stock at 30°C in liquid or solid YPD media (1% yeast
22 extract, 2% peptone, 2% dextrose). Solid media were obtained by adding 2% of agar.
23 Nourseothricin (300 µg/ml) was added to YPD agar plates to select strains containing the
24 SAT1 cassette. *Escherichia coli* TOP10 (Life Technologies) or XL10-Gold cells (Agilent
25 Technologies) were used as a host for plasmid constructions and propagation. *E. coli* cells

1 were cultured on LB (Luria-Bertani broth) or LB plates, supplemented with ampicillin (100
2 µg/ml) or kanamycin (50 µg/ml) when required.

3

4 **MICs determination**

5 The fluconazole and caspofungin MICs of *C. albicans* isolates were determined by using the
6 Etest® method as specified by the manufacturer's and interpreted according to CLSI
7 guidelines (M27-S4. Wayne, 2012). Briefly, RPMI agar plates (bioMérieux, Marcy l'Etoile,
8 France) were inoculated with sterile swabs saturated with suspensions of the study isolates
9 (Overnight culture diluted to an OD600 of 0.02 in sterile water). For mutant strains, the RPMI
10 plates were previously supplemented with uridine at 20 mg/L. After the plates were allowed
11 to dry, Etest® strips (bioMérieux, Marcy l'Etoile, France) were applied. Plates were incubated
12 at 35°C for 48h, and pictures were taken at 24 and 48 hours.

13

14 **Multilocus Sequence Typing**

15 All strains were typed using Multi Locus Sequence Typing (MLST), as described previously
16 (Bougnoux et al., 2003). Briefly, the internal regions of seven housekeeping genes (*AAT1a*,
17 *ACC1*, *ADP1*, *MPIb*, *SYA1*, *VPS13*, and *ZWF1b*) were amplified in each isolate using the Taq
18 polymerase (Invitrogen) in an Eppendorf Mastercycler ep gradient thermal cycler. The
19 sequencing was performed on both strands at Eurofins sequencing facility by using the cycle
20 sequencing technology (dideoxy chain termination/cycle sequencing) on ABI 3730XL
21 sequencing machines (Applied Biosystems). Sequences analysis and SNPs detection were
22 performed using the SeqScape™ v3.0 software. Allele and diploid sequence type (DST)
23 assignments were determined using the *C. albicans* multilocus sequence typing database
24 (<http://pubmlst.org/calbicans>).

25

1 **Extraction and Illumina sequencing**

2 The fungal cells from overnight YPD culture were initially disrupted with glass beads (0.5
3 mm) in a FastPrep24 (MP Biomedicals, Illkirch, France) for 3 cycles of 1 min at 50 Hz with
4 cooling between cycles. DNA was then extracted using the QIAamp® DNA Mini Kit
5 (Qiagen, Courtaboeuf, France), according to the manufacturer's instruction. Libraries were
6 prepared using the NEXTflex™ PCR-Free DNA Sequencing kit (Illumina), according to the
7 manufacturer's recommendations. Sequencing was performed on a HiSeq2000 platform
8 generating 100 bp paired-ends reads (roughly 100x coverage).

9

10 **Genomes analysis**

11 Sequences were analyzed as described previously (Feri et al., 2016; Ropars et al., 2018) using
12 BWA alignment tool v0.7.7 (Li and Durbin, 2009) and the *C. albicans* reference strain
13 SC5314 (genome assembly 22 version A22-s06-m01-r01). The variant calling was performed
14 with GATK v3.1 (McKenna et al., 2010) according to the best practices recommended
15 annotation for hard filters. In each strain, the entire open reading frame sequences of 5 genes
16 involved in the azole (*ERG11*, *TAC1*, *MRR1* and *UPC2*) and echinocandins (*FKSI*) resistance
17 was compared to that of reference strain SC5314. Only non-synonymous mutations were
18 considered. At each position, homozygous and heterozygous substitution frequency was
19 calculated from the set of 151 *C. albicans* susceptible strains. Genotype results are listed in
20 the supplementary **Table S2**.

21

22 **RNA isolation and quantitative RT-PCR**

23 For some resistant isolates, gene expression level was measured by quantitative RT-PCR.
24 Fungal cells from YPD cultures in the mid-log exponential growth phase were collected
25 (5.6×10^7 cells) by centrifugation, resuspended with RLT buffer + β -mercapto ethanol

1 solution or with RNA later® (Ambion) if not used right away. The cells were then
2 mechanically disrupted with glass beads (0.5 mm) in a FastPrep24 (MP Biomedicals, Illkirch,
3 France) for 5 cycles of 1 min at 50 Hz with cooling between cycles.

4 The total RNA was extracted using the RNeasy® Mini Kit (Qiagen, Courtaboeuf, France)
5 according to the manufacturer's instructions and eluted in 40µL of RNA free water. The
6 RNA concentration was determined by NanoVue Plus (GE Healthcare) readings, while the
7 RNA integrity was determined by electrophoresis on 1.5% agarose gel. Contaminating
8 genomic DNA was removed by DNase treatment (DNase I recombinant, RNase-free, Roche,
9 France). The RNA was then diluted to obtain a final concentration of 0.025 µg/µl.

10 The expression level of *CDR2* and *ERG11* was quantitatively determined by a One-Step Real-
11 Time Reverse Transcription TaqMan® PCR (RT-qPCR) in a Mastercycler® ep realplex
12 (Eppendorf). Reactions and cycling conditions were done according to the manufacturer's
13 instructions with the Super Script™ III Platinum One-Step qRT-PCR kit (Invitrogen) using
14 2µl of diluted RNA and 1µl of RNaseOUT Recombinant Ribonuclease Inhibitor
15 (RNaseOUT™, Invitrogen). The housekeeping gene *ACT1* was used as an expression control.

16 Changes (n fold) in gene expression relative to that of control strain (SC5314 reference strain)
17 for *ERG11* and CEC718 resistant strain harboring the known A736T mutation for *CDR2*),
18 were determined from ACT1-normalized expression levels by the method of $2^{-\Delta\Delta C_t}$. The
19 primers and probes used for this assay are described in supplementary **Table S5**. For each
20 strain, two independent biological and three technical replicates were processed.

21

22 ***TAC1* and *UPC2* mutant constructions**

23 To prove the role of the new mutations on gene overexpression and phenotypic resistance to
24 fluconazole, mutants have been reconstructed for the 6 mutations found in *TAC1* and 2 found

1 in *UPC2*. The genotypes of constructed mutants are listed in the supplementary **Table S5**.
2 Primers used for strains and plasmids constructions are listed in supplemental **Table S6**.
3 DNA fragments (*TAC1* and *UPC2*) from SC5314 (wild type strain) and from clinical strains
4 containing mutations into *TAC1* gene (A790V from CEC723) and *UPC2* gene (G657D from
5 CEC718 and D654N from CEC4522) were amplified including the promotor and terminator
6 by PCR and cloned into TOPO-BLUNT cloning vector (Invitrogen). Sequence of complete
7 fragments was checked by Sanger sequencing. The resulting plasmids, harboring a wild type
8 or mutated fragment were digested by selected restriction enzymes and ligated in CIPsAT
9 plasmid. *E. coli* TOP10 (Invitrogen) competent cells were used for cloning and maintaining
10 the plasmid constructs.

11 The CIPsAT-*P_{TAC1}-TAC1* plasmid harboring the wild type *TAC1* allele was used as template
12 for PCR amplification to introduce other mutations by site-directed mutagenesis (H263Y,
13 N823K, H839Y, C858Y, P971S). Primers used were designed using the web-based
14 QuikChange Primer Program available online (www.agilent.com/genomics/qcpd). Reactions
15 and cycling conditions were done according to the manufacturer's instructions with the
16 QuikChange Lightning Site-Directed Mutagenesis kit (Agilent). *E. coli* XL10-Gold
17 (Agilent) Ultracompetent cells were used for plasmids transformation. Complete fragments
18 were checked by Sanger sequencing.

19

20 ***C. albicans* strains transformation**

21 The CIPsAT derivatives were integrated at the *RPS1* locus by transforming *C. albicans*
22 strains DSY4578 or DSY3984 (Supporting Information **Table S5**) using the lithium-acetate
23 transformation protocol as described previously (Walther and Wendland, 2003).
24 Transformants were selected on YPD medium supplemented with nourseothricin (300 µg/ml).

1 Plasmids integrations were checked by PCR with CipUL and CipUR primers (**Table S6**).
2 *TAC1* or *UPC2* ORFs were sequenced to confirm the existence of each mutation.

3

4 **Prediction mutation effect in Erg11p**

5 Putative novel mutations found in Erg11p were analyzed and compared with already
6 described mutations involved in resistance using software that predicts the effect of
7 mutations. The solvent accessibility of amino-acids within the protein Erg11p and the
8 prediction of protein thermodynamic stability and thermostability changes upon single-site
9 mutations ($\Delta\Delta G$ and ΔT_m) as well as the deleterious effect of mutations were evaluated. The
10 5v5z entry (Keniya et al., 2017) from the Protein Data Base (<https://www.rcsb.org/>) was used
11 to extract 3D structure information's of Erg11p. PoPMuSiCv3.1 (Dehouck et al., 2011).
12 HoTMuSiCv0.5 (Pucci et al., 2016) and SNPmuSiC (Ancien et al., 2018) softwares were
13 used to predict $\Delta\Delta G$, ΔT_m and solvent accessibility of amino-acids. The output mutation files
14 were generated by MuSiCv3.0r3 available online at <https://soft.dezyme.com/>. Amino-acids
15 accessibility classification ranged from 0 (buried) to 100 (fully accessible). The mutation
16 effect on protein stability was classified as destabilizing ($\Delta\Delta G > 0$ and $\Delta T_m < 0$) or stabilizing
17 ($\Delta\Delta G < 0$ and $\Delta T_m > 0$).

18

19 **Statistical analysis**

20 **Assessment of the natural genetic polymorphism diversity**

21 The genetic polymorphism diversity in susceptible strains was defined as the cumulative
22 number of distinct mutations that could be identified with an increasing number of strains.
23 The distribution of this genetic diversity was derived using a bootstrapping method where the
24 $N_S = 151$ susceptible strains were randomly permuted with $r = 10,000$ replicates. In each
25 replicate, the cumulative number of distinct mutations was counted as a function of the

1 increasing number of strains. Likewise, the rate of new mutations was derived from these
2 resampled strain sets to illustrate the representativity of the pool of susceptible strains from
3 our experiment.

4

5 **Contribution of resistant strains to genetic diversity**

6 To assess the contribution of the $N_R = 10$ resistant strains to genetic diversity, the above
7 method was adapted to replace the last N_R susceptible strains from the random permutations
8 with random permutations of the N_R resistant strains. This resulted in two sets of bootstrapped
9 distributions: the null with the N_S strains and the alternate with $N_S - N_R$ susceptible strains and
10 last N_R resistant strains. Finally, the number of new mutations arising in the last N_R strains
11 from the alternate distribution (that is the number of mutations from the pool of resistant
12 strains that cannot be identified in susceptible strains) was compared to the distribution of the
13 number of mutations in the null bootstrap using percentiles for credible intervals.

