Loss of heterozygosity : its impact on generating and shaping genetic variations in human fungal pathogen candida albicans

Timea B. Marton

Université de Paris

Ecole doctorale BioSPC

ED 562 - Infectiologie Microbiologie

Unité de Biologie et Pathogénicité Fongiques, Institut Pasteur

Thèse de doctorat de Microbiologie

Timea B. MARTON

Loss of heterozygosity: Its impact on generating and shaping genetic variations in the human fungal pathogen *Candida albicans*

Dirigée par Christophe D'ENFERT et Mélanie LEGRAND

Présentée et soutenue publiquement à l'Institut Pasteur le 23 septembre 2020

Président du jury : SILAR, Philippe / Professeur / Université de Paris
Rapporteur : SELMECKI, Anna / Assistant Professor / University of Minnesota Medical School
Rapporteur : MALAGNAC, Fabienne / Professeure / Université Paris Saclay – CNRS – CEA
Examinateur : FISCHER, Gilles / Directeur de Recherche / Institut de Biologie Paris Seine
Directeur de thèse : D'ENFERT, Christophe / Professeur / Institut Pasteur- INRA
Co-encadrant de thèse : LEGRAND, Mélanie /Chargée de Recherche/ Institut Pasteur- INRA

INSTITUT PASTEUR

LabEx IBEID

INRA SCIENCE & IMPACT

Acknowledgements

I would first like to express my appreciation to the members of my thesis jury Fabienne Malagnac, Anna Selmecki, Gilles Fischer and Philippe Silar. Thank you for you input and for taking the time to be a part of these final steps of my thesis, your participation means a great deal to me.

I thank Christophe d'Enfert for giving me the opportunity to be a part of his research team for the past years and Mélanie Legrand for welcoming me from day one. Thank you both for accompanying me on this journey, sharing your tremendous knowledge and for giving me the freedom which permitted me to develop and lean from scientific questions. You have fueled my curiosity and have made this adventure enjoyable; because of you I will always cherish this venture.

These past years spent at Pasteur have truly been amazing, filled with support, laughter, and numerous new friendships. Indeed, this is thanks to each one of the past and present BFP labmates with whom I shared countless moments of happiness. In particular, thank you, Murielle for her endless encouragement and laughs; Natacha for our escapades in the basement of Francois Jacob, Sophie for sharing her passion for teaching and inviting me to participate, Corinne for her availability to my endless questions, because of you bioinformatic and bioanalysis are no longer cringeworthy, and, Iryna and Elodie because they best understood all the wild emotional states of a PhD student. No section is sufficient to fully express my gratitude to all short- and long-term members thus, please know that I am grateful for the numerous memories we shared together including, humoring of my blonde side and my enthusiasm for pink. I would also like to thank Pierre-Henri, Frédérique, Guilhem, Lucia and everyone located on the 4th floor of Fernbach, for their friendliness and positive attitude. A coveted workspace that I will miss!

A big thank you also goes out to the best of friends, located on both sides of the Atlantic, despite my regular social absence throughout this thesis you have always been there for me! Pauline, Monica and Mélanie, Paris is not be the same without you "la coloc pour toujours !". Cécile, you kept me sane throughout this adventure and made me relativize, thank you for always being there. Finally yet importantly, I thank my family, especially my parents, for their unconditional encouragement in pursuing all my goals, this has pushed me to succeed this far.

Thank you, you all contributed to making this thesis exceptionally memorable!

Summary

Candida albicans is an opportunistic human pathogen possessing a relatively heterozygous diploid genome that is highly tolerant to particular genomic rearrangements, namely loss-of-heterozygosity (LOH). These two genomic hallmarks are thought to participate in the successful adaptation to its fluctuating host-dependent environment. In this thesis, we **first** made an inventory of LOH events within a collection of clinical *C. albicans* isolates and highlighted that LOH are principally short in size, which permits to generate new allelic combinations while maintaining relatively high overall heterozygosity level. **Secondly**, we identified recessive lethal alleles which constrains the directionality of these LOH events, and demonstrated that such constrains could be alleviated by recombinogenic properties of repeat sequences. A major source of LOH are DNA double-strand breaks (DSB), which have been previously shown to primarily be repaired through homologous recombination-mediated repair pathways in *C. albicans*, resulting in various lengths of LOH. **Thirdly**, we focused on break-induced replication (BIR) in *C. albicans*, which leads to long-tract LOH, spanning numerous kilobases until the telomere. Through characterization of different genes involved in DNA repair, we found that BIR was unexpectedly associated with homozygosis tracts between the break site and the centromere caused by intrinsic features of BIR and the activity of the mismatch repair pathway on natural heterozygous positions. **Fourth**, because strains often display numerous concomitant genomic rearrangements upon use of a classical transformation protocol, we investigated the recently developed CRISPR-Cas9 protocol and showed that this method seems to reduce simultaneous rearrangements, namely aneuploidy events. Although LOH are pervasive and implicated in various aspect of *C. albicans* biology, including acquisition of adaptive phenotypes, little is known regarding genome-wide LOH dynamics. **Fifth**, using a molecular approach, we highlighted elevated levels of heterogeneity in terms of genome stability, LOH frequency, at several levels, inter-strain, intra-strain and inter-chromosomes in basal and stress conditions. The presence of such heterogeneity participates in generating a wide-spectrum of genome instability potential within a population, perhaps facilitating the generation of new allelic combinations, a powerful adaptive strategy in a predominantly clonal organism such as *C. albicans*, which is continuously exposed to an evolving environment.

Key words: *Candida albicans,* Loss-of-heterozygosity, Genome stability, Heterogeneity

Résumé

Le pathogène opportuniste de l'homme *Candida albicans*, possède un génome diploïde relativement hétérozygote et tolérant aux réarrangements génomiques, tels que les pertes d'hétérozygotie (LOH). Ces deux caractéristiques génomiques permettent à cet organisme de s'adapter rapidement à un environnement dynamique à l'intérieur de l'hôte. Au cours de cette thèse, nous avons **premièrement** construit un répertoire des évènements de LOH au sein d'une collection d'isolats cliniques de *C. albicans,* en remarquant que la plupart des LOH sont de petite taille, permettant ainsi de générer de nouvelles combinaisons alléliques tout en maintenant un niveau important d'hétérozygotie. **Deuxièmement**, nous avons identifié des allèles récessifs létaux qui limitent la directionnalité de ces évènements de LOH, et avons démontré que ces contraintes peuvent être atténuées par les propriétés recombinogènes de séquences répétées. Une source majeure de LOH sont les cassures double-brin d'ADN (DNA-DSB) qui ont été démontrées comme étant principalement réparées par des voies de réparation par recombinaison homologue (HR) chez *C. albicans*, générant des LOH de tailles diverses. **Troisièmement**, nous nous sommes intéressés au break-induced replication (BIR), un mécanisme qui engendre de longs évènements de LOH, pouvant s'étendre sur plusieurs kilobases jusqu'au télomère. Grâce à la caractérisation de différents gènes impliqués dans la réparation de l'ADN, nous avons constaté que le BIR est également associé de manière inattendue à des régions d'homozygotie entre le site de cassure et le centromère, qui peuvent s'expliquer par les caractéristiques intrinsèques du BIR et l'activité du mismatch repair sur des positions naturelles hétérozygotes. **Quatrièmement**, parce que le processus de transformation classique de cellules *C. albicans* engendre de multiples réarrangements génomiques, nous avons étudié le protocole récemment adapté chez *C. albicans* de CRISPR-Cas9 et montré que l'utilisation de cette méthode semble réduire les réarrangements simultanés, notamment les événements d'aneuploïdie. Bien que les LOH soient omniprésentes et impliquées dans divers aspects de la biologie de *C. albicans*, y compris l'acquisition de phénotypes adaptatifs, on sait peu de choses sur la dynamique des LOH à l'échelle du génome. **Cinquièmement**, en utilisant une approche moléculaire, nous avons mis en évidence des niveaux élevés d'hétérogénéité en termes de stabilité du génome, fréquence de LOH, à plusieurs niveaux, inter-souches, intra-souches et inter-chromosomes en conditions basales et de stress. La présence d'une telle hétérogénéité peut permettre de générer un large spectre de potentiel d'instabilité génomique au sein d'une population, facilitant éventuellement la génération de nouvelles combinaisons alléliques, une stratégie adaptative intéressante dans un organisme à prédominance clonale comme *C. albicans* qui est continuellement exposé à un environnement changeant.

Mots clés : *Candida albicans*, Pertes d'hétérozygoties, Stabilité génomique, Hétérogénéité

Table of contents

PREFACE **12**

INTRODUCTION **13**

A. *C. ALBICANS* IN THE CLINICAL WORLD 13
B. PHYLOGENY, POPULATION STRUCTURE: GETTING TO KNOW *C. ALBICANS* 15
C. (PARASEXUAL) LIFE CYCLE OF *C. ALBICANS* 17
D. GENOMIC FEATURES CONTRIBUTING TO GENETIC VARIATION IN *C. ALBICANS* 20
1. Chromosomes 20
i. Centromeres 21
ii. Telomeres 23
iii. Organization within the nucleus 25
2. Repetitive sequences and repeat regions 26
i. Subtelomeric regions (TLO and LTRs) 26
ii. Major Repeat Sequences (MRS) 28
iii. Ribosomal DNA locus 29
iv. Inverted repeats 30
3. Ploidy 31
4. Heterozygosity 33
E. GENOMIC PLASTICITY – GENOMIC REARRANGEMENTS 35
1. Translocations 35
2. Aneuploidies 37
3. Isochromosomes 39
4. LOH 41
F. TOOLS TO STUDY GENOME STABILITY IN *C. ALBICANS* 47
1. Tools to induce DNA lesions 47
i. Genotoxic compounds and physical stress 47
ii. Inducible DNA-DSB system 48
iii. CRISPR-Cas9 49
2. Tools for monitoring LOH 52
i. *GAL1*/*URA3* system 52
ii. BFP/GFP LOH reporter system 53
3. Tools used to characterize gross chromosomal rearrangement events 55
i. Whole-genome sequencing 55
ii. Characterization through auxotrophic spotting and fluorescent confirmation 56
iii. SNP-RFLP 57
G. DNA BREAKS AND DNA REPAIR MECHANISMS: SOURCES OF GENOMIC REARRANGEMENTS 59
1. Excision and mismatch repairs 60
i. Base excision repair (BER) 60
ii. Nucleotide excision repair (NER) 61
iii. Mismatch repair (MMR) 64
2. DNA-DSB repair 66
i. Non-homologous end-joining (NHEJ) 66
ii. Homologous recombination-dependent repair 68
a. Initiation of homologous recombination 68
b. Homologous recombination completion 70
b1. Single-strand annealing (SSA) 71
b2. Synthesis-dependent strand annealing (SDSA) 72
b3. Gene conversion (GC), with or without crossover 72
b4. Break-induced replication (BIR) 75
iii. No repair 77
H. BIOLOGICAL IMPLICATIONS OF GENOMIC INSTABILITY 78
1. Genomic rearrangements as a response to environmental factors 78
2. Genomic rearrangements and adaptation 80
I. OVERVIEW OF THE AIMS AND ACHIEVEMENTS OF THE PHD THESIS 82

RESULTS AND DISCUSSION **85**

Chapter 1:Patterns of loss-of-heterozygosity within the natural population of Candida albicans 85
Introduction 87
Results 89
Discussion 99
Material and methods 103
Supplementary material 105

Chapter 2: Study of the impact of recessive alleles and buffering role of repeated sequences in Candida albicans *111*
Introduction 113
Results 115
Discussion 123
Material and methods 128
Supplementary material 132
Supplementary methods 139
Chapter 3: Extension of homozygosis towards centromere upon DNA-DSB repair by break-induced replication in Candida albicans *142*
Introduction 144
Results 147
Discussion 160
Material and methods 165
Supplementary material 170
Chapter 4: The use of CRISPR-Cas9 to target homologous recombination limits transformation-induced genomic changes in Candida albicans *182*
Introduction 184
Results 186
Discussion 194
Materials and methods 195
Supplemental materials 204
Chapter 5: How C. albicans chromosomes respond to stress exposure *220*
Introduction 222
Results 225
Discussion 235
Material and methods 238
Supplementary material 244
PERSPECTIVES **251**
BIBLIOGRAPHY **258**
APPENDIX **275**

Figures and Tables

Introduction:

Figure 1: Antifungal drug targets in *C. albicans*. 13
Figure 2: Bacterial dysbiosis can favor *C. albicans* candidemia. 14
Figure 3: Phylogeny of the Saccharomycotina sub-phylum. 15
Figure 4: Structure of the *C. albicans* population. 17
Figure 5: The mating type-like (MTL) locus and the parasexual life cycle of *C. albicans*. 19
Figure 6: Point and regional centromeres. 21
Figure 7: Telomeres 25
Figure 8: Major repeat sequences (MRS) of *C. albicans*. 29
Figure 9: Relationship between genomic heterozygosity and fitness in *C. albicans*. 34
Figure 10: Translocations. 36
Figure 11: Examples of aneuploidy events in a diploid background. 37
Figure 12: Isochromosomes involving Chr5 in *C. albicans* and inverted repeat (IR)-mediated mechanisms of isochromosome formation. 40
Figure 13: Density map of heterozygous single nucleotide polymorphisms (SNPs) across Chr3 of seven *C. albicans* strains. 42
Figure 14: Molecular mechanisms resulting in LOH events upon a DNA-DSB. 43
Figure 15: Rate and nature of LOH events in *C. albicans* vary between stresses. 44
Figure 16: Involvement of LOH in azole resistance. 46
Figure 17: An inducible DNA-DSB system coupled to the BFP/GFP LOH reporter system. 49
Figure 18: CRISPR-Cas9-mediated DNA double-strand break (DSB). 50
Figure 19: Application of the CRISPR-Cas9 system. 51
Figure 20: Monitoring LOH by counterselection. 53
Figure 21: The BFP/GFP LOH reporter system. 54
Figure 22: Identification of gross chromosomal rearrangements using sequencing data. 56
Figure 23: Identification of allelic status by SNP-RFLP. 58
Figure 24: Base excision repair (BER). 61
Figure 25: Nucleotide excision repair (NER). 63
Figure 26: Mismatch repair (MMR). 66
Figure 27: Non-homologous end joining (NHEJ). 67
Figure 28: Initiation of homologous recombination during DNA-DSB repair 70
Figure 29: Homologous recombination-mediated DNA-DSB repair. 71
Figure 30: Gene conversion with crossover during G2 phase of cell cycle can lead to long tract LOH events. 74

Chapter 1: Patterns of Loss-of-heterozygosity within the natural population of *Candida albicans*

Figure 1: Phylogeny representing distance of nucleic acid sequences between the 54 genetic cluster 1 *C. albicans* strains. 90
Figure 2: The root genome, an artificial reference genome built for analysis of LOH events in genetic cluster 1 of the *C. albicans* population. 92
Figure 3: Criteria used to identify LOH events throughout the 54 *C. albicans* strains of genetic cluster 1. 93
Figure 4: Characterization of LOH events within genetic Cluster 1 of the *C. albicans* population. 95
Figure 5: Characterization of distal LOH events within genetic Cluster 1 of the *C. albicans* population. 97
Figure 6: Distribution of LOH in *C. albicans* genetic cluster 1. 98
Figure S1: Distribution of LOH size by chromosomal location identified within the 54 *C. albicans* isolates of Cluster 1. 105
Figure S2: Correlation between the number of open reading frames (ORFs) per chromosome and the median number of LOH events per chromosome. 106
Figure S3: Correlation between the average number of heterozygous SNPs per kilobase and the cumulative LOH size across the 54 genome of *C. albicans* isolates. 107
Figure S4: Correlations of distal LOH and chromosome arm parameters. 108

Table S1: Strains analyzed from genetic cluster 1 of the *C. albicans* population 109
Table S2: Candidate hotspot genomic regions of LOH in cluster 1 of *C. albicans* 110

Chapter 2: Study of the impact of recessive alleles and buffering role of repeated sequences in *Candida albicans*

Figure 1: Coupling of a DNA double-strand-break-inducing system and a FACS-optimized LOH reporter system on Chr7. 117
Figure 2: Complementation with a functional MTR4 allele restores viability of Chr7B homozygous cells. 121
Figure 3: Upon DNA-DSB, Major Repeat Sequences are a source of inter-homolog recombination. 123
Figure 4: Schematic representation of the localization of premature STOP-inducing SNPs in the *C. albicans* SC5314 genome. 125
Figure S1: Overview of Chr7 in engineered strains throughout this study. 132
Figure S2: General characterization of strains. 133

Table 1: Summary of chromosome homozygosis bias observed in the literature……..124
Table 2: Genome-wide RLA candidates in the *C. albicans* strain SC5314……..126
Table S1: *C. albicans* strains used in this study……134
Table S2: Plasmids used in this study……136
Table S3: Primers used in this study. ……137
Table S4: List of heterozygous SNPs inducing premature STOP codons in the *C. albicans* SC5314 genome……..138

Chapter 3: Extension of homozygosis towards centromere upon DNA-DSB repair by break-induced replication in *Candida albicans*

Figure 1: Allele balance at heterozygous position plots for wild type *C. albicans* strains of this study. 148
Figure 2: Results of sequencing data analysis at DNA-DSB site in cells having undergone BIR/MCO in WT genetic background. 149
Figure 3: Characterization of DNA repair KO mutants. 151
Figure 4: Quantification of I-SceI-induced long-tract LOH events in WT and KO strains. 154
Figure 5: Analysis of molecular mechanism used in I-SceI induced conditions. 155
Figure 6: Distribution of genome-wide gross chromosomal rearrangements in wild-type and deletion strains. 156
Figure 7: Results of sequencing data analysis at DNA-DSB site in cells having undergone BIR/MCO. 158
Figure 8: Comparison of ascending LOH-tracts towards CEN4 across genetic backgrounds. 159
Figure S1: Coupling of a DNA double-strand-break-inducing system and a FACS-optimized LOH reporter system on Chr4. 171
Figure S2: Characterization of monofluorescent-sorted individuals leading to the identification of a molecular mechanism. 172
Figure S3: Doubling time of *C. albicans* strains. 173
Figure S4: Filamentation assay..... 174
Figure S5: Allele balance at heterozygous site plots for Chr4 for all knock-out *C. albicans* strains sequenced in this study..... 175
Figure S6: Flow cytometry profile illustrating genomic stability of wild type and knock-out strains. .. 176

Table 1: Fold changes of frequencies of 5-FOA^R colonies obtained between I-*Sce*I induced and non-induced conditions……..153
Table S1: Sequencing depth on chromosome 4 of sequenced isolates displaying I-*Sce*I-dependent BIR/MCO……177
Table S2: Summary of gross chromosomal rearrangements (GCR) within sequenced isolates displaying I-*Sce*I-dependent BIR/MCO……..178
Table S3: Strains used in this study……179
Table S4: List of primers used throughout this study……180

Chapter 4: The use of CRISPR-Cas9 to target homologous recombination limits transformation-induced genomic changes in *Candida albicans*

Figure 1: CRISPR-Cas9-free and -dependent transformation strategies used for strain construction. 187
Figure 2: Doubling times in YPD at 30°C. 188

Figure 3: Identification of GCRs within the 57 *C. albicans* sequenced strains. 190
Figure 4: Quantification of GCRs identified within *C. albicans* strains, comparing the two transformation methods. 192
Figure 5: Nature of GCRs identified within *C. albicans* strains. 194

Table S1: *C. albicans* strains used in this study……205
Table S2: *C. albicans* plasmids used in this study……210
Table S3: List of primers……211
Table S4: Sequencing depth of strains……217
Table S5: Analysis of average allele balance at heterozygous site (ABHet) values……218
Table S6: Supplemental table with identification and characterization of GCRs in *C. albicans* strains whose genome has been sequenced……219

Chapter 5: How *C. albicans* chromosomes respond to stress exposure

Figure 1: Blue-Green collection..... 226
Figure 2: Median basal LOH frequencies across 14 *C. albicans* genomic loci..... 228
Figure 3: Basal LOH frequencies of two independently constructed strains. 229
Figure 4: Inter-strain heterogeneity of basal LOH frequency..... 231
Figure 5: LOH frequency of Blue-Green collection in growth conditions mimicking physiological stresses..... 232
Figure 6: Effect of hydrogen peroxide of the LOH frequency of *C. albicans*..... 233
Figure 7: Assessing LOH frequency of two distinct *C. albicans* chromosomes within a single strain 234
Figure 8: Fluctuation analysis allowing to determine an accurate LOH frequency. 241
Figure S1: Doubling time of *C. albicans* strains used within this study. 244

Table S1: Fold changes of monofluorescent cells upon stress exposure permitting to validate the functionality of the BFP/GFP LOH reporter systems within *C. albicans* strains used in this study ……245
Table S2: List of strains and plasmids used throughout this study……247
Table S3: Primers used throughout this study……249

Abbreviations

2
2-DG: 2-deoxygalactose
2-DG^{R}: 2-deoxygalactose resistant
5
5-FOA: 5-fluoroorotic acid
5-FOA^{R}: 5-fluoroorotic acid resistant
A
ABHet: Allele balance at heterozygous site
ALT: Alternative lengthening of telomeres
alt-NHEJ: Alternative non-homologous end joining
AP: Apurinic/Apyrimidinic
B
BER: Base excision repair
BFP: Blue fluorescent protein
BIR: Break-induced replication
BTS: Bermuda triangle sequence
C
CCL: Concerted chromosome loss
CDE: Centromeric DNA element
CEN: Centromere
CFU: Colony forming unit
CGH: Comparative genome hybridization
Chr: Chromosome
CL: Chromosome loss
CNV: Copy number variation
CO: Crossover
CRISPR: Clustered regularly interspaced short palindromic repeat
CT: Chromosome truncation
D
dCas9: Dead Cas9,
dHJ: Double Holliday junction
D-loop: Displacement loop
DNA: Deoxyribonucleic acid
dNTP: Deoxyribonucleoside triphosphate
DSB: Double-strand break
dsDNA: Double-Strand DNA
E
EMS: Ethyl methanesulfonate
G
GC: Gene conversion
GC w CO: Gene conversion with crossover
GCR: Gross chromosomal rearrangement
GFP: Green fluorescent protein
GG-NER: Global genome - nucleotide excision repair
GI: Gastrointestinal
GIT: Gastrointestinal tract
GOF: Gain-of-function
GWAS: Genome-wide association studies
H
hDNA: Heteroduplex DNA
HJ: Holliday junction
HO: Homothallic switching endonuclease
HR: Homologous recombination
HU: Hydroxyurea

I
IR: Inverted repeats
K
KO: Knock-out
L
LOH: Loss of heterozygosity
LP-BER: Long-patch base excision repair
LTR: Long terminal repeat
M
MCO: Mitotic crossover
MIC: Minimal inhibitory concentration
MLST: Multilocus sequence typing
MMBIR: Micro-homology mediated break-induced replication
MMEJ: Microhomology end joining
MMR: Mismatch repair
MMS: Methyl methanesulfonate
MRS: Major repeat sequence
MTL: Mating type-like
N
NCO: Non-crossover
NER: Nucleotide excision repair
NHEJ: Non-homologous end joining
O
OE: Overexpression
ORF: Open reading frame
P
PAM: Protospacer-adjacent motif
PCR: Polymerase chain reaction
R
rDNA: Ribosomal DNA
RE: Restriction enzyme
RLA: Recessive lethal allele
ROS: Reactive oxygen species
S
SDSA: Synthesis-dependent strand annealing
sgRNA: Single guide RNA
SNP: Single nucleotide polymorphism
SNP-RFLP: Single nucleotide polymorphism – restriction fragment length polymorphism
SNV: Single nucleotide variation
SP-BER: Short-patch base excision repair
SSA: Single-strand annealing
SSB: Single-strand breaks
ssDNA: Single-strand DNA
T
TALEN: Transcription activator-like effector nucleases
tBHP: Tert-butyl hydroperoxide
TC-NER: Transcription-coupled - nucleotide excision repair
TLO: Telomere associated
TRE: TLO recombination element
U
UV: Ultraviolet
W
WGD: Whole-genome duplication
WGS: Whole-genome sequencing
WT: Wild-type

List of publications

d'Enfert C, Bougnoux ME, Feri A, Legrand M, Loll-Krippleber R, **Marton T,** Maufrais C, Ropars J, Sertour N and Sitterlé E. (2017). Chapter genome diversity and dynamics in *Candida albicans*. *Candida albicans*: cellular and molecular biology. Springer. Prasad R.

Marton T, Feri A, Commere PH, Maufrais C, d'Enfert C, Legrand M. Identification of Recessive Lethal Alleles in the Diploid Genome of a *Candida albicans* Laboratory Strain Unveils a Potential Role of Repetitive Sequences in Buffering Their Deleterious Impact. mSphere Feb 2019, 4 (1) e00709-18; DOI: 10.1128/mSphere.00709-18

Marton T, Maufrais C, d'Enfert C, Legrand M. The use of CRISPR-Cas9 to target homologous recombination limits transformation-induced genomic rearrangements in *Candida albicans*. mSphere Sep 2020; 5 (5) e00620-20; DOI: 10.1128/mSphere.00620-20

Marton T, d'Enfert C, Legrand M. Multiple stochastic parameters influence genome dynamics in a heterozygous diploid eukaryotic model. In preparation, to be submitted 2020.

Marton T*, Chauvel M*, Feri A*, Maufrais C, D'Enfert C, Legrand M. Extension of homozygosis towards centromere upon DNA double-strand break repair by break-induced replication in *Candida albicans.* In preparation, To be submitted in Genetics-G3 in September 2020

Marton T, Maufrais C, Bougnoux M-E, D'Enfert C, Legrand M. Patterns of Loss-of-heterozygosity within the natural population of *Candida albicans.* In preparation

Preface

Eukaryotic genomes are naturally dynamic, as DNA content fluctuates during cell division via DNA replication and chromosome segregation. Additionally, all genomes of living organisms evolve over time through changes which may be influenced by internal or external factors. These factors may compromise the genomic integrity voluntarily or sporadically thus, each cell must be able to protect the integrity of its genome to ensure livelihood of the lineage by identifying and executing necessary processes. In general, four major mechanisms govern the maintenance of genomes: (i) high-fidelity DNA replication, (ii) precise segregation of chromosomes during cell division, (iii) coordination of cell cycle progression and (iv) error-free repair of DNA damages. The failure to properly regulate the latter mechanisms often leads to genomic instability where genetic alterations are more frequently encountered. Certainly, many genetic variations have negative consequences such as in humans where they can be associated to numerous genetic diseases (i.e. neurodegeneration, impaired development, or cancer predispositions). Though strict maintenance of the genetic material is crucial, genomic instability is frequently beneficial in unicellular organisms as it can promote rapid evolutionary adaptation to a new environment. In agriculture and food production, numerous strains of *Saccharomyces cerevisiae* are used for fermentation purposes, e.g. production of bread, wine, beer, or sake. Indeed, genome sequencing data revealed that *S. cerevisiae* used in the production of various fermented food process displayed many genomic rearrangements which allowed the adaptation to different substrates (Legras *et al.* 2018). Additionally, genome comparisons of several species of protozoa parasite genus *Leishmania* highlighted that various combinations of genomic rearrangements contribute to the highly adaptive lifestyle of this unicellular organism (Bussotti *et al.* 2018). Indeed, an equilibrium is required between genomic stability which ensures transmission of genetic content to the next generation and genomic plasticity promoting the acquisition on new genetic variations aiding adaptation to a progressive environment.

This fine-tuned balance is especially pivotal in pathogenic species which are often confronted to dynamic environments, continuously necessitating adaptation. Throughout this thesis, we will discuss genome plasticity in the human fungal pathogen *Candida albicans* by exploring which parameters play key roles in generating genetic variation and how these alterations participate to the success of this species, as both a commensal and a human opportunistic fungal pathogen.

Introduction

A. *Candida albicans* in the clinical world

To this day, fungi are much underappreciated human disease-causing agents responsible for a wide array of infections, ranging from superficial lesions to life-threatening invasive infections. The global burdens, associated with these two infection extremes, are estimated at roughly 1 billion cases of superficial mycosis and 1 million cases of invasive fungal infections (Brown *et al.* 2012; Bongomin *et al.* 2017; Papon *et al.* 2020). These multitudinous cases are usually caused by various species within the Aspergillus, Candida and Cryptococcus genera. Amongst the Candida genus, *Candida albicans*, *Candida dubliniensis, Candida parapsilosis* and more recently *Candida auris* are some examples of the causative agents of infections. Over 3 million cases of oral and oesophageal candidiasis, and approximately 700,000 cases of invasive candidiasis occur yearly worldwide (Bongomin *et al.* 2017). Additionally, it is estimated that 50-75% of female individuals will be affected by vulvovaginal candidiasis once in their lifetime, and 5-7% of them will suffer from recurrent vulvovaginal infections (i.e. roughly 75 million women) (Brown *et al.* 2012). However, fungal infections are often efficiently treated, despite the limited number of antifungal molecules available (major classes include azoles: targeting the ergosterol pathway, echinocandins: targeting glucans of the cell wall, polyenes: targeting ergosterol in the cell wall and fluoropyrimidine: inhibition of nucleic acid synthesis) (Figure 1).