14

1 **References**

- 2 Accoceberry, I., Rougeron, A., Biteau, N., Chevrel, P., Fitton-Ouhabi, V., Noël, T., 2017. A CTG Clade
3 *Candida* Yeast Genetically Engineered for the Genotype-Phenotype Characterization of
4 Azole Antifungal Resistance in Human-Pathogenic Yeasts. *Antimicrob. Agents*
5 *Chemother.* 62. <https://doi.org/10.1128/AAC.01483-17>
- 6 Al-Dorzi, H.M., Sakkijha, H., Khan, R., Aldabbagh, T., Toledo, A., Ntinika, P., Al Johani, S.M., Arabi,
7 Y.M., 2018. Invasive Candidiasis in Critically Ill Patients: A Prospective Cohort Study in
8 Two Tertiary Care Centers. *J. Intensive Care Med.* 088506661876783.
9 <https://doi.org/10.1177/0885066618767835>
- 10 Ancien, F., Pucci, F., Godfroid, M., Rooman, M., 2018. Prediction and interpretation of deleterious
11 coding variants in terms of protein structural stability. *Sci. Rep.* 8.
12 <https://doi.org/10.1038/s41598-018-22531-2>
- 13 Asai, K., Tsuchimori, N., Okonogi, K., Perfect, J.R., Gotoh, O., Yoshida, Y., 1999. Formation of azole-
14 resistant *Candida albicans* by mutation of sterol 14-demethylase P450. *Antimicrob.*
15 *Agents Chemother.* 43, 1163–1169.
- 16 Bennett, R.J., Forche, A., Berman, J., 2014. Rapid mechanisms for generating genome diversity:
17 whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harb. Perspect.*
18 *Med.* 4. <https://doi.org/10.1101/cshperspect.a019604>
- 19 Berkow, E.L., Manigaba, K., Parker, J.E., Barker, K.S., Kelly, S.L., Rogers, P.D., 2015. Multidrug
20 Transporters and Alterations in Sterol Biosynthesis Contribute to Azole Antifungal
21 Resistance in *Candida parapsilosis*. *Antimicrob. Agents Chemother.* 59, 5942–5950.
22 <https://doi.org/10.1128/AAC.01358-15>
- 23 Bougnoux, M.-E., Tavanti, A., Bouchier, C., Gow, N.A.R., Magnier, A., Davidson, A.D., Maiden, M.C.J.,
24 d'Enfert, C., Odds, F.C., 2003. Collaborative Consensus for Optimized Multilocus Sequence
25 Typing of *Candida albicans*. *J. Clin. Microbiol.* 41, 5265–5266.
26 <https://doi.org/10.1128/JCM.41.11.5265-5266.2003>
- 27 Chau, A.S., Mendrick, C.A., Sabatelli, F.J., Loebenberg, D., McNicholas, P.M., 2004. Application of
28 Real-Time Quantitative PCR to Molecular Analysis of *Candida albicans* Strains Exhibiting
29 Reduced Susceptibility to Azoles. *Antimicrob. Agents Chemother.* 48, 2124–2131.
30 <https://doi.org/10.1128/AAC.48.6.2124-2131.2004>
- 31 Cornely, O.A., Bassetti, M., Calandra, T., Garbino, J., Kullberg, B.J., Lortholary, O., Meersseman, W.,
32 Akova, M., Arendrup, M.C., Arikan-Akdagli, S., Bille, J., Castagnola, E., Cuenca-Estrella, M.,
33 Donnelly, J.P., Groll, A.H., Herbrecht, R., Hope, W.W., Jensen, H.E., Lass-Flörl, C., Petrikos,
34 G., Richardson, M.D., Roilides, E., Verweij, P.E., Viscoli, C., Ullmann, A.J., ESCMID Fungal
35 Infection Study Group, 2012. ESCMID* guideline for the diagnosis and management of

1 Candida diseases 2012: non-neutropenic adult patients. Clin. Microbiol. Infect. Off. Publ.
2 Eur. Soc. Clin. Microbiol. Infect. Dis. 18 Suppl 7, 19–37. [https://doi.org/10.1111/1469-](https://doi.org/10.1111/1469-0691.12039)
3 0691.12039

4 Coste, A., Turner, V., Ischer, F., Morschhäuser, J., Forche, A., Selmecki, A., Berman, J., Bille, J.,
5 Sanglard, D., 2006. A mutation in Tac1p, a transcription factor regulating CDR1 and
6 CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal
7 resistance in *Candida albicans*. Genetics 172, 2139–2156.
8 <https://doi.org/10.1534/genetics.105.054767>

9 Coste, A.T., Karababa, M., Ischer, F., Bille, J., Sanglard, D., 2004. TAC1, transcriptional activator of
10 CDR genes, is a new transcription factor involved in the regulation of *Candida albicans*
11 ABC transporters CDR1 and CDR2. Eukaryot. Cell 3, 1639–1652.
12 <https://doi.org/10.1128/EC.3.6.1639-1652.2004>

13 Debnath, S., Addya, S., 2014. Structural basis for heterogeneous phenotype of ERG11 dependent
14 Azole resistance in *C. albicans* clinical isolates. SpringerPlus 3, 660.
15 <https://doi.org/10.1186/2193-1801-3-660>

16 Dehouck, Y., Kwasigroch, J., Gilis, D., Rooman, M., 2011. PoPMuSiC 2.1: a web server for the
17 estimation of protein stability changes upon mutation and sequence optimality. BMC
18 Bioinformatics 12, 151. <https://doi.org/10.1186/1471-2105-12-151>

19 Dunkel, N., Blaß, J., Rogers, P.D., Morschhäuser, J., 2008. Mutations in the multi-drug resistance
20 regulator *MRR1*, followed by loss of heterozygosity, are the main cause of *MDR1*
21 overexpression in fluconazole-resistant *Candida albicans* strains. Mol. Microbiol. 69,
22 827–840. <https://doi.org/10.1111/j.1365-2958.2008.06309.x>

23 Dunkel, N., Morschhäuser, J., 2011. Loss of Heterozygosity at an Unlinked Genomic Locus Is
24 Responsible for the Phenotype of a *Candida albicans sap4 Δ sap5 Δ sap6 Δ* Mutant.
25 Eukaryot. Cell 10, 54–62. <https://doi.org/10.1128/EC.00281-10>

26 Favre, B., Ryder, N.S., Didmon, M., 1999. Multiple amino acid substitutions in lanosterol 14 α -
27 demethylase contribute to azole resistance in *Candida albicans*. Microbiology 145, 2715–
28 2725. <https://doi.org/10.1099/00221287-145-10-2715>

29 Feri, A., Loll-Krippleber, R., Commere, P.-H., Maufrais, C., Sertour, N., Schwartz, K., Sherlock, G.,
30 Bougnoux, M.-E., d'Enfert, C., Legrand, M., 2016. Analysis of Repair Mechanisms following
31 an Induced Double-Strand Break Uncovers Recessive Deleterious Alleles in the *Candida*
32 *albicans* Diploid Genome. mBio 7. <https://doi.org/10.1128/mBio.01109-16>

33 Fisher, M.C., Hawkins, N.J., Sanglard, D., Gurr, S.J., 2018. Worldwide emergence of resistance to
34 antifungal drugs challenges human health and food security. Science 360, 739–742.
35 <https://doi.org/10.1126/science.aap7999>

1 Flowers, S.A., Barker, K.S., Berkow, E.L., Toner, G., Chadwick, S.G., Gyax, S.E., Morschhäuser, J.,
2 Rogers, P.D., 2012. Gain-of-function mutations in UPC2 are a frequent cause of ERG11
3 upregulation in azole-resistant clinical isolates of *Candida albicans*. *Eukaryot. Cell* 11,
4 1289–1299. <https://doi.org/10.1128/EC.00215-12>

5 Flowers, S.A., Colón, B., Whaley, S.G., Schuler, M.A., Rogers, P.D., 2015. Contribution of Clinically
6 Derived Mutations in *ERG11* to Azole Resistance in *Candida albicans*. *Antimicrob. Agents*
7 *Chemother.* 59, 450–460. <https://doi.org/10.1128/AAC.03470-14>

8 Gillum, A.M., Tsay, E.Y., Kirsch, D.R., 1984. Isolation of the *Candida albicans* gene for orotidine-5'-
9 phosphate decarboxylase by complementation of *S. cerevisiae* *ura3* and *E. coli* *pyrF*
10 mutations. *Mol. Gen. Genet.* MGG 198, 179–182.

11 Goldman, G.H., da Silva Ferreira, M.E., dos Reis Marques, E., Savoldi, M., Perlin, D., Park, S., Godoy
12 Martinez, P.C., Goldman, M.H.S., Colombo, A.L., 2004. Evaluation of fluconazole resistance
13 mechanisms in *Candida albicans* clinical isolates from HIV-infected patients in Brazil.
14 *Diagn. Microbiol. Infect. Dis.* 50, 25–32.
15 <https://doi.org/10.1016/j.diagmicrobio.2004.04.009>

16 Hargrove, T.Y., Friggeri, L., Wawrzak, Z., Qi, A., Hoekstra, W.J., Schotzinger, R.J., York, J.D.,
17 Guengerich, F.P., Lepesheva, G.I., 2017. Structural analyses of *Candida albicans* sterol
18 14 α -demethylase complexed with azole drugs address the molecular basis of azole-
19 mediated inhibition of fungal sterol biosynthesis. *J. Biol. Chem.* 292, 6728–6743.
20 <https://doi.org/10.1074/jbc.M117.778308>

21 Heilmann, C.J., Schneider, S., Barker, K.S., Rogers, P.D., Morschhäuser, J., 2010. An A643T
22 Mutation in the Transcription Factor *Upc2p* Causes Constitutive ERG11 Upregulation
23 and Increased Fluconazole Resistance in *Candida albicans*. *Antimicrob. Agents*
24 *Chemother.* 54, 353–359. <https://doi.org/10.1128/AAC.01102-09>

25 Hirakawa, M.P., Martinez, D.A., Sakthikumar, S., Anderson, M.Z., Berlin, A., Gujja, S., Zeng, Q.,
26 Zisson, E., Wang, J.M., Greenberg, J.M., Berman, J., Bennett, R.J., Cuomo, C.A., 2015. Genetic
27 and phenotypic intra-species variation in *Candida albicans*. *Genome Res.* 25, 413–425.
28 <https://doi.org/10.1101/gr.174623.114>

29 Jensen, R.H., Astvad, K.M.T., Silva, L.V., Sanglard, D., Jørgensen, R., Nielsen, K.F., Mathiasen, E.G.,
30 Doroudian, G., Perlin, D.S., Arendrup, M.C., 2015. Stepwise emergence of azole,
31 echinocandin and amphotericin B multidrug resistance *in vivo* in *Candida albicans*
32 orchestrated by multiple genetic alterations. *J. Antimicrob. Chemother.* 70, 2551–2555.
33 <https://doi.org/10.1093/jac/dkv140>

34 Kakeya, H., Miyazaki, Y., Miyazaki, H., Nyswaner, K., Grimberg, B., Bennett, J.E., 2000. Genetic
35 analysis of azole resistance in the Darlington strain of *Candida albicans*. *Antimicrob.*
36 *Agents Chemother.* 44, 2985–2990.