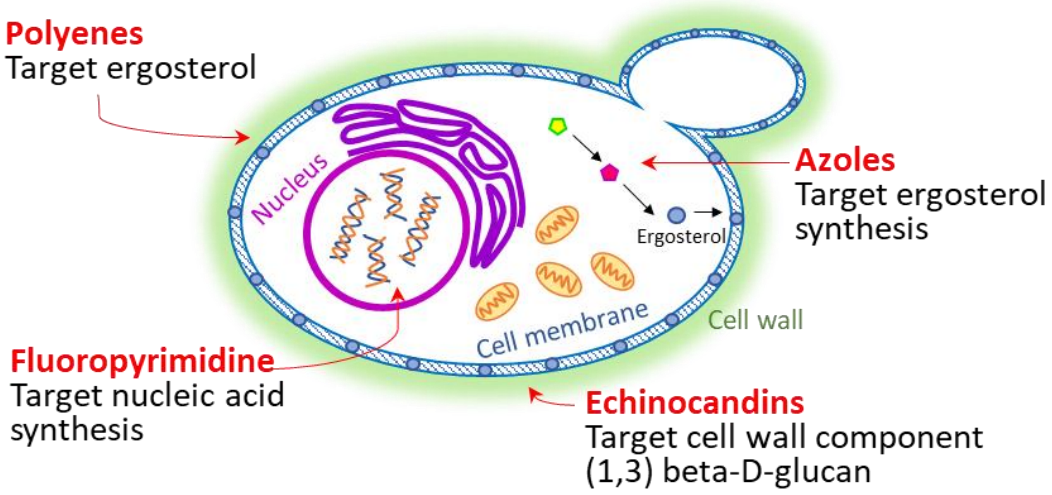

Figure 1: Antifungal drug targets in C. albicans.
Four major classes of antifungals are currently used in the clinical setting in order to treat *C. albicans* infections: Polyenes affects integrity of the cell membrane by targeting ergosterol (e.g. generating pores), fluoropyrimidine affect DNA replication and transcription by targeting nucleic acid synthesis, echinocandins affect integrity of cell wall by targeting the (1,3) beta-D-glucan synthase and azoles affect the cell membrane integrity/fluidity by targeting ergosterol synthesis.

Even though *C. albicans* can cause disease, it is most frequently found as a common commensal yeast of the human gastrointestinal (GI) and genital tracts. This opportunistic pathogenic yeast is polymorphic and can switch from the commensal to the pathogenic state upon alterations of the host environment (Papon *et al.* 2013). *C. albicans* is the dominant agent of infection in terms of incidence (Pfaller and Diekema 2010), within the Candida genus. Upon candidemia, despite the arsenal of antifungal molecules, an associated mortality rate of roughly 20-50% exists (Pappas *et al.* 2018). Individuals at risk of developing life threatening *C. albicans* infection are (i) highly immunocompromised individuals, (ii) individuals who have endured an invasive clinical procedure or (iii) those who have experienced major trauma and whose treatment requires extended stay in intensive care units (Brown *et al.* 2012). This is quite worrisome due to the increasing number of immunocompromised individuals within our evolving society, including transplant and cancer patients. A recent study has also shown that immunocompromised individuals, who develop candidemia are most often infected by the commensal strain in their gut. In a transplant setting, immunosuppression and prophylactic antifungal and antibiotic treatments promote dysbiosis in the GI tract allowing expansion of opportunistic pathogenic Candida species, which can translocate from the GI tract into the bloodstream leading to candidemia (Papon *et al.* 2020; Zhai *et al.* 2020) (Figure 2). This highlights the importance to better understand the biology and pathogenicity of fungal species in order to counterbalance the growing "at-risk" population, in inevitable contact with potentially pathogenic fungal species. Additionally, this emphasizes the importance of understanding, at a genetic level, how commensal Candida yeast, including *C. albicans,* can switch into a pathogen.

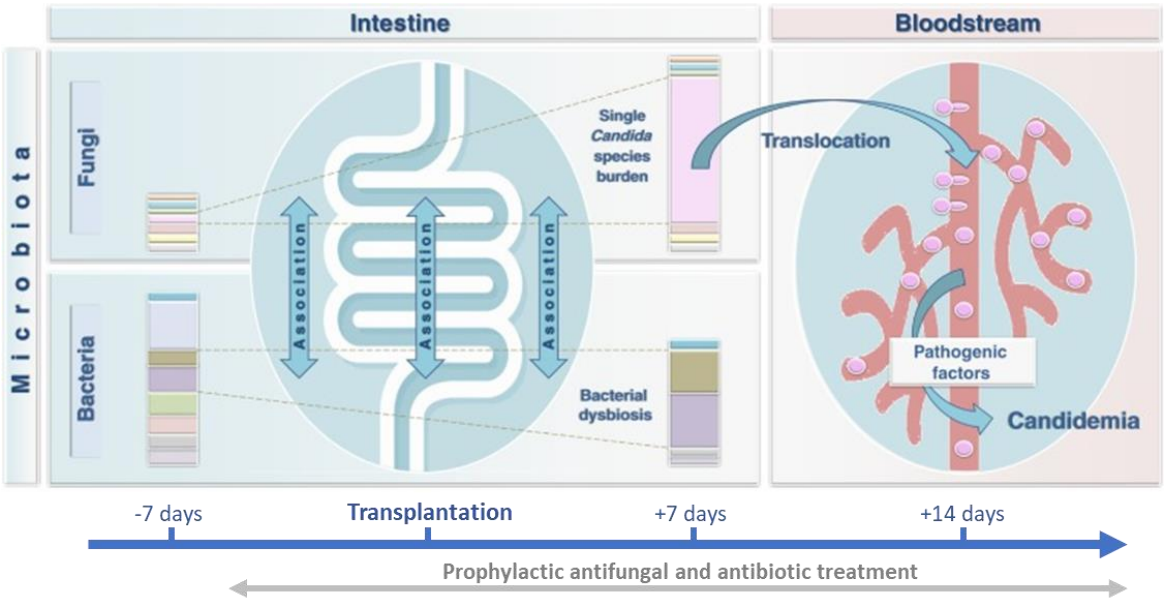

Figure 2: Bacterial dysbiosis can favor C. albicans candidemia.
Prophylactic treatment using an antifungal and antibiotic combination in transplant setting can lead to microbiota dysbiosis, favoring the translocation of resident commensal *Candida spp.* strains from the gut to the bloodstream, potentially promoting candidemia. Figure adapted from (Papon *et al.* 2020).

B. Phylogeny, population structure: getting to know *C. albicans*

Following the ancestry of *C. albicans,* this microorganism is found within the Fungi kingdom in the Ascomycota phylum, sub-phylum Saccharomycotina, which possesses the largest yeast diversity of the Ascomycota (more than 1500 yeast species) (Dujon 2010). Notably, the Saccharomycotina sub-phylum contains the order of Saccharomycetales, regrouping the budding yeasts, which is divided into 16 families. Amongst them is found the *Saccharomycetaceae* comprised of species which have undergone whole-genome duplication (WGD), including the research model organism *Saccharomyces cerevisiae*, also known as the baker's yeast. Additionally found within the Saccharomycotina, the *Saccharomycetales incertae sedis* family includes several medically relevant Candida species, such as *C. albicans, C. dubliniensis*, *C. parapsilosis and C. tropicalis.* More specifically, *C. albicans* is part of a subgroup known as the CTG clade. Within this clade, the member yeasts usually translate the CTG codon as serine instead of a leucine (Figure 3).

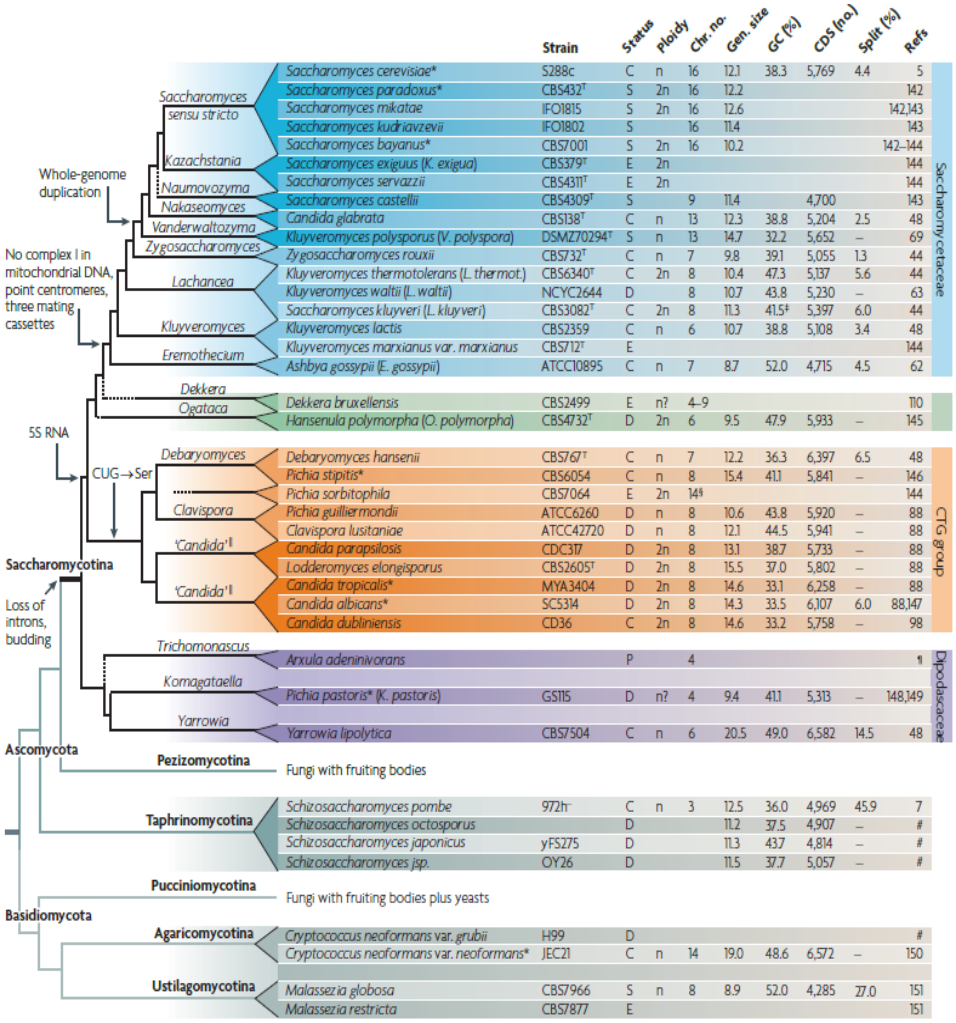

Figure 3: Phylogeny of the Saccharomycotina sub-phylum.
C. albicans is a member of the Saccharomycotina sub-phylum regrouped with other yeast species, these yeasts translate the CTG codon into a serine instead of a leucine amino acid base. Figure from (Dujon 2010).

At the level of the *C. albicans* population, multilocus sequence typing (MLST) (Bougnoux *et al.* 2002; Odds *et al.* 2007; Shin *et al.* 2011) of strains and more recently maximum likely-hood phylogenetic analysis, based on single nucleotide polymorphism (SNP) calling, conducted on a collection of 182 sequenced clinical *C. albicans* isolates, show that the population is divided into 18 genetic clusters. Between these genetic clusters, *C. albicans* strains are regrouped together originating from different sampling sites (commensal, superficial and invasive) and geographically distinct locations. However, strains tend to be regrouped by geographic origin within a cluster. Overall, no obvious biologically relevant particularity acts as a regrouping force of *C. albicans* strains, with two exceptions; clusters 1 and 13. Cluster 1 strains seem to exhibit a resistance phenotype to flucytosine (Pujol et al. 2004). While Cluster 13 strains (also known as *Candida africana*) are unable to use glucosamine or N-acetylglucosamine as carbon sources and do not form chlamydospores (Tietz *et al.* 1995), a particular cell morphotype characterized by its round, thick-wall, globular structure (Staib and Morschhäuser 2007). Additionally, *C. africana* strains have only been reported to cause genital infections, suggesting a niche restriction. In general, divergence between *C. albicans* clusters can be associated to the appearance of heterozygous SNPs, notably cluster-specific heterozygous SNPs, except for Cluster 13 (Figure 4). Genome sequencing of strains from Cluster 13 shows low levels of heterozygosity in comparison to the strains from other clusters. It is proposed that the ancestor strain of Cluster 13 underwent heavy loss of heterozygosity (LOH) thus diverging from the population. Additionally, within these LOH tracts, lower levels of heterozygous SNPs reappearance is observed amongst Cluster 13 strains suggesting the “recent” genetic cluster formation (Ropars *et al.* 2018) (Figure 4). Overall, genetic differences (or variants) between *C. albicans* strains are predominantly due to SNP and LOH events. However, no genome-wide association studies (GWAS) have been conducted linking genetic variations to specific phenotypic properties. Thus, it is still unclear what the biological implications of either cluster- or strain-specific genetic variants are. Overall, analysis of the *C. albicans* population structure has revealed a primarily clonal mode of reproduction although evidence for occasional gene flow between genetic clusters have been reported (Ropars *et al.* 2018; Wang *et al.* 2018) (Figure 4).

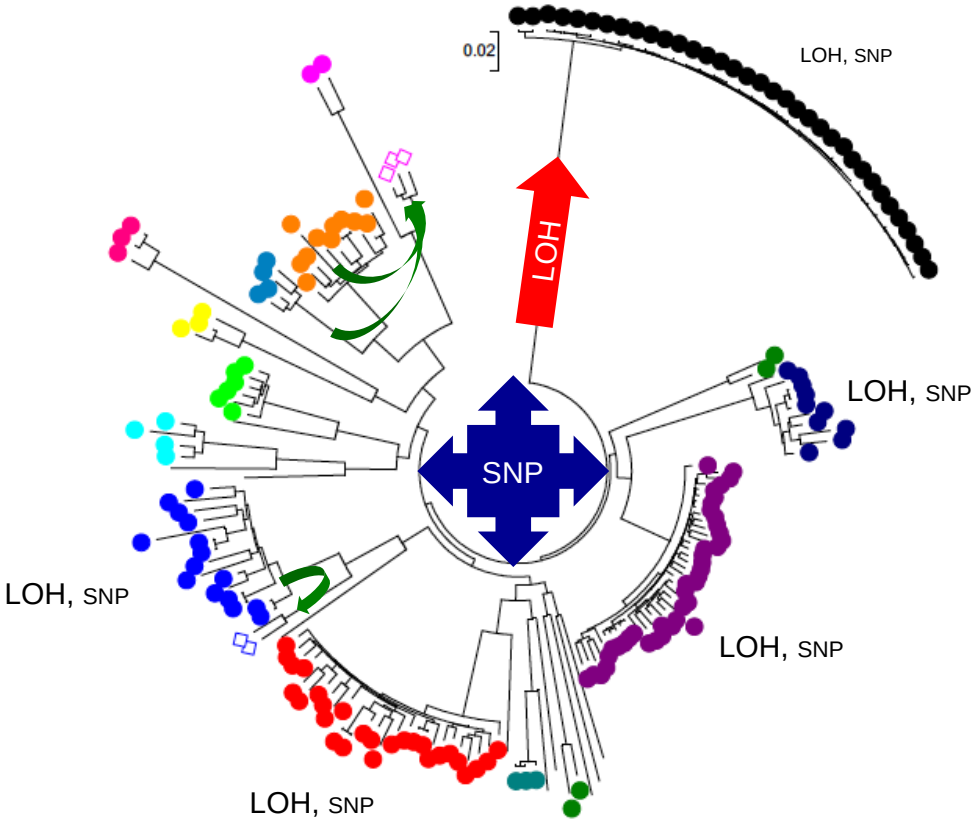

Figure 4: Structure of the C. albicans population.
The *C. albicans* population is distributed amongst 18 clusters (unique colors). The appearance of SNPs and LOH events has led to the diversification of the populations (central blue arrows and red arrow) while within-clade diversification is associated to both SNP and LOH events. The presence of admixture between genetic clusters is illustrated by green arrows, evidence of (para)sexuality.