1 Keniya, M.V., Sabherwal, M., Wilson, R.K., Sagatova, A.A., Tyndall, J.D.A., Monk, B.C., 2017.
2 Structure of CYP51 from the pathogen *Candida albicans*.
3 <https://doi.org/10.2210/pdb5v5z/pdb>

4 Kullberg, B.J., Arendrup, M.C., 2015. Invasive Candidiasis. *N. Engl. J. Med.* 373, 1445–1456.
5 <https://doi.org/10.1056/NEJMra1315399>

6 Lackner, M., Tscherner, M., Schaller, M., Kuchler, K., Mair, C., Sartori, B., Istel, F., Arendrup, M.C.,
7 Lass-Flörl, C., 2014. Positions and Numbers of *FKS* Mutations in *Candida albicans*
8 Selectively Influence *In Vitro* and *In Vivo* Susceptibilities to Echinocandin Treatment.
9 *Antimicrob. Agents Chemother.* 58, 3626–3635. <https://doi.org/10.1128/AAC.00123-14>

10 Lamb, D.C., Kelly, D.E., White, T.C., Kelly, S.L., 2000. The R467K amino acid substitution in
11 *Candida albicans* sterol 14 α -demethylase causes drug resistance through reduced
12 affinity. *Antimicrob. Agents Chemother.* 44, 63–67.

13 Lee, M.-K., Williams, L.E., Warnock, D.W., Arthington-Skaggs, B.A., 2004. Drug resistance genes
14 and trailing growth in *Candida albicans* isolates. *J. Antimicrob. Chemother.* 53, 217–224.
15 <https://doi.org/10.1093/jac/dkh040>

16 Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler
17 transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760.
18 <https://doi.org/10.1093/bioinformatics/btp324>

19 Löffler, J., Kelly, S.L., Hebart, H., Schumacher, U., Lass-Flörl, C., Einsele, H., 1997. Molecular
20 analysis of *cyp51* from fluconazole-resistant *Candida albicans* strains. *FEMS Microbiol.*
21 *Lett.* 151, 263–268.

22 Lohberger, A., Coste, A.T., Sanglard, D., 2014. Distinct roles of *Candida albicans* drug resistance
23 transcription factors TAC1, MRR1, and UPC2 in virulence. *Eukaryot. Cell* 13, 127–142.
24 <https://doi.org/10.1128/EC.00245-13>

25 MacPherson, S., Akache, B., Weber, S., De Deken, X., Raymond, M., Turcotte, B., 2005. *Candida*
26 *albicans* zinc cluster protein Upc2p confers resistance to antifungal drugs and is an
27 activator of ergosterol biosynthetic genes. *Antimicrob. Agents Chemother.* 49, 1745–
28 1752. <https://doi.org/10.1128/AAC.49.5.1745-1752.2005>

29 Marichal, P., Koymans, L., Willemsens, S., Bellens, D., Verhasselt, P., Luyten, W., Borgers, M.,
30 Ramaekers, F.C.S., Odds, F.C., Vanden Bossche, H., 1999. Contribution of mutations in the
31 cytochrome P450 14 α -demethylase (Erg11p, Cyp51p) to azole resistance in *Candida*
32 *albicans*. *Microbiol. Read. Engl.* 145 (Pt 10), 2701–2713.
33 <https://doi.org/10.1099/00221287-145-10-2701>

34 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K.,
35 Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The Genome Analysis Toolkit: a

1 MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*
2 *Res.* 20, 1297–1303. <https://doi.org/10.1101/gr.107524.110>

3 Morio, F., Jensen, R.H., Le Pape, P., Arendrup, M.C., 2017. Molecular basis of antifungal drug
4 resistance in yeasts. *Int. J. Antimicrob. Agents* 50, 599–606.
5 <https://doi.org/10.1016/j.ijantimicag.2017.05.012>

6 Morio, F., Loge, C., Besse, B., Hennequin, C., Le Pape, P., 2010. Screening for amino acid
7 substitutions in the *Candida albicans* Erg11 protein of azole-susceptible and azole-
8 resistant clinical isolates: new substitutions and a review of the literature. *Diagn.*
9 *Microbiol. Infect. Dis.* 66, 373–384. <https://doi.org/10.1016/j.diagmicrobio.2009.11.006>

10 Noble, S.M., Johnson, A.D., 2007. Genetics of *Candida albicans*, a diploid human fungal pathogen.
11 *Annu. Rev. Genet.* 41, 193–211.
12 <https://doi.org/10.1146/annurev.genet.41.042007.170146>

13 Oliver, B.G., Song, J.L., Choiniere, J.H., White, T.C., 2007. cis-Acting elements within the *Candida*
14 *albicans* ERG11 promoter mediate the azole response through transcription factor
15 Upc2p. *Eukaryot. Cell* 6, 2231–2239. <https://doi.org/10.1128/EC.00331-06>

16 Pappas, P.G., Kauffman, C.A., Andes, D.R., Clancy, C.J., Marr, K.A., Ostrosky-Zeichner, L., Reboli,
17 A.C., Schuster, M.G., Vazquez, J.A., Walsh, T.J., Zaoutis, T.E., Sobel, J.D., 2016. Clinical
18 Practice Guideline for the Management of Candidiasis: 2016 Update by the Infectious
19 Diseases Society of America. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* 62, e1-50.
20 <https://doi.org/10.1093/cid/civ933>

21 Perea, S., Lopez-Ribot, J.L., Kirkpatrick, W.R., McAtee, R.K., Santillan, R.A., Martinez, M., Calabrese,
22 D., Sanglard, D., Patterson, T.F., 2001. Prevalence of Molecular Mechanisms of Resistance
23 to Azole Antifungal Agents in *Candida albicans* Strains Displaying High-Level Fluconazole
24 Resistance Isolated from Human Immunodeficiency Virus-Infected Patients. *Antimicrob.*
25 *Agents Chemother.* 45, 2676–2684. [https://doi.org/10.1128/AAC.45.10.2676-](https://doi.org/10.1128/AAC.45.10.2676-2684.2001)
26 [2684.2001](https://doi.org/10.1128/AAC.45.10.2676-2684.2001)

27 Pfaller, M.A., Messer, S.A., Woosley, L.N., Jones, R.N., Castanheira, M., 2013. Echinocandin and
28 triazole antifungal susceptibility profiles for clinical opportunistic yeast and mold
29 isolates collected from 2010 to 2011: application of new CLSI clinical breakpoints and
30 epidemiological cutoff values for characterization of geographic and temporal trends of
31 antifungal resistance. *J. Clin. Microbiol.* 51, 2571–2581.
32 <https://doi.org/10.1128/JCM.00308-13>

33 Pucci, F., Bourgeas, R., Rومان, M., 2016. Predicting protein thermal stability changes upon
34 point mutations using statistical potentials: Introducing HoTMuSiC. *Sci. Rep.* 6.
35 <https://doi.org/10.1038/srep23257>

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

7 Sagatova, A.A., Keniya, M.V., Wilson, R.K., Monk, B.C., Tyndall, J.D.A., 2015. Structural Insights into
8 Binding of the Antifungal Drug Fluconazole to *Saccharomyces cerevisiae* Lanosterol 14 α -
9 Demethylase. *Antimicrob. Agents Chemother.* 59, 4982–4989.
10 <https://doi.org/10.1128/AAC.00925-15>

11 Sanglard, D., Bille, J., 2002. Action of and resistance to antifungal agents. In: *Candida and*
12 *candidiasis*. Calderone RA, p. 370.

13 Sanglard, D., Ischer, F., Koymans, L., Bille, J., 1998. Amino acid substitutions in the cytochrome P-
14 450 lanosterol 14 α -demethylase (CYP51A1) from azole-resistant *Candida albicans*
15 clinical isolates contribute to resistance to azole antifungal agents. *Antimicrob. Agents*
16 *Chemother.* 42, 241–253.

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

19 Selmecki, A., Gerami-Nejad, M., Paulson, C., Forche, A., Berman, J., 2008. An isochromosome
20 confers drug resistance in vivo by amplification of two genes, ERG11 and TAC1. *Mol.*
21 *Microbiol.* 68, 624–641. <https://doi.org/10.1111/j.1365-2958.2008.06176.x>

22 Selmecki, A.M., Dulmage, K., Cowen, L.E., Anderson, J.B., Berman, J., 2009. Acquisition of
23 aneuploidy provides increased fitness during the evolution of antifungal drug resistance.
24 *PLoS Genet.* 5, e1000705. <https://doi.org/10.1371/journal.pgen.1000705>

25 Walther, A., Wendland, J., 2003. An improved transformation protocol for the human fungal
26 pathogen *Candida albicans*. *Curr. Genet.* 42, 339–343. [https://doi.org/10.1007/s00294-](https://doi.org/10.1007/s00294-002-0349-0)
27 [002-0349-0](https://doi.org/10.1007/s00294-002-0349-0)

28 Wang, H., Kong, F., Sorrell, T.C., Wang, B., McNicholas, P., Pantarat, N., Ellis, D., Xiao, M., Widmer,
29 F., Chen, S.C., 2009. Rapid detection of ERG11 gene mutations in clinical *Candida albicans*
30 isolates with reduced susceptibility to fluconazole by rolling circle amplification and
31 DNA sequencing. *BMC Microbiol.* 9, 167. <https://doi.org/10.1186/1471-2180-9-167>

32 Whaley, S.G., Berkow, E.L., Rybak, J.M., Nishimoto, A.T., Barker, K.S., Rogers, P.D., 2017. Azole
33 Antifungal Resistance in *Candida albicans* and Emerging Non-*albicans* *Candida* Species.
34 *Front. Microbiol.* 7. <https://doi.org/10.3389/fmicb.2016.02173>

- 1 White, T.C., 1997. The presence of an R467K amino acid substitution and loss of allelic variation
2 correlate with an azole-resistant lanosterol 14 α demethylase in *Candida albicans*.
3 *Antimicrob. Agents Chemother.* 41, 1488–1494.
- 4 Wu, Y., Gao, N., Li, C., Gao, J., Ying, C., 2017. A newly identified amino acid substitution T123I in
5 the 14 α -demethylase (Erg11p) of *Candida albicans* confers azole resistance. *FEMS Yeast*
6 *Res.* 17. <https://doi.org/10.1093/femsyr/fox012>
- 7 Xu, Y., Chen, L., Li, C., 2008. Susceptibility of clinical isolates of *Candida* species to fluconazole
8 and detection of *Candida albicans* ERG11 mutations. *J. Antimicrob. Chemother.* 61, 798–
9 804. <https://doi.org/10.1093/jac/dkn015>

10

11 **Acknowledgements**

12 **Funding**

13 E.S. was the recipient of a Ph.D. grant from DIM Malinf -Région Ile-de-France.

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

17

18 **Transparency declarations**

19 The authors declare no competing interests

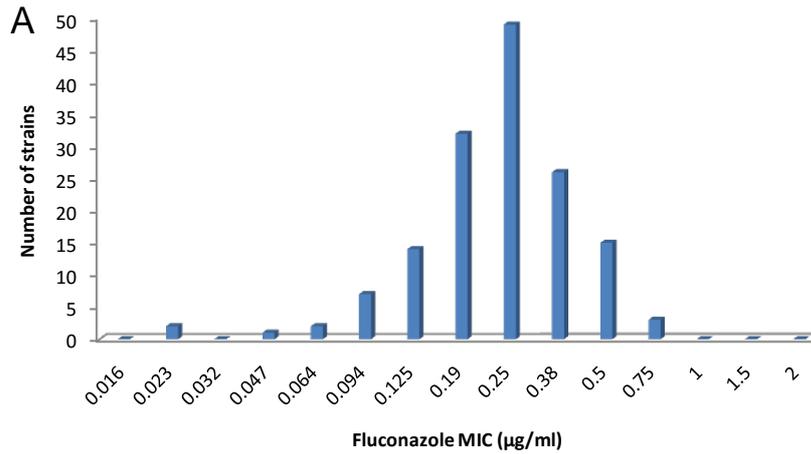
20

21

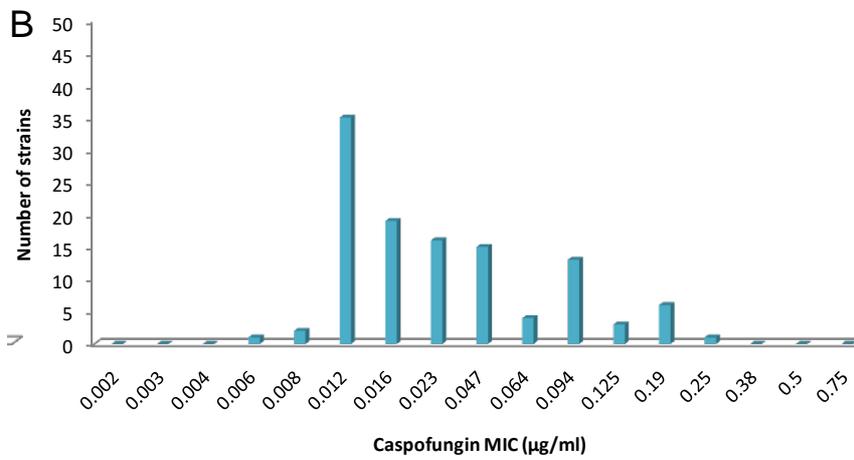
1 **Table and Figures**

2 **Figure 1: Distribution of fluconazole (A) and caspofungin (B) MICs for the 151**
3 **susceptible clinical strains of *C. albicans*, determined by Etest® method in RPMI**
4 **medium.**

5



6



7

1 **Table 1: Characteristics of natural polymorphisms in genes involved in resistance to**
2 **azoles and echinocandins.**

3 The repertoire includes 126 non-synonymous natural polymorphism positions (ORF position
4 and amino acid correspondence) detected in the 5 genes involved in resistance to azoles
5 (*ERG11*, *TAC1*, *MRR1* and *UPC2*) and echinocandins (*FKSI*) through the use of the 151
6 susceptible strains. **For each strain the genotype at each position is compared to that of**
7 **the reference strain SC5314 and resumed by the homozygous and heterozygous**
8 substitution frequency. h = substitution found only at the heterozygous state.

Gene name	ORF position	SC5314 reference strain genotype	Substitution frequency (%)				Amino acid substitutions
			Heterozygous		Homozygous		
TAC1	261	TT	TG	8.4	GG	0.65	K87N
TAC1	268	CC	CT	0.65			A90T ^h
TAC1	310	AC	AC	49.4	CC	2.6	F104V
TAC1	323	CC	CT	16.1	TT	3.2	S108N
TAC1	391	GG	GT	23.2	TT	1.3	L131I
TAC1	445	AA	AT	1.3			L149I ^h
TAC1	508	TT	TC	20.65	CC	3.2	M170V
TAC1	510	CC	CA	9.0	AA	0.65	M170I
TAC1	520	TT	TC	9.0	CC	0.65	N174D
TAC1	566	AA	AG	28.8	GG	9.2	F189S
TAC1	593	TT	TC	1.3			H198R ^h
TAC1	596	CT	CT	33.3	CC	0.65	S199N
TAC1	617	CT	CT	36.1	CC	2.0	R206H
TAC1	620	AG	AG	36.6	AA	2.0	V207A
TAC1	628	CC	CA	0.65			A210S ^h
TAC1	1121	GG	GA	0.65			T374I ^h
TAC1	1130	GG	GA	12.3			A377V ^h
TAC1	1187	TC	TC	33.1	TT	1.3	N396S
TAC1	1924	TT	TC	1.3			T642A ^h
TAC1	1994	AA	AG	0.65			F665S ^h
TAC1	2083	GG	GA	13.5	AA	1.3	L695F
TAC1	2123	TT	TG	0.65			K708T ^h
TAC1	2140	TT	TC	0.65			S714G ^h
TAC1	2276	AA	AG	0.65			L759S ^h
TAC1	2291	CC	CT	2.0			S764N ^h
TAC1	2298	TT	TA	1.3			K766N ^h
TAC1	2316	AT	AT	51.0	TT	5.2	N772K
TAC1	2326	CT	CT	31.6	CC	0.65	D776N
TAC1	2344	TT	TC	0.65			T782A ^h
TAC1	2485	CG	CG	43.4	CC	9.2	E829Q
TAC1	2495	AA	AG	0.65			F832S ^h
TAC1	2606	CC	CT	2.6			R869Q ^h
TAC1	2684	AA	AG	14.2			I895T
TAC1	2687	TT	TC	49.4	CC	4.5	N896S
TAC1	2723	CC	CT	1.3	TT	0.65	C908Y
TAC1	2804	GA	GA	23.9	AA	1.3	S935L
TAC1	2810	GG	GA	8.4			S937L ^h
TAC1	2822	AG	AG	73.4	AA	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	AT	32.9	TT	7.1	D116E
<i>ERG11</i>	383	TG	TG	19.4	GG	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
<i>UPC2</i>	1519	CC	CT	1.3			V507I ^h
<i>UPC2</i>	1904	AA	AC	0.65			L635W ^h