C. (Parasexual) Life cycle of *C. albicans*

C. albicans is a polymorphic microorganism, meaning it has the capacity to switch between multiple yeast and hyphal morphologies depending on the environmental conditions. In agreement with the clonal nature of its population, *C. albicans* primarily reproduces asexually through budding. During the budding process, a genetically identical daughter cell will bud from a particular site of the mother cell through cell division, alternatively known as vegetative growth. Nevertheless, as stated above, evidences of admixture and potentially inter-clade hybrid strains within the natural population of *C. albicans* suggest the existence of a sexual reproductive cycle (Ropars *et al.* 2018; Wang *et al.* 2018) (green arrows in Figure 4). Comparable to the *MAT* locus in *S. cerevisiae, C. albicans* is heterothallic and possesses a mating type-like (*MTL*) locus with two idiotypes, *MTL*a and *MTLα,* located on chromosome 5 (Chr5). However, the regulatory network governing the reproductive cycle of *C. albicans* differs from that of *S. cerevisiae* (Dujon *et al.* 2004). In *C. albicans,* the locus encodes four mating-related transcriptional factors (*MTLa*1/*MTLa*2 or *MTLα*1/*MTLα*2) and 6 additional genes that are not involved in cellular identity (*PIKa* or *PIKα*: phosphatidylinositol kinase - *PAPa* or *PAPα*:

poly(A) polymerase and *OBPa* or *OBPα*: oxysterol binding protein) (Figure 5A). Additionally, sequences encoding for transcriptional regulators controlling expression of mating-type specific genes and elements of the pheromone response pathways exist in *C. albicans* (Hull and Johnson 1999; Tzung *et al.* 2001). Genes involved in mating appear to be under purifying selection, thus suggesting functionality (Zhang *et al.* 2015). Indeed, after mating of *C. albicans* cells was initially observed in both laboratory conditions and in a mammalian host (Hull *et al.* 2000; Magee and Magee 2000), a parasexual life cycle devoid of meiosis was subsequently described in *C. albicans* under laboratory conditions and genetic analysis of progenies demonstrates the occurrence of allele shuffling (Bennett and Johnson 2003; Forche *et al.* 2008; Hickman *et al.* 2013; Zhang *et al.* 2015).

The *MTL* locus in *C. albicans* strains is naturally heterozygous, where ~92% strains are *MTL*a/α (Odds *et al.* 2007). In order to enter the parasexual life cycle, *C. albicans* cells must first undergo LOH at the *MTL* locus, where *MTL*a/α cells become *MTL*a/a or *MTL*α/α, rendering cells mating competent (Lockhart *et al.* 2002). Moreover, mating is regulated by a phenotypic switch called the white to opaque switch (Miller and Johnson 2002). The switch from round sterile yeast cells, forming white dome colonies, to opaque fertile cells, characterized by darker yeast cells which form flatter colonies, is frequent and inheritable (Slutsky *et al.* 1985). The white-opaque switch is influenced by environmental factors mimicking the host environment such as, acidic pH, presence of carbon dioxide (CO_2) and N-acetylglucosamine (mono-saccharide primarily produced by GI tract bacteria), promoting the phenotypic switch (Huang *et al.* 2009, 2010; Sun *et al.* 2015). While host temperature (37°C) reduces the switch (Slutsky *et al.* 1985), CO_2 at physiological levels stabilizes the opaque phenotype at this temperature (Huang *et al.* 2009). Indeed, this suggests that mating may be favored in certain niches in the *in vivo* setting. Additionally, this epigenetic phenomenon is restrained by *MTL*a1 and *MTLα*2, which repress *WOR1*, the major regulator of this phenotypic switch. Indeed, *MTL*-homozygous cells undergo the phenotypic switch at a higher frequency, as expression of *WOR1* increases and is maintained by a positive feedback loop, favoring the phenotypic transition of white to opaque cells (Zordan *et al.* 2006). In fact, *MTL* homozygous opaque *C. albicans* cells are 10^6 times more efficient at mating than white cells (Miller and Johnson 2002). Opaque *MTL*-homozygous diploid *C. albicans* cells of opposite mating type can then undergo fusion resulting in a tetraploid cell. In a non-meiotic process named concerted chromosome loss (CCL), *C. albicans* tetraploids will lose chromosomes in order to return to a near-diploid state (Figure 5B) (Forche *et al.* 2008; Hickman *et al.* 2015). During CCL, in tetraploid cells, recombination events occur at a rate three times higher than during mitotic growth. Moreover, it has been shown that both the meiosis-specific Spo11 and Rec8 proteins play roles in DNA double-strand break (DSB) and recombination and, in promoting

chromosome instability during CCL, respectively. The CCL is thought to be a parameiotic process, perhaps an intermediate step of the mitosis to meiosis evolution (Anderson *et al.* 2019). *C. albicans* strains are most frequently found in the diploid state although haploid and tetraploid strains (further discussed in section D.3 Ploidy, p.31) have also been described. Haploid *C. albicans* cells are also mating competent, undergoing the white-opaque switch, cellular fusion and CCL to return to near haploid state, resulting in allele shuffled progenies (Figure 5B) (Hickman *et al.* 2013).

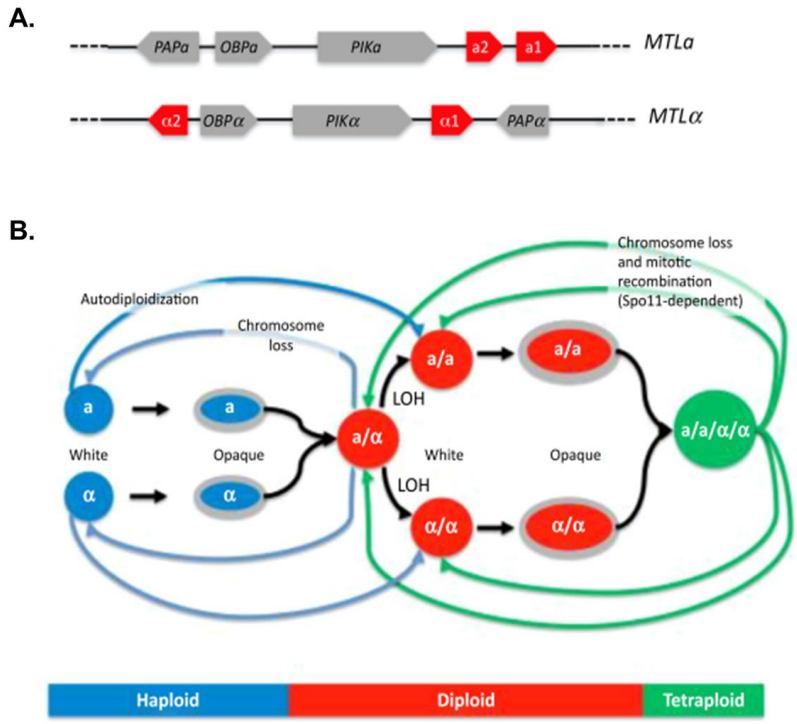

Figure 5: The mating type-like (MTL) locus and the parasexual life cycle of C. albicans.
A. Organization and composition of the *MTL* locus. **B.** Schematic view of the parasexual life cycle in *C. albicans.* Figure adapted from (Legrand *et al.* 2019)

C. albicans possesses and benefits from both asexual and (para)sexual life cycles, even though predominantly described as a clonal organism. Especially true for multicellular organisms, it is thought that asexual strains are evolutionary dead-end in nature though, within the fungal world, both types of life cycles have their advantages and inconveniences. Asexual fungal reproduction is energy efficient (compared to sexual reproduction), has no risk of genetic or organelle conflicts and maintains well adapted genomic conformations. However, asexual reproduction leads to low genetic diversity thus lowered adaptability to a fluctuating environment and the accumulation of random single nucleotide polymorphisms (SNPs) with a less efficient purging of deleterious mutations and selection of beneficial mutations. Although sexual reproduction is a lengthier and energetically costly process, which risks genetic conflicts between mates and may break apart well adapted genomic configurations, it is

commonly associated with a higher genetic diversity due to allele shuffling. This results into better removal of harmful mutations, efficient selection of beneficial mutations and better adaptation to a fluctuating environment (Sun and Heitman 2011). Overall, strictly referring to life cycle, an equilibrium between both life cycles is ideal in order to (i) generate new allelic combinations by mating or hybridization (mating between two varieties or species) (ii) while clonality can maintain and quickly expand within the population the advantageous genomic conformation. Nevertheless, this does not seem to be completely descriptive of the *C. albicans* population structure, due to mix of strain origins within described genetic clusters (Figure 4), thus suggesting the presence and the importance of other factors exercising a role on genetic variation.

D. Genomic features contributing to genetic variation in *C. albicans*

The genome is defined as the genetic material of an organism, a string of nucleic acids encoding for diverse interconnecting components allowing the cell to function and thrive. Genomes across the tree of life are highly heterogenous in terms of size, organization, and content, each one marked by evolution, driving them to constantly be modulated in order to continue to survive in their progressing environments. Indeed, an organism can benefit from genomic alterations but not all genomic modifications lead to positive outcomes thus, an equilibrium between acquisition of genomic variations and genomic stability is essential in order to allow the appearance of variants yet ensure the maintenance of genetic material and favorable genotypes. Sexual reproduction is not the only important contributor to the acquisition of genetic variation within a population, numerous genomic characteristics, including genomic content, organization, and copies, allow and/or promote recombination in both mitotic and meiotic contexts. Within clonal populations, these features are crucial for the longevity of the lineage. The following sections tackle the structure and organization of chromosomes, the various repeat sequences, ploidy, and heterozygosity found in the genome of *C. albicans* and how they contribute to generating and shaping genomic variations in this human fungal pathogen.

1. Chromosomes

The diploid genome of *C. albicans* has been sequenced and its composition has been described. It is a genome of roughly 16 Mb (haploid) which is poor in introns (6% intron-containing genes) and encodes for 6,198 open reading frames (ORFs), of which merely 27% have a characterized function. The genome is organized in 8 chromosome pairs numbered in

decreasing order of size, from Chr1 (3.2 Mb) to Chr7 (0.9 Mb). The 8th chromosome is named ChrR (2.2 Mb) because it possesses the ribosomal DNA repeats on its right arm (Jones *et al.* 2004; Hirakawa *et al.* 2015; Skrzypek *et al.* 2017). Each chromosome possesses a centromere, and a telomere at each chromosomal extremity.

i. Centromeres

Centromeres (CEN) are regions of chromosomes which play a role in spindle microtubules attachment, via the multiprotein complex named the kinetochore, permitting the correct segregation of replicated chromosomes during mitosis and meiosis. Centromere location is not only dictated by DNA sequence in eukaryotes but can also be associated to epigenetic markers (Buscaino *et al.* 2010). Two types of CENs have been described, “point” and “regional” CENs (Figure 6). Point CENs are defined by a specific and usually conserved DNA sequence composed of a short AT rich repeated sequence (roughly 125-400 bp) and are bound by proteins specifically recognizing this sequence of DNA. *S. cerevisiae* possesses point CENs composed of three centromeric DNA elements (CDE), each defined by specific sequences (Figure 6) (Westermann *et al.* 2007; Roy and Sanyal 2011). In contrast, regional centromeres, found in most organisms, are much larger (roughly 10-10 000 kb) and are epigenetically specified by the presence of a centromere-specific histone H3 variant, CENP-A. They are associated with various elements including: repeated DNA, inverted repeats, retroelements and unique DNA sequences (Buscaino *et al.* 2010).

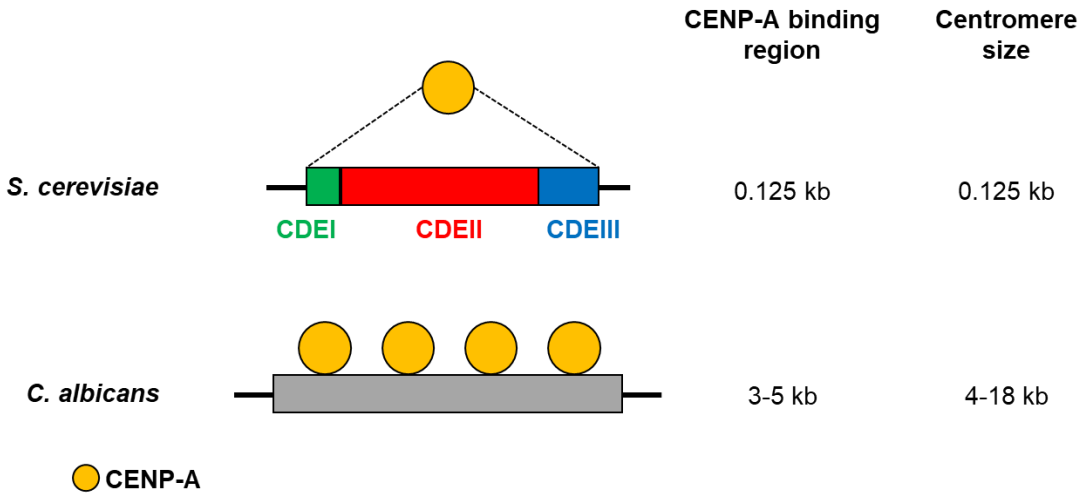

Figure 6: Point and regional centromeres.
A. Illustration of point centromeres composed of centromeric DNA elements (CDE): CDEI (8-bp [PuTCACPuTG]), CDEII (AT-rich region) and CDEIII (25-bp [TGTTT(T/A)TGNTTTCCGAAANNNAAAAA]). CENP-A recognizes and binds to the repeat DNA sequence. **B.** Representation of regional centromeres such as those found in *C. albicans*. These centromeres are of greater size and their localization is dictated by epigenetic marks recognized by CENP-A. Figure adapted from (Roy and Sanyal 2011).

In *C. albicans,* the identification and delimitation of centromeric regions were not directly described upon the advent of sequencing. Indeed, it is only upon identification of a homolog of CENP-A, *Ca*Cse4,(Sanyal and Carbon 2002) within the *C. albicans* genome, that researchers were then able to define and characterize its small regional CENs (Sanyal *et al.* 2004). Using chromatin immunoprecipitation of *Ca*Cse4, Mishra and collaborators (Mishra *et al.* 2007) have concluded that the CENP-A homolog only binds to centromeric regions in *C. albicans*. While the characterization of *Ca*Cse4 bound fragments determined that these CENs are composed of unique 3-4.5 kb long sequences found within gene-free regions (Figure 6). Pericentromeric regions of most *C. albicans* chromosomes show presence of inverted repeats (Chr1, 4, 5 and R) or long terminal repeats (Chr2, 3 and 6); only Chr7 presents no repeat sequences surrounding its CEN (Mishra *et al.* 2007; Ketel *et al.* 2009). Multiple other studies have suggested that CEN function and propagation are independent of DNA sequence. Construction of an artificial chromosome containing a CEN sequence and its integration into *C. albicans,* revealed that introduction of an artificial CEN DNA sequence is not enough to obtain a functional and mitotically stable CEN (Baum *et al.* 2006). Despite the heterogeneity of centromeric sequences between the 8 *C. albicans* chromosomes, centromeres form at the same position in divergent *C. albicans* isolates (Mishra *et al.* 2007) and even in the closely related species *C. dubliniensis* (Thakur and Sanyal 2013). This hints towards an epigenetic control of centromeres in terms of function and propagation. Chromatin status of repeats in pericentromeric regions of *C. albicans* is defined as intermediate chromatins states composed of interspersed euchromatin (high histone acetylation) and heterochromatin (hypomethylated) which show weak suppression of gene expression. This chromatin state is independent of Sir2 protein, a histone deacetylase allowing for recruitment and heterochromatin assembly. Thus, it is hypothesized that the partial silencing of genes in pericentromeric regions may be due to either (i) low levels of CENP-A binding, or (ii) the presence of the heterochromatin marks (H3K4 hypomethylation), both known to repress expression (Freire-Benéitez *et al.* 2016). Furthermore, CEN function is also relying on the proper functioning of the kinetochore. The knock-out (KO) of the crucial kinetochore protein Dam1 leads to disruption of the kinetochore architecture (Thakur and Sanyal 2012), resulting in CENP-A delocalization and degradation from centromeric chromatin (Roy and Sanyal 2011). Additionally, *RAD51-RAD52* proteins, involved homologous recombination (HR), mediate fork stalling at centromere proximal replication origins on *CEN7* thus, influencing CENP-A distribution levels across this regional centromere ultimately epigenetically maintaining centromere function (Mitra *et al.* 2014).

Upon deletion of *C. albicans* native CENs, neocentromeres will form within large intergenic regions flanked by repeated DNA sequences (Ketel *et al.* 2009; Thakur and Sanyal 2013; Burrack *et al.* 2016). Most often, proximal neocentromeres will form at <100 kb from the

native CEN but more distal neocentromeres (200-450kb from native CEN) have also been described in *C. albicans* (Ketel *et al.* 2009). However, as it was shown for Chr7, upon partial deletion of *CEN7*, the maintenance of the native centromere is favored through gene conversion (GC) at the CEN locus switching the deleted *CEN7* region to the native CEN and with gene conversion tracts encompassing the neocentromere, inactivating the neocentromere in favor of the native *CEN7*. Thereby, GC can act upon and suppress neocentromeres in order to maintain the native centromere positions conserved during evolution. GC at CENs may also explain the low levels of heterozygous SNPs observed across *C. albicans* CENs (Thakur and Sanyal 2013). The formation of neocentromeres in *C. albicans* emphasizes the important role of centromeres, ensuring the proper and complete division of genetic content between daughter cells thus allowing longevity of the lineage.

ii. Telomeres

Telomeres maintain genomic integrity through numerous cell divisions by promoting chromosome end replication and protecting chromosome ends from degradation (capping), end-to-end fusion and irregular recombination. These repetitive sequences are composed of repeat units which are usually 6 bp GT rich motifs. Fungi, including yeasts, show significant diversity in terms of repeats composition, complexity and length. Indeed, *Candida spp.* possess highly divergent telomere repeats, as repeat units are unusually long and not particularly G-rich, although within this sequence they all have shared 6 bp GT motif rich in guanine. *C. albicans* telomeres (500 bp - 5 kb) are composed of 23 bp-long 5'-ACTTCTTGGTGTACGGATGTCTA-3' repeat sequences (underlined 6 bp GT motifs) (Shampay *et al.* 1984; McEachern and Hicks 1993). At the very end of telomeres, the guanine rich DNA strand is often longer than its complementary strand, thus forming a 3' single-strand overhang, termed G-tail (Figure 7A). Telomere length is maintained by a ribonucleoprotein reverse transcriptase, named telomerase. Its catalytic protein subunit TERT extends telomere repeats using its RNA component (TERC) as a template to bind the G-tail (Nugent and Lundblad 1998; Hsu *et al.* 2007). Preservation of the G-tail, upon multiple rounds of G-tail extension by the telomerase, also permits the formation of protective T-loops at chromosome ends, shielding it from deterioration and end-to-end fusion. T-loops are formed upon the invasion of the 3' single-strand overhang into a duplex telomere region (Figure 7A). Furthermore, *C. albicans* telomerase core components also possess protection roles against aberrant overhang accumulation, independently of its catalytic activity (Hsu *et al.* 2007). *C. albicans* telomere structure and maintenance are reviewed in (Yu 2012).

Additionally, telomere length has been shown to be regulated independently of telomerase activity, in a process named alternative lengthening of telomeres (ALT), commonly thru recombination events. These events are facilitated by the repetitive nature of telomeres. For example, recombination-dependent events, such as GC and Break-Induced Replication (BIR) (further discussed in section G.2 DNA-DSB repair, p.66), have been observed intra-telomere or between chromosomes, in telomerase defective cells. This allows the addition of telomere repeat units (Lundblad and Blackburn 1993; Chen *et al.* 2001; Zhang *et al.* 2019a) in yeast and human cell lines (Figure 7B). Telomere length may also be regulated by extrachromosome telomere fragments, named T-circles. T-circles, circular telomere repeat DNA fragments, are generated by either an intra-telomere recombination event or by the excision of a T-loop (Figure 7A). Indeed, recombination events between T-circles and shorter telomeres lead to telomere repeat elongation (Figure 7B) (Bhattacharyya and Lustig 2006; Tomaska *et al.* 2009). In *C. albicans,* homozygous mutant strain of *RAD52* exhibited heterogenous and longer telomere lengths as compared to wildtype background, suggesting the occurrence of telomere recombination events (Ciudad *et al.* 2004). Additionally, homozygous mutant strain of *KU70* homologue displays accumulation of G-tail and T-circles, phenotypes which can be associated to telomere dysfunction (Chico *et al.* 2011). Overall, ATL has not specifically been investigated in *C. albicans* though upon KO of *C. albicans* gene homologues involved in ATL and telomere characterization data suggests presence of ATL in this opportunistic yeast.

Overall, telomeres contribute to the protection of chromosomal DNA content. Multiple rounds of DNA replication during cellular divisions result in shortening of telomere length. Telomere shortening can drive to cellular senescence, the irreversible cell cycle arrest. Thus, the maintenance of telomeres through regulation of their length is important for genome stability and cellular lifespan. Telomere length maintenance is predominantly and efficiently regulated by the telomerase. However, as previously mentioned, multiple recombination mediated ALT events may rescue telomerase deficient cells (Figure 7B). Furthermore, telomere structure (T-loops) (Figure 7A) and telomerase machinery also play a role in stabilization and protection of telomeres. Uncapped telomeres can result in chromosome end-to-end fusion giving rise to dicentric chromosomes (Figure 7B). Moreover, *C. albicans* telomeres are also associated with transcriptionally repressive heterochromatin, dependent on Sir2-mediated deacetylation of histones. However, the heterochromatin status at telomere can vary between weak and strong based on the environment (temperature 30°C vs. 39°C) influencing the expression of telomere proximal genes (Freire-Beneitez *et al.* 2016). This multi-factorial protection of DNA content is essential for any cell, giving it a chance to further thrive and/or adapt to its environment.

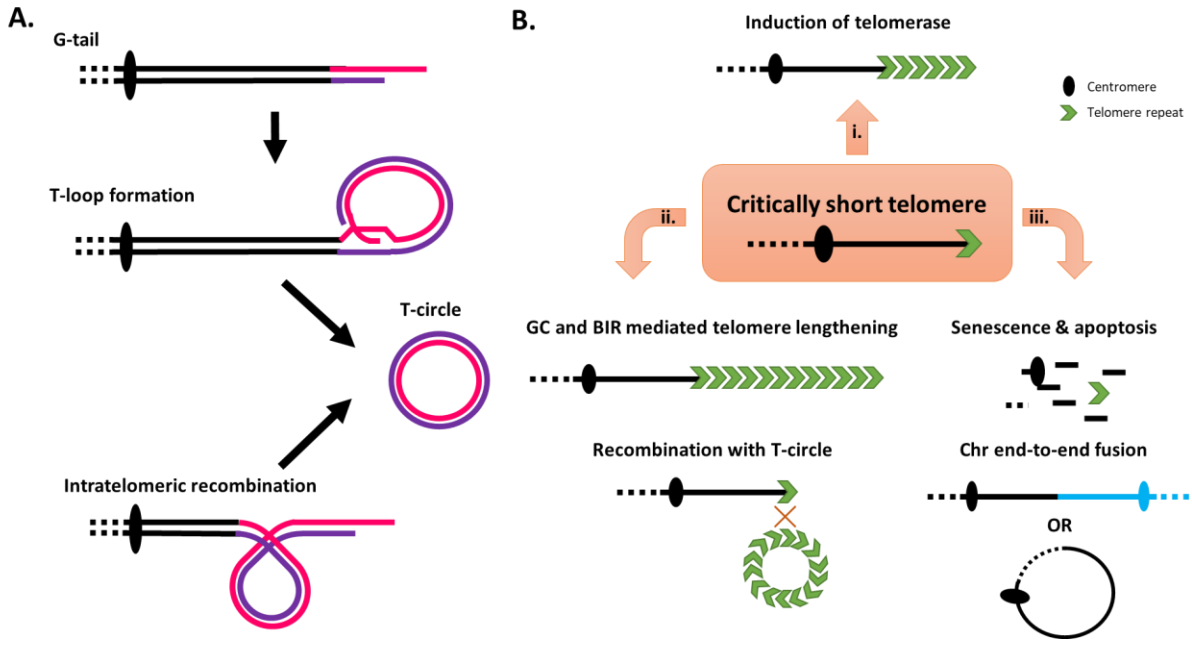

Figure 7: Telomeres
A. Representation of telomeric end structures and formation of T-circles. **B.** Multiple cellular fates arising from critically short telomeres, i. maintenance of telomere length in telomerase-proficient cells, ii. maintenance of telomere length via recombination (most frequently in telomerase-deficient cells), iii. uncapping of chromosomes resulting in apoptosis or end-to-end fusion of chromosomes. Figure inspired from (Bhattacharyya and Lustig 2006; Tomaska *et al.* 2009)

iii. Organization within the nucleus

The complete genome sequence of a eukaryotic cell, often spanning hundreds of thousands of bases, fits within its nucleus due to DNA packaging, where the DNA molecule is wrapped around proteins called nucleosomes which are coiled into chromatin fibers and these are further condensed into chromosomes. Within the nucleus, though initially thought that chromosomes where randomly organized microscopy experiments have shown that CENs tend to cluster together. The recent development of chromosome capture experiments is revealing the importance and the consequences of special genome architecture. In the nucleus, *C. albicans* centromeres are clustered together near the spindle pole bodies, the microtubule-organizing center, permitting physical interactions between chromosome arms (Burrack *et al.* 2016; Yadav *et al.* 2018). This spatial organization may also influence the determination of centromere regions, where a CENP-A rich nuclear region may exist and favor the formation of centromeres (Anderson *et al.* 2009; Thakur and Sanyal 2013). This CENP-A cloud could also explain the appearance of neocentromeres proximal to the native CEN locus (Burrack *et al.* 2016). Additionally, the clustering of centromeres could consequently also promote genetic mixing between homologs or chromosomes, via crossovers or translocations (Guin *et al.* 2020).

Ultimately, the genome of *C. albicans* is distributed across 8 linear chromosome pairs. Each one possesses a centromere, facilitating proper segregation of genetic material during mitosis, and a telomere at each end, ensuring the protection of the DNA content over numerous cell divisions. Indeed, recombination events at centromeres allow the preservation of native centromeres and the formation of neocentromeres upon native centromere loss in *C. albicans*. Additionally, recombination events could also maintain telomere length and potentially ensure livelihood of telomerase deficient cells. Evidently, the exchange of genetic material, by means of recombination events, is involved in assuring the transmission of a complete genome to the next generation. Moreover, the organization of chromosomes within the nuclei may promote inter-chromosomal and intra-homolog recombination events, permitting the shuffling of genetic material and potentially resulting in new advantageous allelic combinations. Thus, chromosomes are the initial blocks which can be modulated, potentially leading to beneficial genetic variations.

2. Repetitive sequences and repeat regions

By definition, a repetitive sequence of DNA is a defined series of nucleic acids which can be found multiple times within a given genome. Repetitive sequences can involve large portions of a genome, thus referred to as repeat regions. The genome of *C. albicans* possesses 4 major types of repeat regions including, telomeres (presented above), subtelomeric regions, major repeat sequences and the ribosomal DNA locus. However, repetitive sequences are not exclusively found within repeat regions but can also be identified across the genome. Below are briefly described the major repetitive sequences and repeat regions found in *C. albicans* and their contributions to generating genomic variation.

Repetitive sequences and repeat regions have recently been reviewed in (Dunn and Anderson 2019).

i. Subtelomeric regions (TLO and LTRs)

Subtelomeric regions are defined as the most telomere proximal region, 10-15 kb, on each chromosomal arm. They are rich in repetitive genetic elements including, transposable elements, expanding gene family members and remnants of ORFs which were disrupted by telomere length variations, and telomerase. Retrotransposons are particularly enriched near telomeres. A total of 7 transposable element families have been identified within subtelomeric regions in the genome of the reference *C. albicans* strain SC5314. They are defined by their

unique flanking sequences of Long Terminal Repeats (LTRs), facilitating their genomic excision and incision (jumps). Lone LTRs, no longer associated to transposons or transposable elements, are also identifiable throughout the genome. These lone LTRs are remnants (or scars) of transposable element jumps. The enrichment of LTRs in subtelomeric regions is probably due to the lowered fitness cost, in contrast to the insertion of a transposable element within the core chromosome DNA. Additionally, transposable element jumps within subtelomeric regions can also rescue telomerase deficient cells through the addition of DNA fragments at chromosome end, acting as an additional buffer (Goodwin and Poulter 2000). LTRs are not restricted to transposable elements and can also be associated to telomere associated (TLO) genes.