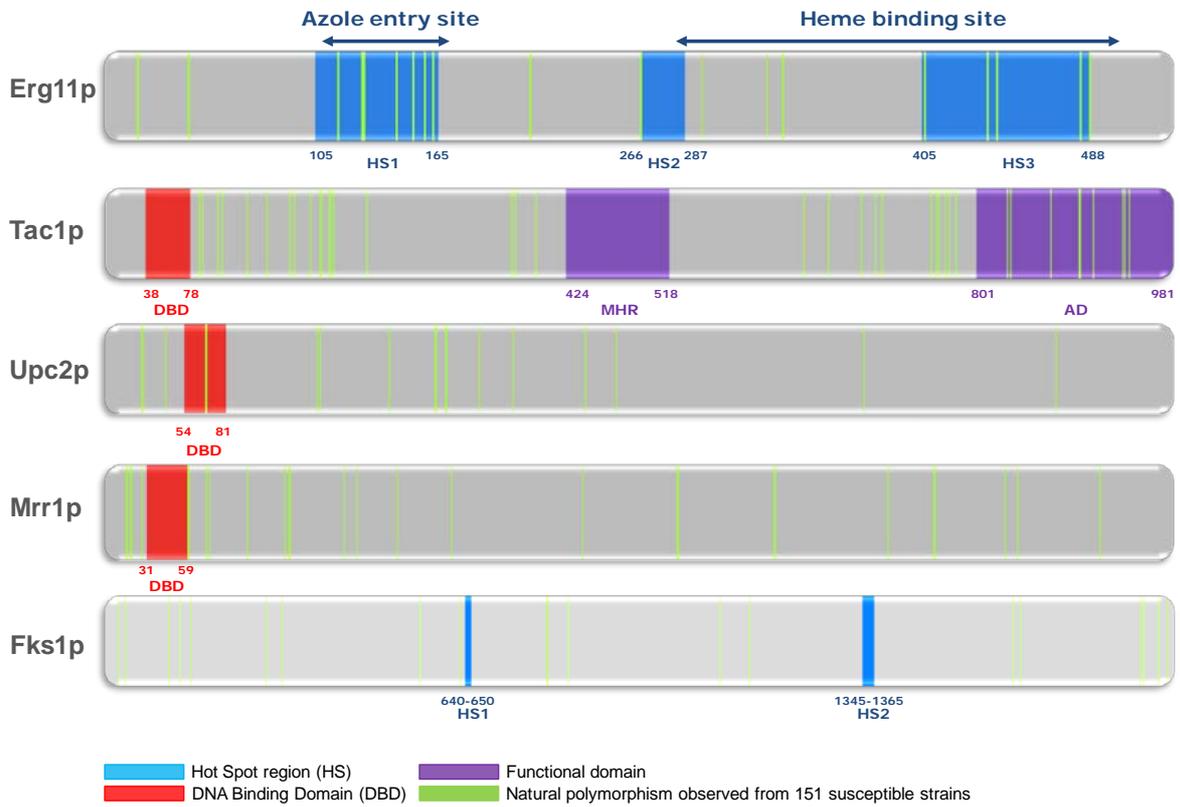
1

2

MRR1	47	GG	GT	3.6	TT	1.3	S16I
MRR1	56	CC	CT	2.0	TT	18.7	P19L
MRR1	79	GG	GA	27.1	AA	3.9	V27I
MRR1	179	TT	TC	1.3	CC	0.65	V60A
MRR1	218	CC	CA	3.6	AA	1.3	T73N
MRR1	223	GG	GA	2.0	AA	16.0	G75R
MRR1	304	GG	GA	2.0			A102T ^h
MRR1	387	TT	TG	0.65			N129K ^h
MRR1	394	CC	CA	0.65			Q132K ^h
MRR1	511	CC	CT	0.65			S171L ^h
MRR1	512	TT	TC	37.2	CC	18.2	S171P
MRR1	539	CC	CT	0.65			P180L ^h
MRR1	626	AA	AC	0.65			Q209P ^h
MRR1	742	TT	TG	28.0	GG	3.9	L248V
MRR1	1022	TT	TA	32.3	AA	11.8	V341E
MRR1	1226	GG	GA	1.3			G409E ^h
MRR1	1432	AA	AT	1.3			I478L ^h
MRR1	1675	AA	AT	1.3			T559S ^h
MRR1	1774	CC	CT	30.7	TT	4.5	L592F
MRR1	1925	CC	CT	0.65			A642V ^h
MRR1	1951	GG	GA	2.0			D651N ^h
MRR1	2128	GG	GA	0.65			E710K ^h
MRR1	2497	CC	CT	1.3			R833* ^h
MRR1	2750	CC	CT	9.4	TT	1.3	T917M
MRR1	2756	CC	CT	2.0			A919V ^h
MRR1	2768	CC	CT	9.4	TT	1.3	T923I
MRR1	2811	CC	CG	0.65	GG	18.1	N938K
MRR1	2900	GG	GA	0.65			G967D ^h
MRR1	3058	GG	GC	30.7	CC	7.9	E1020Q
MRR1	3082	GG	GA	0.65			E1028K ^h
MRR1	3096	TT	TA	2.1	AA	18.8	F1032L
MRR1	3110	CC	CT	2.1	TT	17.2	S1037L
MRR1	3265	GG	GT	0.65			A1089S ^h
FKS1	71	CC	CT	7.1	TT	0.65	G24E
FKS1	109	CC			TT	0.65	G37S
FKS1	340	AA	AC	0.65			Y114D ^h
FKS1	400	CC			TT	0.65	G134S
FKS1	457	CC	CT	0.65			A153T ^h
FKS1	859	CC	CT	0.65			A287T ^h
FKS1	943	CC	CT	0.65			G315S ^h
FKS1	1679	TT	CT	0.65			Y560C ^h
FKS1	1891	AA	AG	0.65			F631L ^h
FKS1	2355	TT	TA	0.65			E785D ^h
FKS1	2356	TT	TC	0.65			I786V ^h
FKS1	2496	CC	CA	0.65			M832I ^h
FKS1	3274	CC	CA	0.65			A1092S ^h
FKS1	3430	CC	CT	0.65			D1144N ^h
FKS1	4834	AA			TT	1.3	L1612M
FKS1	4873	CC	CT	0.65			D1625N ^h
FKS1	5512	GC	GC	45.4	GG	9.1	P1838A
FKS1	5531	TT	TG	0.65			N1844T ^h
FKS1	5609	CC	CT	1.3			R1870K ^h
FKS1	5657	GC	GC	36.1	CC	24.5	T1886S

1

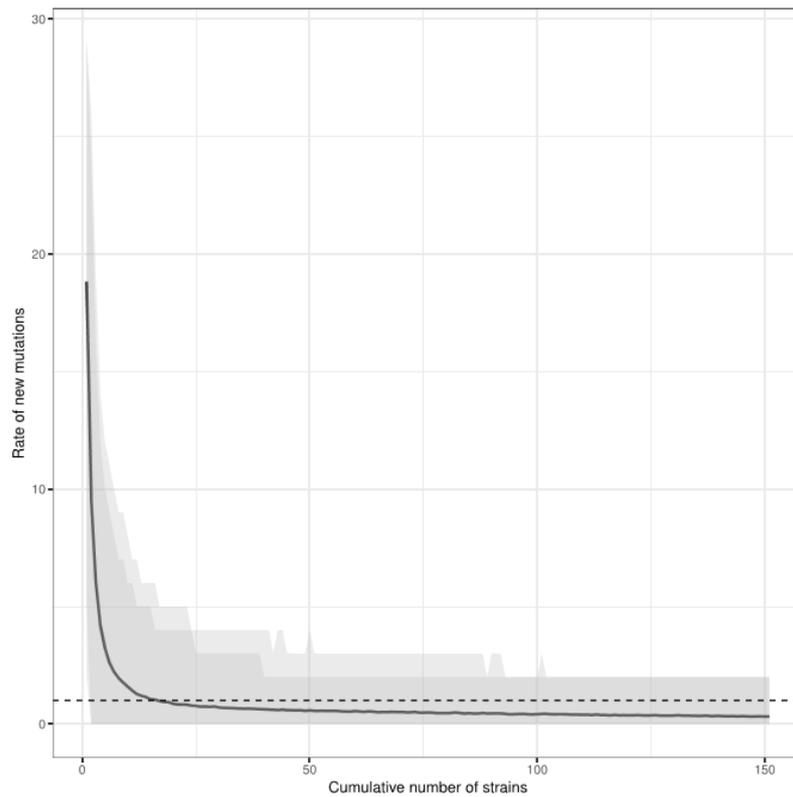
2 **Figure 2: Cartography of the natural amino acid substitutions for the different proteins**
3 **analyzed (Erg11p, Tac1p, Upc2p, Mrr1p and Fks1p) observed from the 151 *C. albicans***
4 **susceptible clinical strains.**



5

6

1 **Figure 3: Rate of new mutations detected in the set of 151 susceptible strains.** The solid
2 line shows the mean number of new mutations that can be found for each new strain explored
3 in the randomly resampled sets from the developed bootstrap method. The grey envelopes
4 correspond to 95% and 90% credible intervals. Dashed line shows the threshold of 1
5 mutation/strain.



6

7

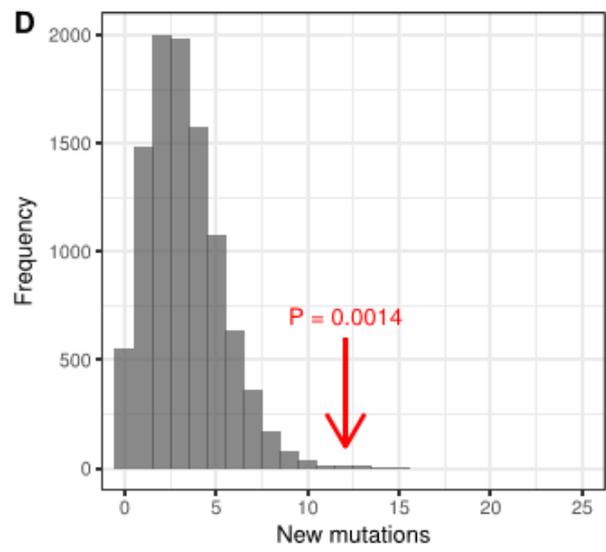
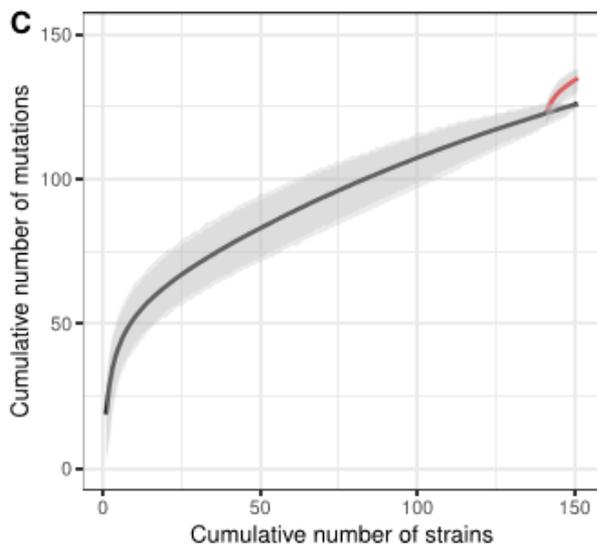
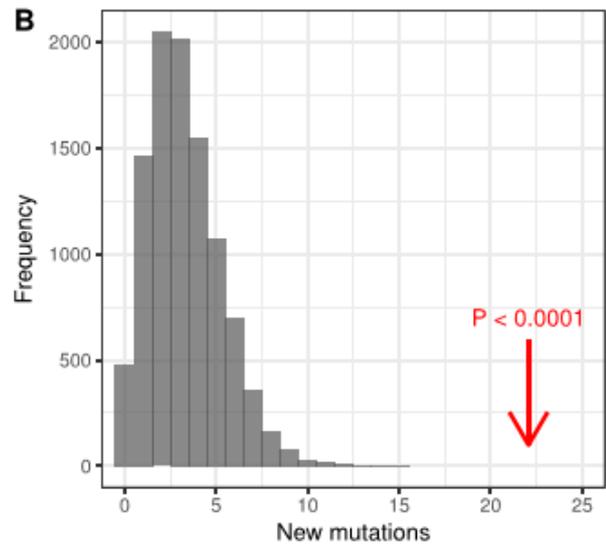
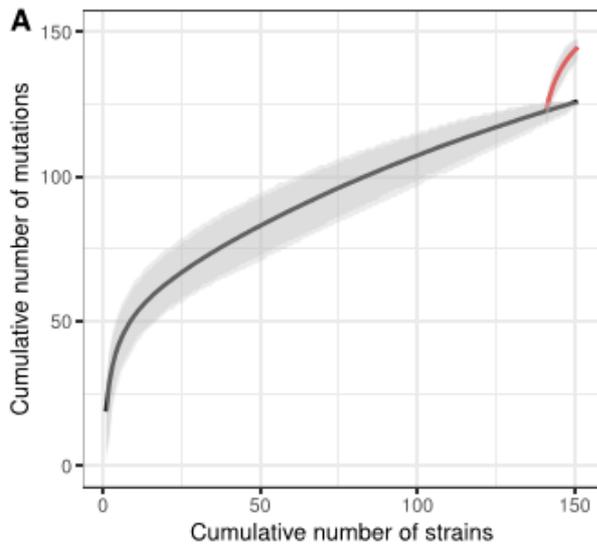
1 **Table 2: Results of *in vitro* fluconazole and caspofungine susceptibility testing, amino**
2 **acid substitutions and overexpression level of genes (*CDR2* and *ERG11*) for the 10**
3 **resistant strains of *C. albicans*.** Only amino acid substitutions selected by our tool as
4 possibly being involved in antifungal resistance are shown. These one includes resistance
5 mutations previously described in the literature and the new amino acid substitutions selected
6 (boldface) in the protein Tac1p, Upc2p, Erg11p and Fks1p. New resistant mutations validated
7 in this study are marked with an asterisk (*). h = heterozygous mutation found in a single
8 allele; FLC = fluconazole; CAS = caspofungin.

9

10

Isolate	MIC ($\mu\text{g/ml}$)			Amino acid substitutions					Overexpression level (n fold)		
	FLC	CAS		Erg11p	Tac1p	Upc2p	Fks1p	CDR2 (compared to CEC718)	ERG11 (compared to SC5314)		
CEC718	>256	0.047		Y132H, P236S, S405F	A736T	G657D		Used as control strain	7.5		
CEC723	16	0.064		P30S ^h , S405F	A790V*			1.7			
CEC4504	24	0.047		R157K ^h	H263Y*, H839Y*, C858Y			12.1			
CEC4503	>256	0.0125		D9V ^h , R157K,	H263Y*, H839Y*, C858Y						
CEC4515	>256	0.094		Y132H, G450E	H263Y*, P971S*	G648D ^h		2.7			
CEC4516	>256	0.125		Y132H, G450E	H263Y*, P971S*						
CEC4521	>256	0.064		Y132H, R467I	H263Y*, P971S*	D654N ^h			4.5		
CEC4522	>256	0.19		Y132H, R467I	H263Y*, P971S*	D654N			5.9		
CEC4524	>256	0.5		Y64H, T123I, Y132H, R467I	H263Y*, P971S*	D654N ^h	R647G				
CEC4527	8	0.047		G448E	A736T, N823K, C858Y			3.9			

1 **Figure 4: Contribution of resistant strains to genetic diversity of *C. albicans*.**
2 Accumulation curves of mutations as a function of the random selection of clinical isolates of
3 *C. albicans*, accounting for all mutations present in resistant strains (A) or restricting those to
4 undocumented mutations only (C). Solid lines present the mean accumulation of mutations
5 with increasing number of strains (black: random resampling of all susceptible strains
6 (n=151); red: random samples of 141 susceptible strains with final introduction of the 10
7 resistant strains). Grey envelopes correspond to 95% and 90% credible intervals. Histograms
8 (B, D) show the distribution of new mutations in the last 10 susceptible strains, with red
9 arrows identifying the amount brought when considering resistant strains instead under
10 scenarios used for panels A and C, respectively.