Roughly 50 intact protein-coding ORFs have been identified in subtelomeric regions of the *C. albicans* reference strain, SC5314. The latter encode for genes involved in various functions, often associated with pathogenicity of *C. albicans,* such as growth in filamentation inducing conditions, biofilm formation and metabolic processes (Dunn and Anderson 2019). Nevertheless, subtelomeric regions remain poor in ORF compared to internal chromosome regions. Several of these ORFs belong to the *TLO* gene family. Up to fifteen, highly similar (>98% homology), *TLO* gene family members have been identified in *C. albicans* across all 8 chromosomes (van het Hoog *et al.* 2007; Butler *et al.* 2009; Anderson *et al.* 2012). In contrast, the closely related *C. dubliniensis* possess only two *TLO* genes. *TLO* genes encode for Mediator complex subunits (Haran *et al.* 2014), Mediator being a multiprotein complex implicated in transcription coactivation in eukaryotes by interacting with transcription factors and RNA polymerase. *TLO* genes participate in the regulation of distinct virulence properties (Flanagan *et al.* 2018; Dunn *et al.* 2018) potentially explaining why *C. dubliniensis* is less virulent than *C. albicans* (Haran *et al.* 2014). Molecular investigations have revealed recombination-promoting elements in the vicinity of *TLO* genes such as repetitive genetic elements, TLO recombination element (TRE) and the Bermuda Triangle sequence (BTS). These LTRs are found in 11/15 of telomere-associated *TLO* genes, excluding Chr7 and ChrR *TLO* members. A TRE sequence spans over roughly 300 bp and includes the 50 bp BTS. Moreover, these repeats are associated to subtelomeric recombination events modifying the number and sequence of *TLO* gene family members (Anderson *et al.* 2015).

Undoubtedly, subtelomeric regions are rich in repeated elements including gene family members and transposable elements. Though, an epigenetic study of repeat sequences revealed that subtelomeric genes were transcriptionally silenced, by the Sir2 dependent heterochromatin state of telomeres (Freire-Beneitez *et al.* 2016). These regions display high recombination rates and common LOH events, suggesting that genetic shuffling is frequent in

subtelomeric regions (Anderson *et al.* 2015; Wang *et al.* 2018; Todd *et al.* 2019). Indeed, the high incidence of repeats in subtelomeres and their recombinogenic properties, suggest a potential mechanism of generating genotypic diversity through the number and sequence of transposable elements and of TLO gene family members.

ii. Major Repeat Sequences (MRS)

Major repeat sequences (MRS) are unique to *C. albicans* and *C. dubliniensis* and are distributed throughout their genomes. MRS are composed of three subregions: RB2 which contains the *FGR6* gene, the RPS region which varies in repeat numbers (and thus, in size), and the HOK region (Figure 8A). These repeat regions are on average 50 kb long. Additionally, the repeat sequence within the RPS contains a *Sfi*I restriction site which has been useful for strain karyotyping (Chu *et al.* 1993). *C. albicans* possesses eight MRS, one on each chromosome with the exceptions of Chr7 where the presence of one MRS is observed on each arm and Chr3 where only an incomplete MRS is found (Chibana *et al.* 2000; Lephart *et al.* 2005) (Figure 8B). Indeed, such partial MRS sequences, often RB2 (14) and HOK (2) sequences, have also been identified within the genome. Interestingly, MRS are all oriented towards CENs where the HOK region is most proximal to the CEN. Transposon insertion in one of the eight *FGR6* copies found within each MRS is sufficient to give an altered colony morphology phenotype. Of interest, the phenotypes will vary based on the individual *FGR6* allele targeted (Uhl *et al.* 2003). Being large repetitive sequences, MRS are subject to expansion and contraction and this has previously been shown to be involved in chromosome loss (CL) where the chromosome copy containing the longer MRS region is spontaneously lost (Lephart *et al.* 2005). Furthermore, MRS have also been shown to be involved in chromosome translocation (Chibana *et al.* 2000), when two different chromosomes exchange large regions of an arm. Mitotic crossover (MCO) events appear at a similar rate across the MRS (2.82×10^{-6}) versus a non-repetitive genomic region on Chr5, suggesting no enrichment of intra-chromosomal rearrangements at *mrs-5* (Lephart and Magee 2006). However, Pujol and collaborators (Pujol *et al.* 1999) evaluate the recombination frequency (1 in 1000 cells *in vitro*) at MRS, more specifically RPS region of MRS on Chr7, and propose that these recombination events are consistent with intra-chromosomal recombination events (Pujol *et al.* 1999). Chromatin state of MRS sequences is both hypomethylated and highly acetylated, a mix between heterochromatin and euchromatin characteristics. The insertion of a marker gene within the MRS does not lead to its silencing, thus suggesting that these genomic regions are not transcriptionally silenced (Freire-Beneitez *et al.* 2016). This conclusion supports the hypothesis of genomic instability at MRS, which may be beneficial to generate genotypic variation in *C. albicans.* However, it is not excluded that hypomethylation of MRS may be

enough to promote genomic stability by blocking recombination, supported by the work of Lephart and Magee (Lephart and Magee 2006) describing non-significant differences of MCO between MRS and other genomic loci. Despite numerous publications involving MRS, their biological importance and evolutionary maintenance is to this day unclear.

MRS in *C. albicans* have been reviewed in (Chibana and Magee 2009).

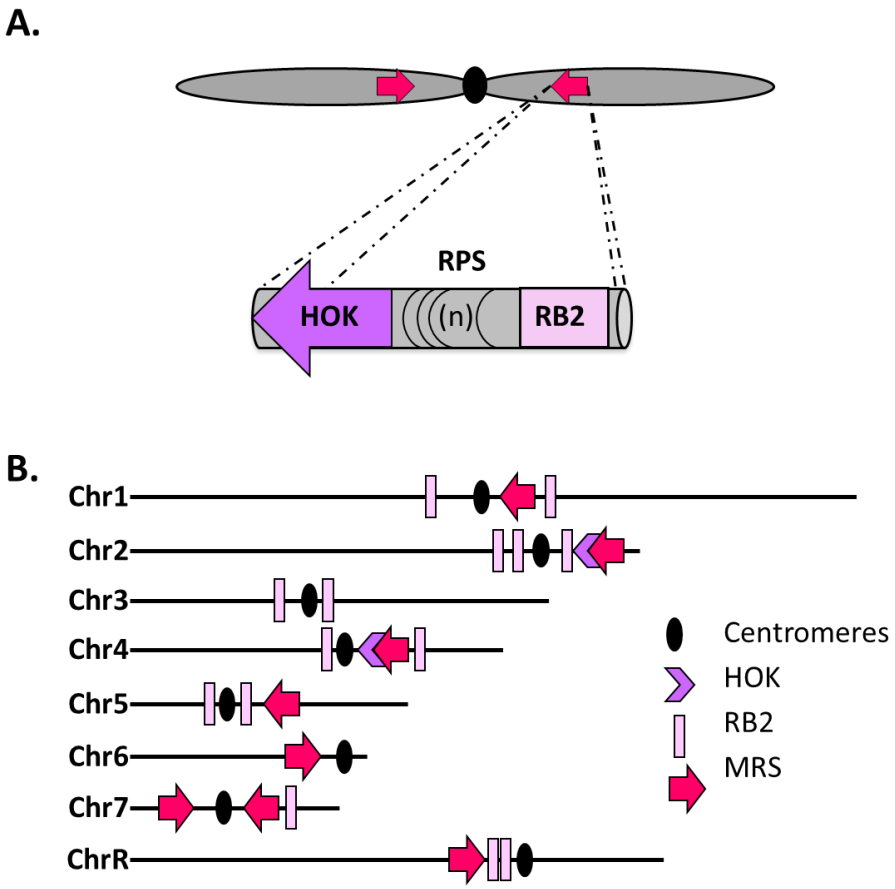

Figure 8: Major repeat sequences (MRS) of C. albicans.
A. Structure of major repeat sequences (MRS) in *C. albicans*. **B.** Distribution of MRS throughout the genome.

iii. Ribosomal DNA locus

The ribosomal DNA (rDNA) or *RDN1* locus is located on the right arm of ChrR in *C. albicans* and encodes for the 18s, 5.8s, 25s and 5s ribosomal RNAs sequences. These sequences are found in tandem repeats at the rDNA locus, which can span 11.6 kb to 12.5 kb and contain 21 to 176 copies depending on the strain and growth conditions (Rustchenko *et al.* 1993). These great variations in rDNA repeats largely influence the absolute size of ChrR (Wickes *et al.* 1991) but also render proper rDNA assembly upon genome sequencing a cumbersome task. It has been shown that monopolin, the recruiter of condensin to CENs and

repetitive sequences during chromosome segregation in *C. albicans*, maintains and stabilizes the length of rDNA repeats (Burrack *et al.* 2013). The rDNA locus is often involved in recombination events, even though the exact rate of recombination is still unknown. This is also illustrated upon genome assembly, post sequencing, where most often the sequence downstream on the rDNA locus on the right arm of ChrR is homozygous (illustrated in Figure 1 of (Ropars *et al.* 2018)).

iv. Inverted repeats

Until recently, the inventory of repeat sequences found within the *C. albicans* genome was relatively limited to the 309 obvious annotated elements in the Candida Genome Database (Skrzypek *et al.* 2017): 129 long terminal repeats, 42 repeat regions,12 retrotransposons and 126 tRNA. A more recent comprehensive study has identified an additional 1,978 repeat sequence pairs (size ranging from 69 bp to 6,499 bp, ≥ 80% sequence identity), named long repeat sequences which cover roughly 3% of the genome (Todd *et al.* 2019). Matches are found intra- (within) and inter- (between) chromosome(s), in tandem, inverted, or mirrored organizations. Often, intra-chromosome repeats are found in tandem, resulting from recombination between chromatids or replication slippages (Todd *et al.* 2019). These repeat sequences do not correlate with GC content or ORF density, but are distributed quasi evenly throughout the genome, within ORFs as well as in intergenic regions (Todd *et al.* 2019). Additionally, this study highlighted the involvement of inverted repeats (IR) on generating genotypic diversity within 33 *C. albicans* isolates, through chromosome inversions, LOH breakpoints and CNV (copy number variation) breakpoints (Todd *et al.* 2019).

The genome of *C. albicans* encompasses diverse repeated sequences, sometimes clustered together forming repeat regions. Indeed, amongst them, repeat sequences are divergent at the DNA sequence level, are regrouped into multiple subcategories and their precise roles are seldom understood. However, it comes as no surprise that repeat sequences and their adjacent DNA are frequently the substrate for generating genotypic diversity. Thus, the abundance of repeat sequences and their associations to events resulting in genetic shuffling is consistent with the observations made by Fischer *et al.* (Fischer *et al.* 2006), where *C. albicans* harbor's one of the most unstable genomes within the Hemiascomycetous yeast lineages.

3. Ploidy

Ploidy, the number of complete genome copies found within a cell, varies amongst organisms, although haploid (1 copy (1n)) and diploid (2 copies (2n)) states are most frequent. While the incidence of polyploidy, more than 2 sets of chromosomes, is highest in plants, certain fungal and animal species also possess polyploid genomes (Otto and Whitton 2000; Otto 2007). Indeed, throughout evolution multiple events of ploidy changes have been observed, such as the WGD event undergone by the *Saccharomycetaceae* family (Dujon 2010). Diploidization or polyploidization, often arise from failure of cell division following mitotic doubling or mating between two varieties or species, called hybridization. Nevertheless, ploidy changes may also result in a reduction of the number of genomic copies, through mechanisms such as meiosis or random loss of chromosomes (Bennett and Johnson 2003). While mechanisms driving ploidy changes are not yet fully understood, a recent review provides insight on the current knowledge of mechanisms of asexual ploidy changes (Todd *et al.* 2017). Additionally, ploidy changes are not restricted to euploidy events, multiples of a haploid number of chromosomes, but may also result in an abnormal number of chromosomes, aneuploidy. In this section I principally focus on the implications of euploid changes while aneuploidy events are discussed in section E.2. Aneuploidy, p.37.

Ploidy changes are frequently observed during the sexual reproductive cycles of eukaryotes or in highly differentiated tissues of organs (e.g. liver and heart muscle) (Gentric and Desdouets 2014; Bergmann *et al.* 2015). *C. albicans* is predominantly found in a diploid state (Jones *et al.* 2004), though haploid and tetraploid strains have also been described (Bennett and Johnson 2003; Hickman *et al.* 2013). Nonetheless, haploid and tetraploid *C. albicans* strains often return to, a more stable, diploid state by undergoing spontaneous whole genome duplication (autodiploidization) or CCLs, respectively (Figure 5). Ploidy convergence towards the baseline ploidy state is a common observation within multiple fungi including *C. albicans, S. cerevisiae, Cryptococcus neoformans* and *Aspergillus nidulans* (Gerstein *et al.* 2006, 2015; Schoustra *et al.* 2007; Hickman *et al.* 2015). The force acting on a lineage to asexually revert back to its baseline ploidy level is also termed “ploidy drive” (Gerstein *et al.* 2017). This process has often been demonstrated (using haploid, diploid, triploid and tetraploid strains) in evolutionary experiments studying ploidy levels in control condition and diverse stresses (salt, ethyl methanesulfonate (EMS) and nutriment poor environments), demonstrating that stress conditions act as an environmental force on ploidy and suggesting that ploidy is implicated in adaptation (Gerstein *et al.* 2006, 2017).

Possessing multiple copies of the genome can be beneficial to mask the effect of recessive and deleterious mutations. The acquisition of a deleterious mutation within a haploid genome has an immediate consequence, while diploid or polyploid organisms can buffer its effects due to genomic redundancy. This may also explain the instability of haploid strains, readily becoming diploid upon stress (Gerstein *et al.* 2006, 2017). In *C. albicans,* upon haploidization of a naturally diploid strain, haploid and autodiploid strains possess a longer doubling times as compared to the heterozygous diploid parental strain, thus suggesting the uncovering of recessive alleles (Hickman *et al.* 2013). Both recessive deleterious mutations, impacting fitness of a strain, and recessive lethal alleles (RLA) have been identified in *C. albicans* (Gómez-Raja *et al.* 2008; Ciudad *et al.* 2016; Feri *et al.* 2016; Segal *et al.* 2018). However, over time, polyploid lineages tend to accumulate deleterious mutations, a serious concern for clonal populations. These mutations may be purged through (para)sexual reproduction, habitually including a transitory ploidy change. For example, during the temporarily elevated ploidy step (tetraploidy) of the parasexual life cycle of *C. albicans*, recombination events allow allele shuffling between parental cells (Anderson *et al.* 2019) while the subsequent ploidy regression by CCL may permit the loss of undesirable allele combinations. Thus, temporary ploidy changes are also beneficial, by promoting genetic variation and allowing (i) the maintenance of favorable allele combinations and/or, (ii) the purge of deleterious or incompatible allele compositions.