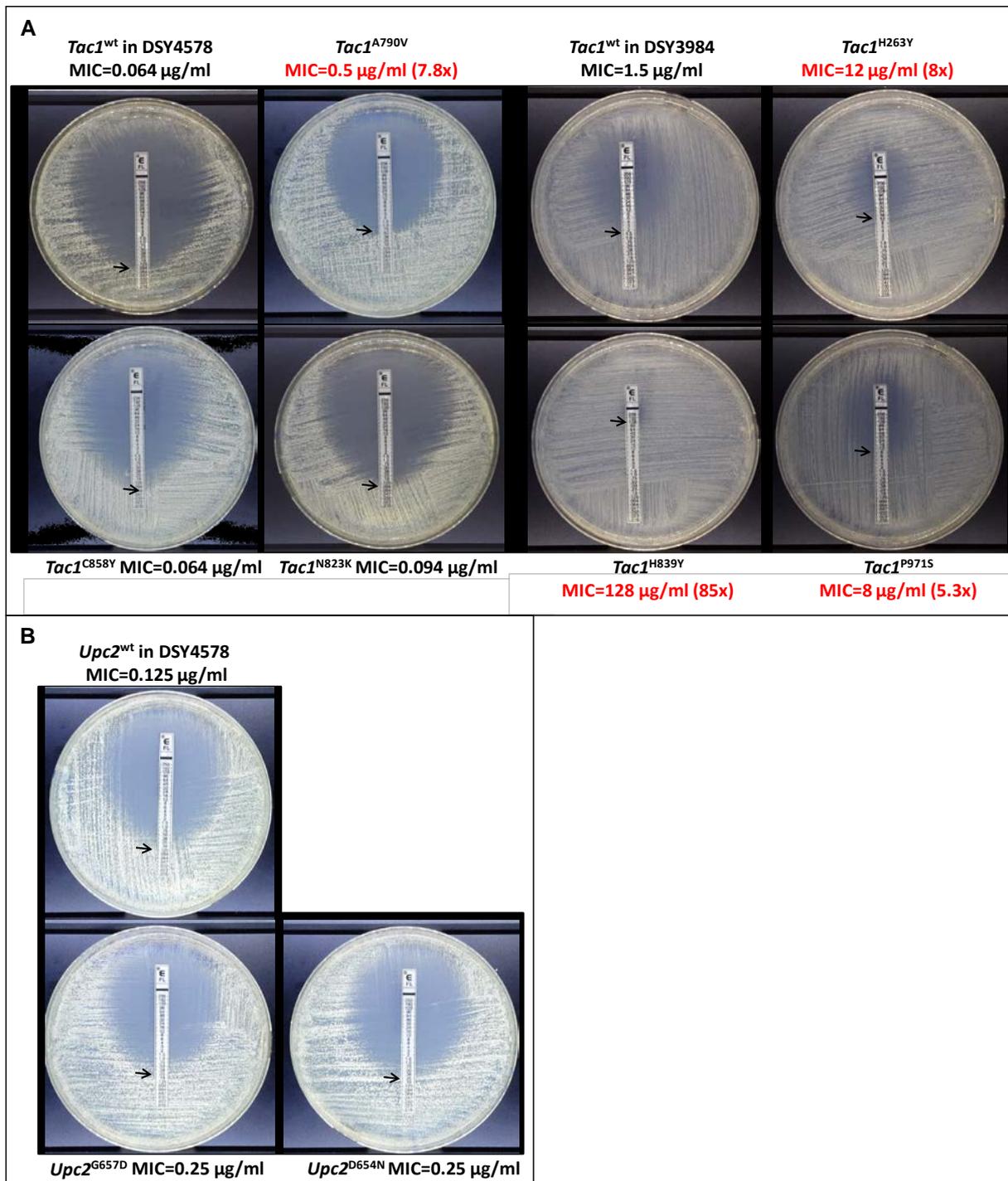


Sample — 141 S, 10 R — 151 S

1

2

- 1 **Figure 5: Fluconazole MICs determination of the mutant constructed strains by Etest®**
- 2 **method. A. Mutants obtained for *Tac1* gene. B. Mutants obtained for *Upc2* gene.**



3

1 **Supplementary Tables S1 to S6**

2
3 **A Useful Tool to Highlight New Resistance Mutations in Genes Involved in**

4 ***Candida albicans* Resistance to Azoles and Echinocandines**

5
6 *Emilie Sitterlé*^{1,2,3}, *Alix Coste*⁴, *Thomas Obadia*^{5,6}, *Corinne Maufrais*⁵, *Murielle Chauvel*¹,
7 *Natacha Sertour*¹, *Anne Puel*^{7,8}, *Dominique Sanglard*⁴, *Christophe d'Enfert*¹, *Marie-*
8 *Elisabeth Bougnoux*^{1,3}

1 **Table S1: Strain characteristics and summary of the sequencing data results from the**
 2 **151 susceptible clinical isolates of *C. albicans* and the reference strain SC5314.**

Strain	Origin	Country	MIC (µg/ml)		MLST		Illumina	
			FLC	CAS	DST	Clade	Sequencing depth (mean)	Read (No.)
SC5314	inv	unknown	0,25	0,032	52	1	112	8707032
CEC1424	sup	Niger	0,38	0,032	949	1	66	5470365
CEC3534	com	Morocco	0,25	0,032	79	1	56	5136349
CEC3623	com	Morocco	0,5	0,023	1364	1	113	9170218
CEC3627	com	Brazil	0,25	0,032	435	1	42	3939244
CEC3544	com	Belgium	0,38	0,023	572	1	74	6285370
CEC3609	com	Morocco	0,25	0,047	1399	1	62	5662630
CEC3672	inv	France	0,25	0,064	69	1	43	4238221
CEC4108	sup	Senegal	0,19	0,032	3044	1	45	3587633
CEC3682	inv	France	0,25	0,023	408	1	59	5306089
CEC3673	inv	France	0,25	0,032	69	1	51	5317899
CEC3560	sup	Niger	0,25	0,016	24	1	54	5677630
CEC3621	com	Brazil	0,25	0,094	1400	1	146	11408364
CEC3678	sup	France	0,75	0,032	2273	1	42	4214744
CEC4106	sup	Senegal	0,19	0,032	2603	1	85	6580720
CEC2020	inv	Morocco	0,38	0,047	69	1	46	3633721
CEC3662	inv	France	0,19	0,064	2231	1	96	8366048
CEC3603	com	Morocco	0,25	0,023	1393	1	46	4554364
CEC3605	com	Morocco	0,5	0,047	1388	1	42	4428884
CEC4029	sup	UK	0,38	0,047	69	1	164	12747576
CEC4035	sup	UK	0,25	0,19	2326	1	143	10941419
CEC4256	sup	France	0,75	0,023	803	1	97	7596500
CEC3554	sup	France	0,25	0,047	69	1	64	6008461
CEC3559	sup	France	0,19	0,016	129	1	57	5519385
CEC4479	NA	France	0,25	0,125	408	1	101	7729673
CEC4480	NA	France	0,19	0,25	364	1	182	13535475
CEC4484	NA	France	0,38	0,125	3103	1	111	8454367
CEC4488	NA	France	0,125	0,094	3098	1	140	10439551
CEC4489	NA	France	0,19	0,064	237	1	75	5809128
CEC4496	NA	France	0,25	0,125	3108	1	178	13696025
CEC4499	NA	France	0,125	0,25	3110	1	152	12021248
CEC4500	NA	France	0,25	0,125	3111	1	205	16263673
CEC3636	sup	UK	0,25	0,032	232	2	131	10436389
CEC4107	sup	UK	0,094	0,125	208	2	47	3735732
CEC2021	inv	France	0,38	0,047	916	2	83	6641033
CEC3531	com	France	0,25	0,032	216	2	73	6927270
CEC3669	sup	France	0,25	0,047	2235	2	73	6823220
CEC3549	com	Belgium	0,25	0,016	999	2	74	6412696
CEC3612	com	Morocco	0,25	0,047	1396	2	56	5541089
CEC3553	com	Belgium	0,25	0,023	910	2	52	5002968
CEC3614	com	Morocco	0,5	0,064	1398	2	66	6322170
CEC4022	sup	UK	0,25	0,032	232	2	168	13012182
CEC3615	com	Morocco	0,25	0,023	1369	2	80	7534932
CEC4034	sup	UK	0,094	0,125	208	2	73	5582345
CEC3558	com	Belgium	0,5	0,064	194	2	46	4637902
CEC4482	NA	France	0,19	0,25	3102	2	171	13153847
CEC4487	NA	France	0,38	0,19	3105	2	168	12532368
CEC4493	NA	France	0,125	0,25	3100	2	160	13110883
CEC4494	NA	France	0,125	0,25	3106	2	203	15977152
CEC3658	sup	France	0,25	0,023	2194	3	47	4521756
CEC4259	com	Guiana	0,5	0,032	1953	3	40	3073804

3

4 (continued on next page)

1 **Table S1: (Continued)**

Strain	Origin	Country	MIC (µg/ml)		MLST		Illumina	
			FLC	CAS	DST	Clade	Sequencing depth (mean)	Read (No.)
CEC1289	inv	France	0,5	0,023	15	3	36	3445232
CEC3626	com	Belgium	0,5	0,032	865	3	48	4820388
CEC3689	sup	France	0,25	0,023	2194	3	71	6458912
CEC3681	inv	France	0,19	0,016	2198	3	69	6053177
CEC3557	com	Belgium	0,019	0,032	868	3	65	6143946
CEC3596	inv	France	0,38	0,016	33	3	73	7093326
CEC709	sup	France	0,19	0,032	1643	3	71	6301192
CEC3597	inv	France	0,38	0,016	17	3	84	7884850
CEC3631	com	Guiana	0,5	0,032	1953	3	42	4349585
CEC3638	com	Belgium	0,25	0,016	1000	3	254	25402156
CEC3579	com	France	0,38	0,016	1769	3	110	8684792
CEC3548	com	Belgium	0,38	0,032	461	3	78	6754659
CEC3715	sup	France	0,25	0,016	2279	3	80	7746400
CEC4038	sup	UK	0,75	0,023	2276	3	54	4207575
CEC3529	sup	UK	0,25	0,016	2278	4	66	6192413
CEC3540	inv	France	0,5	0,023	27	4	74	6596123
CEC1492	unknown	unknown	0,25	0,016	124	4	108	9096726
CEC3532	com	France	0,38	0,016	1766	4	62	6130025
CEC2022	inv	France	0,25	0,016	95	4	53	4197852
CEC3535	com	Morocco	0,25	0,125	605	4	50	4566710
CEC3541	inv	France	0,023	0,38	28	4	64	5650726
CEC3530	com	France	0,38	0,032	1771	4	57	5163333
CEC3551	sup	France	0,064	0,125	1004	4	76	6793503
CEC3602	com	Morocco	0,25	0,016	504	4	39	3669968
CEC3611	com	Morocco	0,25	0,016	1389	4	64	5876995
CEC3659	sup	France	0,19	0,032	914	4	50	4748712
CEC3716	sup	France	0,125	0,016	924	4	49	4738167
CEC3675	inv	France	0,094	0,023	423	4	63	5951944
CEC4026	sup	UK	0,25	0,016	2278	4	230	17759454
CEC3536	com	Morocco	0,19	0,032	423	4	36	3575381
CEC708	sup	France	0,25	0,032	1642	4	59	5516754
CEC3676	sup	France	0,094	0,032	95	4	75	7113137
CEC3610	com	Morocco	0,25	0,016	1386	4	82	7488787
CEC3556	com	Morocco	0,25	0,016	124	4	56	5235993
CEC3607	com	Morocco	0,25	0,032	1405	4	62	6174715
CEC704	sup	France	0,38	0,016	504	4	55	7392463
CEC711	sup	France	0,19	0,047	1644	4	39	3836645
CEC3674	sup	France	0,064	0,016	2274	4	79	7408293
CEC3683	inv	France	0,19	0,047	423	4	58	5465892
CEC712	sup	France	0,25	0,016	1645	4	57	5350740
CEC3555	inv	France	0,38	0,094	13	4	59	5755895
CEC4497	NA	France	0,125	0,125	924	4	159	12546328
CEC3637	sup	Niger	0,38	0,064	950	5	101	9658293
CEC3671	inv	France	0,19	0,047	840	7	50	4921679
CEC2023	com	Guiana	0,19	0,032	602	8	138	10998631
CEC3622	com	Brazil	0,094	0,016	602	8	58	5770762
CEC3634	com	Brazil	0,094	0,023	1360	8	53	5131292
CEC3613	com	Morocco	0,5	0,064	1394	8	54	5450432
CEC2018	sup	France	0,25	0,032	529	9	90	7401988
CEC3533	com	France	0,25	0,016	1767	9	50	4734751
CEC3668	sup	France	0,38	0,016	2234	9	71	6421866

2

3 (continued on next page)

1 **Table S1: (Continued)**

Strain	Origin	Country	MIC ($\mu\text{g/ml}$)		MLST		Illumina	
			FLC	CAS	DST	Clade	Sequencing depth (mean)	Read (No.)
CEC4261	sup	France	1	0,016	1717	9	71	5710362
CEC3679	sup	UK	0,38	0,032	2275	10	63	5797140
CEC3616	com	Morocco	0,25	0,032	1363	10	79	7290642
CEC3706	sup	France	0,25	0,064	2227	10	63	5879460
CEC3711	sup	France	0,19	0,016	2195	10	61	5706530
CEC4024	sup	UK	0,38	0,032	2275	10	198	15710470
CEC3617	inv	France	0,19	0,012	2197	11	62	5796879
CEC1426	sup	France	0,5	0,008	1002	11	59	4831040
CEC3494	com	Belgium	0,125	0,032	903	11	38	3124008
CEC3547	com	Belgium	0,38	0,016	907	11	77	6678861
CEC3561	com	Belgium	0,38	0,016	906	11	70	6331884
CEC3601	com	Belgium	0,25	0,016	908	11	62	5596818
CEC3687	inv	France	0,19	0,012	2197	11	95	8617466
CEC3704	sup	France	0,19	0,012	912	11	71	6441472
CEC3618	com	Belgium	0,25	0,032	912	11	51	5077456
CEC4492	NA	France	0,25	0,047	3099	11	126	9497094
CEC4495	NA	France	0,19	0,064	3107	11	157	12218003
CEC2019	sup	Niger	0,25	0,032	947	12	84	6915795
CEC3686	inv	France	0,25	0,047	2193	12	66	6151544
CEC716	sup	France	0,19	0,016	1646	12	62	5605992
CEC3537	com	Morocco	0,38	0,032	601	12	66	5771341
CEC3685	inv	France	0,19	0,023	947	12	60	5852693
CEC3680	inv	France	0,094	0,032	299	12	80	7279876
CEC4103	sup	Senegal	0,38	0,125	182	13	45	3489420
CEC4104	sup	Senegal	0,125	0,125	182	13	49	3829872
CEC3550	starling	France	0,25	0,032	942	16	87	7642786
CEC3600	starling	France	0,38	0,064	943	16	89	8228040
CEC2871	inv	S. Korea	0,19	0,023	732	18	51	4033383
CEC2872	inv	S. Korea	0,25	0,19	727	18	77	6387004
CEC3786	inv	S. Korea	0,094	0,032	1598	18	78	6136963
CEC2876	inv	S. Korea	0,19	0,032	1596	18	48	3922049
CEC4032	sup	UK	0,125	0,125	2280	NA	59	4636390
CEC3660	sup	UK	0,125	0,064	2324	NA	42	4092222
CEC4031	sup	UK	0,125	0,125	2280	NA	214	16803734
CEC4023	sup	UK	0,125	0,064	2324	NA	187	14846163
CEC1427	inv	France	0,25	0,016	11	NC	62	5128506
CEC3664	sup	France	0,19	0,023	2232	NC	51	5126010
CEC3708	sup	France	0,125	0,032	2228	NC	108	9466569
CEC3661	sup	France	0,19	0,064	2230	NC	57	5309861
CEC4039	sup	UK	0,5	0,094	2327	NC	78	6104047
CEC3707	sup	France	0,125	0,016	2226	NC	59	5285165
CEC3712	sup	France	0,19	0,023	367	NC	55	5189928
CEC3619	com	Belgium	0,38	0,023	911	NC	57	4937609
CEC3663	inv	France	0,19	0,047	2281	NC	99	8954225
CEC3665	sup	France	0,047	0,047	2229	NC	64	6021953
CEC4481	NA	France	0,19	0,064	3101	NC	126	9486589
CEC4485	NA	France	0,19	0,064	3104	NC	115	9045276
CEC4486	NA	France	0,125	0,25	3097	NC	153	11754158
CEC4498	NA	France	0,25	0,125	3109	NC	149	11610114
CEC4254	inv	France	0,125	0,016	2425	NC	70	5610427

2

3 *Sup= superficial; com= commensal; inv=invasive; NA = not applicable; NC = no cladable; UK=*

4 *United Kingdom; FLC= fluconazole; CAS= caspofungin, DST= Diploid Sequence Type.*

1 **Table S1: Strain characteristics and summary of the sequencing data results from the 10**
 2 **resistant clinical isolates of *C. albicans*.**

Strain	Origin	Country	MIC ($\mu\text{g/ml}$)		MLST		Illumina	
			FLC	CAS	DST	Clade	Sequencing depth (mean)	Read (No.)
CEC718	sup	France	>256	0.047	69	1	51	4635247
CEC723	sup	France	16	0.064	2277	NC	64	5825178
CEC4504	sup	France	24	0.047	912	11	313	19788044
CEC4503	sup	France	>256	0.0125	912	11	78	6579903
CEC4515	sup	France	>256	0.094	1327	11	95	8022291
CEC4516	sup	France	>256	0.125	1327	11	139	11061922
CEC4521	sup	France	>256	0.064	1327	11	176	14126635
CEC4522	sup	France	>256	0.19	1327	11	84	6464431
CEC4524	sup	France	>256	0.38	1327	11	97	7822930
CEC4527	sup	France	8	0.047	2886	11	141	11859035

3
 4 *Sup= superficial; NC = no cladable; FLC= fluconazole; CAS= caspofungin, DST= Diploid Sequence*
 5 *Type.*

- 1 **Table S3: *C. albicans* clinical strains genotypes for the 5 genes analyzed.** For each
- 2 polymorphism position (non-synonymous) the genotype obtained for the different clinical
- 3 strains (n=161) was noted when it was different than the SC5314 sequenced reference strain.
- 4 h=heterozygous mutation in a single allele. (Full excel file is available)