Indeed, upon stress conditions polyploid strains tend to become diploid illustrating that genetic doubling is costly in terms of energy for a cell. Thus, temporary or stably maintained ploidy changes must theoretically arise only when beneficial. For example, stably maintained genetic doubling may lead to beneficial outcomes by (i) generating an increase in gene expression (Seoighe and Wolfe 1999), (ii) allowing one copy to evolve a new function (neofunctionalization) or (iii) distributing the function between the two copies thus necessitating both copies to maintain the function (subfunctionalization) (He and Zhang 2005). Certainly, polyploidization does not directly result into these outcomes, but they are influenced by selection, mutation and genetic drift. Nevertheless, polyploidization facilitates the generation of genetic variations and mutations which could subsequently be selected, hence providing additional substrate for the latter processes (Gerstein and Otto 2009). *S. cerevisiae* tetraploid strains do undergo faster adaptation, driven by a higher rate of beneficial mutations associated to stronger fitness effects (Selmecki *et al.* 2015). Alternatively, polyploidization can be incited by environmental changes to favor adaptation. For example, the *C. albicans* parasexual cycle is triggered by growth on nutriment poor media (Lockhart *et al.* 2003). Additionally, the formation of polyploid titan cells in *C. neoformans* is also environmentally

triggered and these large cells play a role in promoting genomic diversity and rapid adaptation in stress conditions (Gerstein *et al.* 2015; Hommel *et al.* 2018).

Evidently, each ploidy level is associated with unique advantages and drawbacks. Haploid *C. albicans* strains have been engineered in the laboratory, with multiple rounds of *in vitro* passaging allowing an increased stabilization of the haploid state. From a research prospective, haploid *C. albicans* cells may facilitate certain studies despite a high fitness cost and a reduced or absent virulence potential in murine systemic infections (Hickman *et al.* 2013). Likewise, although *C. albicans* tetraploid strains are predominantly parasexual cycle intermediates and possess growth rates comparable to diploid strains in laboratory conditions, these strains do exhibit a decrease in virulence and an increase in genetic instability (Ibrahim *et al.* 2005). Altogether, 91% of natural *C. albicans* strains are diploid while the remaining 9% display aneuploidy events, partial ploidy changes (Ropars *et al.* 2018). Indeed, this opportunistic pathogenic yeast tolerates various ploidy states, however they are usually transient suggesting a role in generating genetic variation. These ploidy changes may potentially help *C. albicans* to overcome the absence of meiosis and enhance its adaptation potential as it is confronted to divergent *in vitro* and *in vivo* environments.

4. Heterozygosity

Heterozygosity is defined as possessing two different alleles within an individual. These differences are usually due to the accumulation of SNPs overtime. Diploid *C. albicans* is a naturally heterozygous organism, possessing a heterozygous SNP every ~ 250 bp on average (Hirakawa *et al.* 2015; Ropars *et al.* 2018). In the reference strain SC5314, 56% of all ORFs possess allelic polymorphisms, the majority being affected by non-synonymous SNPs altering the protein sequence (Jones *et al.* 2004). Some of these SNPs may negatively affect the function of an ORF, resulting in the appearance of a deleterious or lethal allele. A drastic example would be the introduction of a premature stop codon within one of the alleles of a coding sequence, leading to a non-functional truncated protein. However, these alleles are often recessive, meaning that their negative effects are masked by the dominant functional allele, ensuring a complete function (*HIS4* (Gómez-Raja *et al.* 2008), *MBP1* (Ciudad *et al.* 2016), *GPI16*/*MRF2* (Feri *et al.* 2016)). Additionally, heterozygosity in *C. albicans* is also associated with differences in levels of RNA expression and ribosomal occupancy between alleles (Muzzey *et al.* 2013, 2014). Overall, allelic differences may generate different phenotypes and bring versatility, with a given environment favoring one allele while the second allele may be beneficial in another condition. Undoubtedly, heterozygosity is a rich source of genetic variation.

Comparative analyses have shown that highly heterozygous strains tend to grow faster in nutriment poor media, clearly illustrating the relationship between heterozygosity and fitness (Hirakawa *et al.* 2015). Contrarily, both haploid and autodiploid strains are outcompeted by heterozygous *C. albicans* strains in systemic murine infection model as they possess significantly reduced growth rates (Hickman *et al.* 2013) (Figure 9A-B). Characterization of autodiploids also revealed that the fitness cost associated with homozygosis drastically increases at the host temperature of 37°C, and that autodiploids are poor commensals in a murine model of gastrointestinal (GIT) colonization (Marton *et al.* unpublished data) (Figure 9C-D). Additionally, strains with naturally low levels of heterozygosity, such as cluster 13 *C. africana* strains, show slower growth in various conditions and lower virulence in infection models (Borman *et al.* 2013; Ropars *et al.* 2018). All *C. africana* strains have been isolated from genital tracts suggesting that the elevated levels of homozygosis, and their related fitness costs, may be restricting cluster 13 strains to this unique niche (Ropars *et al.* 2018). Altogether, this suggests a positive correlation between heterozygosity and strain fitness in *C. albicans*.

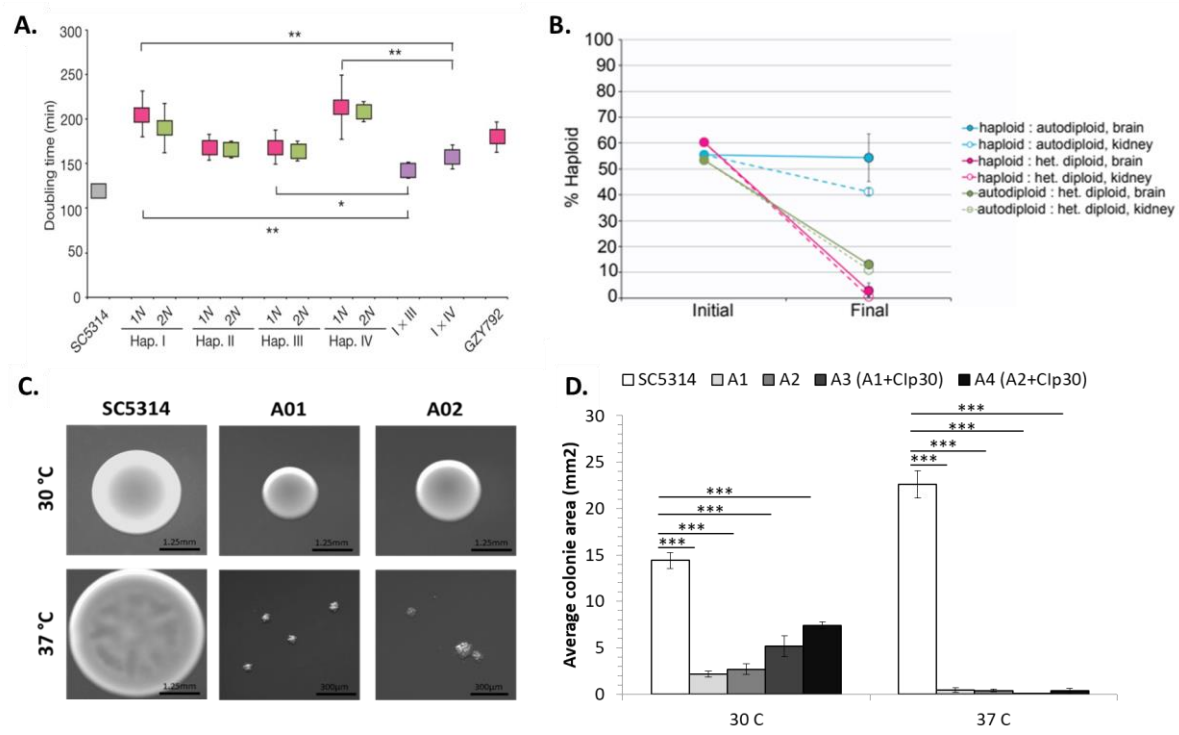

Figure 9: Relationship between genomic heterozygosity and fitness in C. albicans.
A. Doubling times of homozygous haploid and autodiploid strains compared to heterozygous diploid SC5314 reference strain. **B.** Competition assays between haploid, autodiploid and heterozygous diploid *C. albicans* strains in *in vivo* murine infection model. Percentage of haploid (blue and pink) or autodiploid (green) *C. albicans* cells introduced by tail vein injection in mice (initial) versus the percentage recovered in brain and kidneys after 48h of infection. **C.** Colony size on YPD after 48H incubation at 30°C or 37°C. Both autodiploid A01 and A02 (uri-, his-, arg-) strains derive from the haploid GZY792 strain. **D.** Average colony area in mm^2 (n=10) on YPD after 48H incubation at 30°C or 37°C. A03 and A04 autodiploid strains correspond to prototroph A01 and A02 strains, respectively. Measurements taken using Fiji. Panels A and B are taken from (Hickman *et al.* 2013) and C and D correspond to personal unpublished data (Marton *et al*. unpublished data).

Heterozygosity is beneficial in terms of multiplying genetic variation within a genome without necessitating the duplication of generating material. Though the complex relationship between genotype and phenotype in *C. albicans* is not fully understood, the genetic variations observed at the nucleotide level inevitably tickles the question of their impact from a phenotypic point of view. Allelic variety, by amplifying its genetic panel, may be a potential for adaptability in certain organisms such as *C. albicans*.

E. Genomic plasticity – Genomic rearrangements

Ensuring the maintenance of a cell's karyotype is crucial for its viability although, karyotype variabilities are frequent in some species. As discussed in the previous chapter, the genome of *C. albicans* possesses several genomic characteristics promoting genomic plasticity. Genomic plasticity can be described as the alterable nature of a genome, allowing for DNA exchanges and rearrangements to occur. Certainly, genomic plasticity is a major particularity of *C. albicans*, as this yeast is highly tolerant to multiple genomic rearrangements events (Fischer *et al.* 2006). The following chapter will present the major gross chromosomal rearrangement (GCR) events described in *C. albicans* strains, including translocations, aneuploidies, formation of isochromosomes and LOH.

1. Translocations

Chromosome translocation are GCR events resulting in an abnormal chromosome organization. Two major types of translocations exist, reciprocal and non-reciprocal translocations. A reciprocal translocation event arises from the exchange of chromosome portions between two different chromosomes, while an event which does not have a mutual exchange of chromosome portions is called a non-reciprocal translocation (Figure 10). These genomic rearrangement events do not necessarily impact a cell's ploidy or genetic content but, frequently alter the sizes of involved chromosomes. Thus, translocation events are often identified using methods involving pulsed-field gel electrophoresis, permitting the separation of chromosomes by size.

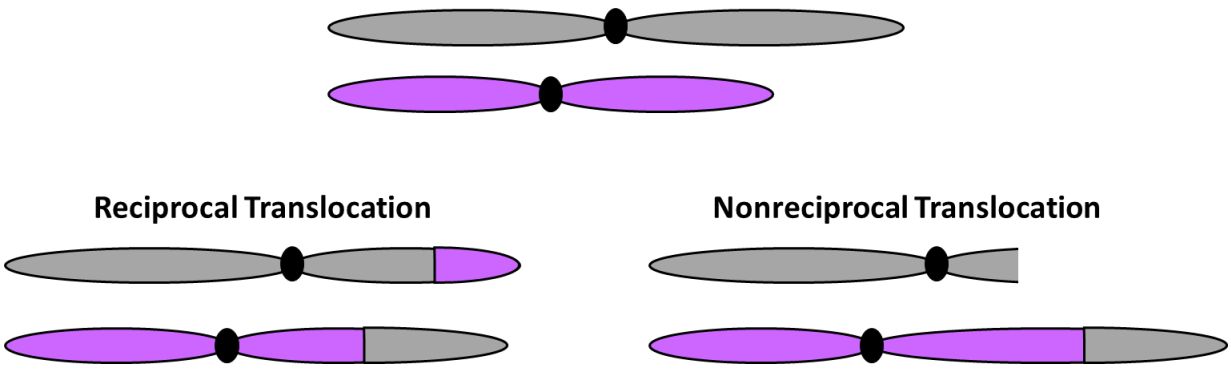

Figure 10: Translocations.
Translocation events involve the exchange of genetic material between two distinct chromosomes (illustrated in gray and purple). This exchange may be reciprocal, characterized by no loss of genetic material, or nonreciprocal, resulting in loss of genetic material.

Potentially due to the lack of research tools, translocation events in *C. albicans* are scarcely studied. Overall, translocation events are rarely reported in *C. albicans* but when they are, they often involve Chr2, Chr4 and Chr7 (Chu *et al.* 1993; Navarro-Garcia *et al.* 1995; Iwaguchi *et al.* 2000, 2001), though those events are most likely underestimated. During the construction of the *Sfi*I macrorestriction map of the *C. albicans* genome, 3 reciprocal translocation events have been detected and described for the first time in the WO-1 strain (Chu *et al.* 1993), strain in which the “white-opaque” switch was first described. Since then, other reciprocal translocations have been described in both clinical and laboratory *C. albicans* strains (Chu *et al.* 1993; Navarro-Garcia *et al.* 1995; Iwaguchi *et al.* 2000, 2001). This exchange of genetic material between chromosomes tends to occur in proximity to a MRS. Upon the integration of exogenous DNA (*URA3* auxotrophic marker), within the *RB2* sub-region of the MRS, reciprocal translocations occurred in roughly 9% of transformants (Iwaguchi *et al.* 2004), suggesting that reciprocal translocation events in *C. albicans* are associated with MRS. Indeed, MRS are good substrates for the occurrence of translocation events because they are found on seven of the eight chromosomes and are substantial regions of interhomolog homology. Certain MRS’s have been proposed as hotspots of recombination (Pujol *et al.* 1999), where mitotic recombination events can take place upon the pairing of heterologous chromosomes at MRS sequences, resulting in a reciprocal translocation.

The implication of translocation events in *C. albicans* are not well understood, though some of these chromosome reorganization events were associated with an altered phenotype, suppressor of ploidy shift (defects in regulating ploidy, phenotype leading to mixed ploidy states within a culture) and incapacity to form chlamydospores (Iwaguchi *et al.* 2000, 2001). Nevertheless, translocation events are better understood in humans, and many of them have

been associated to several diseases, notably cancers (Lin *et al.* 2012). The advent of long-read sequencing, such as PacBio or Nanopore sequencing, may facilitate the study of translocations events in *C. albicans* and may aid to better understand the extent and impact of translocations on generating genetic diversity by creating new linkage relationships in *C. albicans*.

2. *Aneuploidies*

In the previous chapter (section D.3 Ploidy, p.31), it was highlighted that ploidy changes affecting the entire set of chromosomes was costly for a cell in terms of energy. In contrast, an aneuploidy event would be a more energetically affordable ploidy change and thus potentially readily accessible for a cell. This is certainly the case in *C. albicans* since aneuploidy events are the primarily observed polyploidization events. An aneuploidy may impact a whole chromosome or be restricted to a fraction of a chromosome, a segmental aneuploidy. Both gains and losses of chromosomes lead to aneuploidy, an abnormal number of chromosomes within the cell (Figure 11).