		Strain name and genotype													
		SC5314													
Gene name	ORF position	Amino Acid substitution position	CEC001	CEC718	CEC723	CEC4504	CEC4503	CEC4515	CEC4516	CEC4521	CEC4522	CEC4524	CEC4527	CEC3658	CEC4259
TAC1	261	K87N	TT												
TAC1	268	A90T ^h	CC												
TAC1	310	F104V	AC	CC	CC			CC	CC	CC	CC	CC	AA		
TAC1	323	S108N	CC												
TAC1	391	L131I	GG												
TAC1	445	L149I ^h	AA												
TAC1	508	M170V	TT												
TAC1	510	M170I	CC												
TAC1	520	N174D	TT												
TAC1	566	F189S	AA											GG	
TAC1	593	H198R ^h	TT												
TAC1	596	S199N	CT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT		
TAC1	617	R206H	CT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT		
TAC1	620	V207A	AG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG		
TAC1	628	A210S ^h	CC												
TAC1	787	H263Y	GG			AA									
TAC1	1121	T374I ^h	GG												
TAC1	1130	A377V ^h	GG												
TAC1	1187	N396S	TC	CC	CC	CC	CC	CC	CC	CC	CC	CC	CC		
TAC1	1924	T642A ^h	TT												
TAC1	1994	F665S ^h	AA												
TAC1	2083	L695F	GG												
TAC1	2123	K708T ^h	TT												
TAC1	2140	S714G ^h	TT												
TAC1	2206	A736T	CC	TT										TT	
TAC1	2276	L759S ^h	AA												
TAC1	2291	S764N ^h	CC												
TAC1	2298	K766N ^h	TT												
TAC1	2316	N772K	AT	TT	AA	AA	AA	AA	AA	AA	AA	AA	TT		
TAC1	2326	D776N	CT	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT		
TAC1	2369	A790V	GG		AA										
TAC1	2344	T782A ^h	TT												
TAC1	2469	N823K	AA											CC	
TAC1	2485	E829Q	CG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	GG	
TAC1	2495	F832S ^h	AA												
TAC1	2515	H839Y	GG			AA	AA								
TAC1	2573	C858Y	CC			TT	TT							TT	
TAC1	2606	R869Q ^h	CC												
TAC1	2684	I895T	AA												
TAC1	2687	N896S	TT		CC			CC	CC	CC	CC	CC			
TAC1	2723	C908Y	CC												
TAC1	2804	S935L	GA	AA	GG	GG	GG	GG	GG	GG	GG	GG	GG		
TAC1	2810	S937L ^h	GG												
TAC1	2822	S941P	AG	GG	GG	AA	AA	GG	GG	GG	GG	GG	AA		
TAC1	2911	P971S	GG					AA	AA	AA	AA	AA			
ERG11	26	D9V ^h	TT				TA								
ERG11	88	P30S ^h	GG		GA										
ERG11	49	L17F ^h	GG												
ERG11	125	S42L ^h	GG												
ERG11	190	Y64H	AA									GG			
ERG11	348	D116E	AT	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA		
ERG11	383	K128T	TG	TT	TT	TT	TT	TT	TT	TT	TT	TT	TT		
ERG11	368	T123I	GG										AA		
ERG11	394	Y132H	AA	GG				GG	GG	GG	GG	GG			
ERG11	386	G129A ^h	CC												
ERG11	433	F145I ^h	AA												
ERG11	459	D153E	AA												
ERG11	470	R157K	CC			CT	TT								
ERG11	475	V159I ^h	CC												
ERG11	488	R163K ^h	CC												
ERG11	632	R211K	CC												
ERG11	706	P236S	GG	AA											
ERG11	798	E266D	TT												
ERG11	886	E296* ^h	CC												
ERG11	982	H328Y	GG												
ERG11	1007	E336G ^h	TT												
ERG11	1008	E336D ^h	TT												
ERG11	1214	S405F	GG	AA	AA										
ERG11	1217	P406L ^h	GG												
ERG11	1309	V437I	CC		TT										
ERG11	1325	S442F ^h	GG												
ERG11	1343	G448E	CC											TT	
ERG11	1349	G450E	CC					TT	TT						
ERG11	1400	R467I	CC							AA	AA	AA			
ERG11	1447	I483V ^h	TT												
ERG11	1462	V488I	CC												
UPC2	73	L25F ^h	GG												
UPC2	76	K26E	TT												
UPC2	122	T41N ^h	GG												
UPC2	203	R68K	CC												
UPC2	425	I142S	AA												
UPC2	431	T144I ^h	GG												
UPC2	569	S190N	CC												
UPC2	661	A221P	CC												
UPC2	683	S228N	CC												
UPC2	749	N250S ^h	TT												
UPC2	817	T273A	TT												
UPC2	961	G321S ^h	CC												
UPC2	1024	L342V ^h	GG												
UPC2	1519	V507I ^h	CC												
UPC2	1904	L635W ^h	AA												
UPC2	1943	G648D ^h	CC					CT							
UPC2	1960	D654N	CC							CT	TT	CT			
UPC2	1970	G657D	CC	TT											

1 ...

2 **Table S4: List of the 22 amino acid substitutions selected by our tool as potentially**
3 **involved in antifungal resistance.** These positions were selected from the analysis of the 5
4 genes from the 10 resistant clinical isolates of *C. albicans* compared to the natural
5 polymorphism repertoire. Bold formatting indicates the new putative substitutions involved in
6 antifungal resistance. h=heterozygous mutation in a single allele.

Protein name	Amino acid substitution	Reference	
Tac1p	H263Y	New	In this study
Tac1p	A736T	Known	Coste <i>et al.</i> Eukaryotic cell 2009
Tac1p	A790V	New	In this study
Tac1p	N823K	New	In this study
Tac1p	H839Y	New	In this study
Tac1p	C858Y	New	In this study
Tac1p	P971S	New	In this study
Erg11p	D9V^h	New	In this study
Erg11p	P30S^h	New	In this study
Erg11p	Y64H	New	In this study
Erg11p	T123I	Known	YongQin Wu <i>et al.</i> FEMS Yeast Research 2017
Erg11p	Y132H	Known	Chau <i>et al.</i> , 2004; Favre <i>et al.</i> , 1999; Kakeya <i>et al.</i> , 2000; Marichal <i>et al.</i> , 1999; Sanglard <i>et al.</i> , 1998; Xu <i>et al.</i> , 2008
Erg11p	R157K	Known	Sanglard and Bille 2002
Erg11p	P236S	New	In this study
Erg11p	S405F	Known	Wang <i>et al.</i> BMC Mic. 2009
Erg11p	G448E	Known	Loffler <i>et al.</i> 1997
Erg11p	G450E	Known	Chau <i>et al.</i> , 2004; Favre <i>et al.</i> , 1999; Goldman <i>et al.</i> , 2004; Loffler <i>et al.</i> , 1997; Perea <i>et al.</i> , 2001
Erg11p	R467I	Known	Lamb <i>et al.</i> , 2000; Lee <i>et al.</i> , 2004; Sanglard <i>et al.</i> , 1998; White <i>et al.</i> , 1997
Upc2p	G648D ^h	Known	Flowers <i>et al.</i> AAC 2015
Upc2p	D654N	New	In this study
Upc2p	G657D	New	In this study
Fks1p	R647G	Known	M. Lackner <i>et al.</i> , AAC 2014

7

8

1 **Table S5: Strains and mutant constructions used in this study**

Strain Name	Genotype	Auxotrophies	Parental strain	References
SC5314	Reference strain	Prototroph		(1)
DSY4578	<i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri-	CAF4-2 (Isogenic strain construction and gene mapping in <i>Candida albicans</i> . Fonzi WA, Genetics. 1993).	Received from D. Sanglard's lab
CEC5216	<i>RPS1/RPS1::ClpSAT-P_{UPC2}-UPC2</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY4578	This study
CEC5218	<i>RPS1/RPS1::ClpSAT-P_{UPC2}-upc2^{G657D}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY4578	This study
CEC5220	<i>RPS1/RPS1::ClpSAT-P_{UPC2}-upc2^{D654N}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY4578	This study
CEC5221	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-TAC1</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY4578	This study
CEC5222	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-tac1^{A790V}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY4578	This study
CEC5224	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-tac1^{N823K}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY4578	This study
CEC5225	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-tac1^{C858Y}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>upc2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY4578	This study
DSY3984	<i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>UPC2::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri-	DSY2321	Received from D. Sanglard's lab
CEC5226	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-TAC1</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>UPC2Δ::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY3984	This study
CEC5228	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-tac1^{H263Y}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>UPC2/FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY3984	This study
CEC5230	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-tac1^{H839Y}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>UPC2::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY3984	This study
CEC5232	<i>RPS1/RPS1::ClpSAT-P_{TAC1}-tac1^{P971S}</i> <i>tac1Δ::FRT/tac1Δ::FRT</i> <i>mrr1Δ::FRT/mrr1Δ::FRT upc2Δ::FRT/</i> <i>UPC2::FRT ura3::hisG/ura3::hisG</i>	ARG+ HIS+ LEU+ uri- Nsn ^R	DSY3984	This study

2

3

1 **Table S6: Primers used in this study**

	Primer name	Sequence
Zinc cluster transcription factor Transcriptional Activator of CDR genes PCR amplification for cloning	TAC1-NcoI_F	cctggAGCAACAACAGCAGGAGTACC
	TAC1-SbfI_R	cctgcaggTGCTCTACCCATAATGTATCC
Zinc cluster transcription factor Transcriptional Activator of CDR genes PCR and Sequencing	TAC1_P1_F	ATAGAGCCTTTCTCCTTCTC
	TAC1_P2_F	TTCTCTCGGGAATCATTAG
	TAC1_P3_F	GAATCCAATTGGTTGACGAG
	TAC1_P4_F	TGATTGCTACCGAAGATTCC
	TAC1_P4_R	CGCTTTCACCAATTACAACCTC
	TAC1_P3_R	GGAATCTTCGGTAGCAATCA
	TAC1_P2_R	ACATCAACAATGCTTCTAC
	TAC1_P1_R	TCTTCACCGTATGAACCTA
Zinc cluster transcription factor Transcriptional Activator of CDR genes Directed mutagenesis	TAC1_H263Y_F	catgatatatggtttaactctgaatagaacactaaaggtaaattcctatgg
	TAC1_H263Y_R	ccataaggaattacctttagtgttctattcagaggttaaaccatatacatg
	TAC1_N823K_F	ttctctgagctctgtgaaatgctttgtaaagctggattattagc
	TAC1_N823K_R	gctaataatccagctttcacaagaagcatttacaagaagctcaagagaaa
	TAC1_H839Y_F	gacgtgaattccgcatagttttggaaaatcacagggaatttc
	TAC1_H839Y_R	gaaattccctgatattttccaaaactatcggaattcacgctc
	TAC1_C858Y_F	ttctgctaccataaccaccaaattgggaagcacca
	TAC1_C858Y_R	tggtgctcaccatttgggttatgtagcaggaa
	TAC1_P971S_F	ccaaattattgcaagaaaaaattggataaattataaattgactaaatggcaaattacagc
	TAC1_P971S_R	gctgtaaattgccatttagcaatttaaatatccaattttctttgacaataattgg
Zinc cluster transcription factor PCR amplification for cloning	UPC2_BspHI_F	tcatgaTGCTGAAAAGGTTTACCTTGACC
	UPC2_SbfI_R	cctgcaggCACACAGTAACGAATCACATTTG
Zinc cluster transcription factor PCR and Sequencing	UPC2_P1_F	CTGCAGAGAATCACAGTGAAGTTC
	UPC2_P2_F	CTCAGCCGGTGATTCTCCA
	UPC2_P3_F	CGGTCAAACCTCAATATGCTTGAC
	UPC2_P4_F	GTTTCCAGTGCTTTTGGACTCTCC
	UPC2_P1_R	TGGAGGAATCACCGGCTGAG
	UPC2_P2_R	GTCAAGCATATTGAGGTTTGACCG
	UPC2_P3_R	GGAGAGTCCAAAAGCACTGGAAAC
	UPC2_P4_R	CCTATCATCTACGCGGTATTGACC
PCR amplification to check the integration	ClpSAT_checkF	ATACTACTGAAAATTTCTGACTTTC
	ClpSAT_checkR	CTAGTGGATCCCCAGATCATTATCC
ABC drug efflux transporter gene 2 expression analysis	CDR2-RT-F	CTAGATATTTGAGCCACATG
	CDR2-RT-R	ATCATTAGCAATACCGTAAG
	CDR2-RT-P	TCCATTCAACGGCAACATTAGTCA
Lanosterol 14-alpha demethylase gene expression analysis	ERG11-RT-F	ATTCTTATGGGTGGTCAA
	ERG11-RT-R	CAACAATTCAACAACCTTCTT
	ERG11-RT-P	ACTTCTGCTTCTACTTCTGCTTGG
Actin 1 gene expression analysis	ACT1-RT-F	ATAACGGTTCTGGTATGT
	ACT1-RT-R	CCTTGATGTCTTGGTCTA
	ACT1-RT-P	CGGTGACGACGCTCCAAG

2