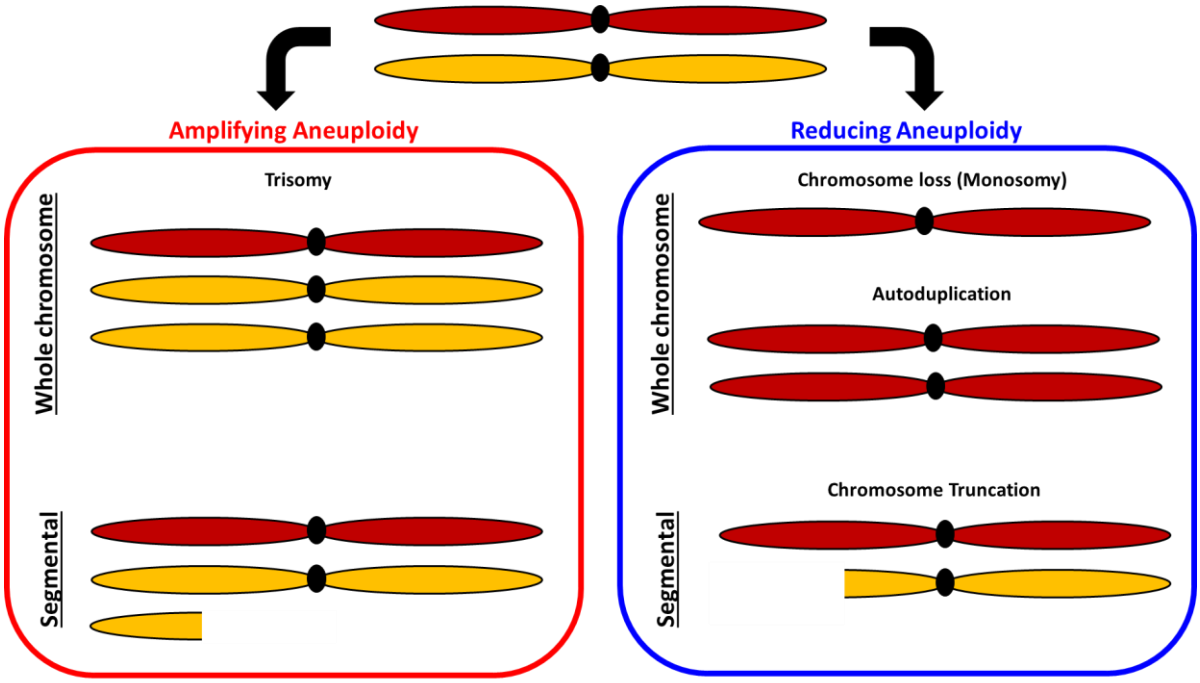

Figure 11: Examples of aneuploidy events in a diploid background.
Aneuploidy events are copy number variations (CNVs) which affect chromosomes by either amplifying or reducing the number of copies. These GCRs may implicate whole chromosomes or be restricted to a fraction of a chromosome, segmental. The two homologues for a given chromosome are depicted in burgundy and yellow.

Numerous aneuploidy events have been described upon chromosome segregation defects, imperfect CCL, partial endoduplication and the absence of DNA-DSB repair, however

we currently lack a complete understanding of underlying mechanisms driving aneuploidy. Whole chromosome aneuploidies, resulting in monosomy or trisomy, are the most frequently observed aneuploidy events in *C. albicans* strains (Selmecki *et al.* 2005). They often arise from chromosome segregation defects during mitosis, following the retention of both copies of a given chromosome homolog by one of the two cells. Thus, both cells originating from this cell division are consequently aneuploid, where one cell undergoes a reducing aneuploidy event (losses a chromosomal homologue, monosomy) and the other undergoes an amplifying aneuploidy event (gain of an extra chromosomal homologue, trisomy). Additionally, aneuploid progenies of the parasexual lifecycle of *C. albicans* are often observed since CCL does not always lead to perfectly diploid progenies (Forche *et al.* 2008; Hickman *et al.* 2015). Amplifying segmental chromosome aneuploidies have also been described in *C. albicans*, resulting in more than 3 copies of a specific chromosome section. Their breakpoints are often located in intergenic regions (Selmecki *et al.* 2005) and coincide with inverted repeats, suggesting that these genomic features may facilitate their formation (Todd *et al.* 2019). Segmental reducing aneuploidy are commonly referred to as chromosome truncations (CTs) because they usually span from a given point of a chromosome arm to the telomere. In order to prevent further DNA degradation, CTs are capped with a 9 nt.-long repeat sequence possessing a pattern similar to telomeric repeat sequences (Selmecki *et al.* 2005). Furthermore, strains possessing defects in the DNA repair machinery are predisposed to the acquisition of reductional aneuploidy events. Larriba and coworkers utilized the heterozygous *HIS4* locus, composed of a functional and non-functional *HIS4* alleles (Gómez-Raja *et al.* 2008) on Chr4 in the CAI-4 *C. albicans* strain, to demonstrate that homozygous deletion mutants for homologous recombination (HR) DNA repair components (*RAD51* or *RAD52*), promote genomic instability resulting in reductional aneuploidies, CL and CT (Andaluz *et al.* 2011; Bellido *et al.* 2019; Ciudad *et al.* 2020). By monitoring the appearance of histidine auxotroph colonies, they calculated a greater rate ($5.4x10^{-3}$ events per cell per generation) of loss (or disruption) of the functional *HIS4* allele in *rad52*$^{-/-}$ mutants, as compared to wild-type strains. Upon characterization of 80 *rad52*$^{-/-}$ histidine auxotrophs, 34% of colonies underwent CT while the remaining 66% underwent CL (Andaluz *et al.* 2011).

Aneuploidies are also frequently associated to several environmental shifts (Selmecki *et al.* 2006a; Forche *et al.* 2011; Yona *et al.* 2012). In particular, CL has often been described in strains following heat shock, exposure to fluconazole and prolonged growth on glucose-rich or sorbose media (Hilton *et al.* 1985; Janbon *et al.* 1998; Bennett and Johnson 2003; Bouchonville *et al.* 2009; Forche *et al.* 2011). Additionally, both fluconazole and heat shock exposure may also result in amplifying aneuploidy events (Bouchonville *et al.* 2009; Arbour *et al.* 2009; Harrison *et al.* 2014). Whole chromosome aneuploidies are frequently unstable and

therefore a monosomy will often tend to reduplicate in order to revert to a stable diploid state. This aneuploid genomic instability is also observed upon whole chromosome polyploidy, as passaging of these strains habitually reverts to a diploid or principally diploid state (Hickman et al. 2015).

Nevertheless, these genomic rearrangements may be a transient evolutionary solution to cope with stress, by rapidly generating new beneficial phenotypes (Berman 2016). Upon a given stress, increasing the number of copies of a specific gene to increase its expression level may turn out to be crucial. As it was shown upon antifungal treatment, an amplifying aneuploidy can result in higher expression of drug efflux pumps, allowing the cell to survive (Selmecki *et al.* 2006a). Alternatively, reductional aneuploidies may lower the drug targets, which would also be potentially beneficial for cell survival. As previously discussed, heterozygosity between alleles in *C. albicans* is associated to differential allele expression. Consequently, CL followed by its reduplication can serve as a mechanism of beneficial allele selection, upon environmental shifts.

3. Isochromosomes

Isochromosomes are supernumerary chromosomes, extra chromosomes possessing a structural abnormality where the chromosome arms are mirror images of each other. These genomic rearrangement events result in two additional copies of a given chromosome arm and thus, they may also be classified as aneuploidies. Isochromosomes in *C. albicans* were first described in the work of Selmecki *et al.* (Selmecki *et al.* 2006a), during karyotype characterization of fluconazole sensitive and resistant strains using comparative genome hybridization (CGH) array. Subsequently, diverse isochromosome organizations of Chr5 were observed in fluconazole exposed strains, i.e. multiple occurrences of isochromosomes simply composed of inverted copies of a single chromosome arm (left arm of Chr5 [i(5L)] or right arm of Chr5 [i(5R)]), a dicentric isochromosome composed of a mirror image of the left arm of Chr5 fused to an intact Chr5 [i(5L)+Chr5], or a monocentric isochromosome possessing mirrored images of the left arm of Chr5 fused to the right arm of Chr3 [i(5L)+3R] (Selmecki *et al.* 2005, 2006a, 2008, 2009) (Figure 12A). More recently, an isochromosome composed of the right arm of Chr4 has also been described (Todd *et al.* 2019), suggesting that isochromosome formation is not limited to Chr5 in *C. albicans.* Isochromosome formation is suggested to be facilitated by repeat sequences flanking CENs, such as IR (Selmecki *et al.* 2006a; Todd *et al.* 2019). Therefore, Chr1, 4, 5 and R should be theoretically more susceptible of forming isochromosomes in *C. albicans*, as they possess IR flanking their CENs. Two IR-mediated mechanisms of isochromosome formation are possible; (i) by the repair of a DNA-DSB, in

proximity of CEN-associated IR, via break-induced replication (BIR) (section G.2.ii.b.b4 Break-induced replication, p.75) using the intact chromosome arm as a repair template (Figure 12B) and (ii) thru an inter-chromatid recombination event between IR (Figure 12C).

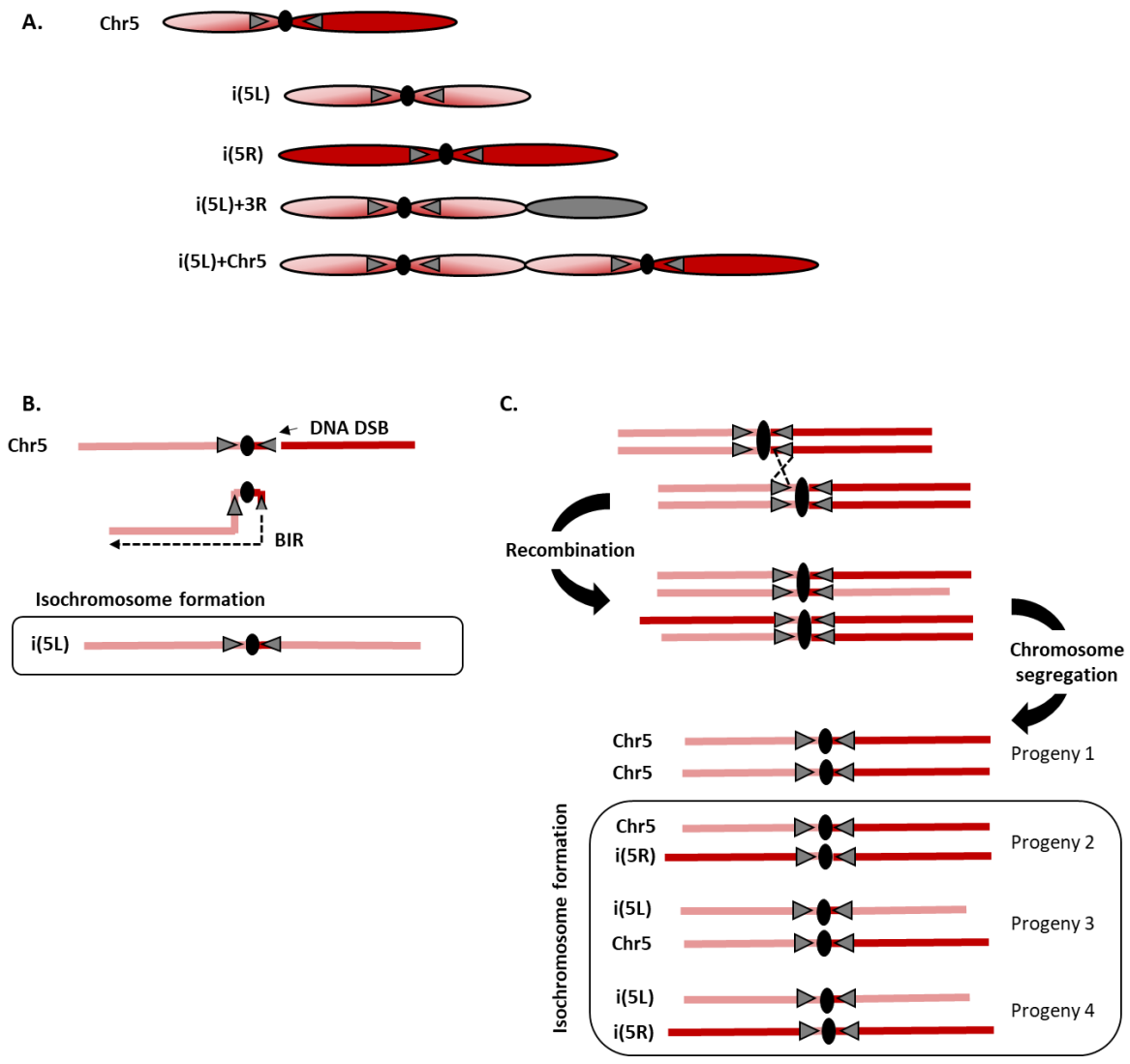

Figure 12: Isochromosomes involving Chr5 in C. albicans and inverted repeat (IR)-mediated mechanisms of isochromosome formation.
A. Summary of Chr5 isochromosomes described in *C. albicans* fluconazole-exposed strains, duplication of left or right arms of Chr5, i(5L) and i(5R) respectively, duplication of left arm of Chr5 and fusion to the right arm of Chr3, i(5L)+3R and duplication of the left arm of Chr5 and fusion to an intact Chr5 i(5L)+Chr5. **B.** Mechanism of isochromosome formation mediated by IR sequences flanking centromeres upon DNA-DSB repair by BIR, using the left arm of Chr5 as a repair template. **C.** Isochromosome formation resulting from an inter-chromatid recombination event followed by chromosome segregation. Black ovals: centromeres, gray triangles: inverted repeats, dashed black line: break-induced replication (BIR).

Isochromosomes have only been described in strains which have undergone an environmental shift i.e. exposure to an antifungal, suggesting adaptive implications. Benefits associated to isochromosome acquisition are identical to those of amplifying aneuploidy events, previously discussed (section E.2 Aneuploidy, p.37). Duplication of chromosomal arms leads to multiple gene copies, which heighten their expression levels, potentially leading to beneficial phenotypes, such as increased minimal inhibitory concentrations (MIC) of antifungal or growth rate (Selmecki *et al.* 2008, 2009; Todd *et al.* 2019). All strains possessing the previously described isochromosomes, notably i(5L), exhibit an increased fitness in presence of fluconazole, while loss of these GCRs restores parental fitness levels (Selmecki *et al.* 2009). Isochromosomes may be more advantageous than aneuploidies, as they bring two additional copies of a genome segment, as compared to the more often single additional copy resulting from aneuploidies. Additionally, isochromosomes could be more stable in time than segmental aneuploidy events, as they possess centromeres. For instance, i(4R) can be maintained throughout roughly 300 generations in absence of selective pressure, i.e. presence of fluconazole (Todd *et al.* 2019).

4. LOH

Loss of heterozygosity (LOH) events are GCRs resulting in the loss of one of two genomic copies. Upon sequencing and mapping of the heterozygous SNP density across the eight *C. albicans* chromosomes, nearly all strains possess at least one region that is poor in heterozygous SNP amongst the overall highly heterozygous genome (Figure 13). LOH are observed ubiquitously across the genome of *C. albicans*, affecting all chromosomes of laboratory and clinical isolates. These LOH events vary in size with certain events being limited to a single chromosomal region, while others can affect an entire chromosomal arm or even cover the entirety of a chromosome (Abbey *et al.* 2014; Ropars *et al.* 2018).

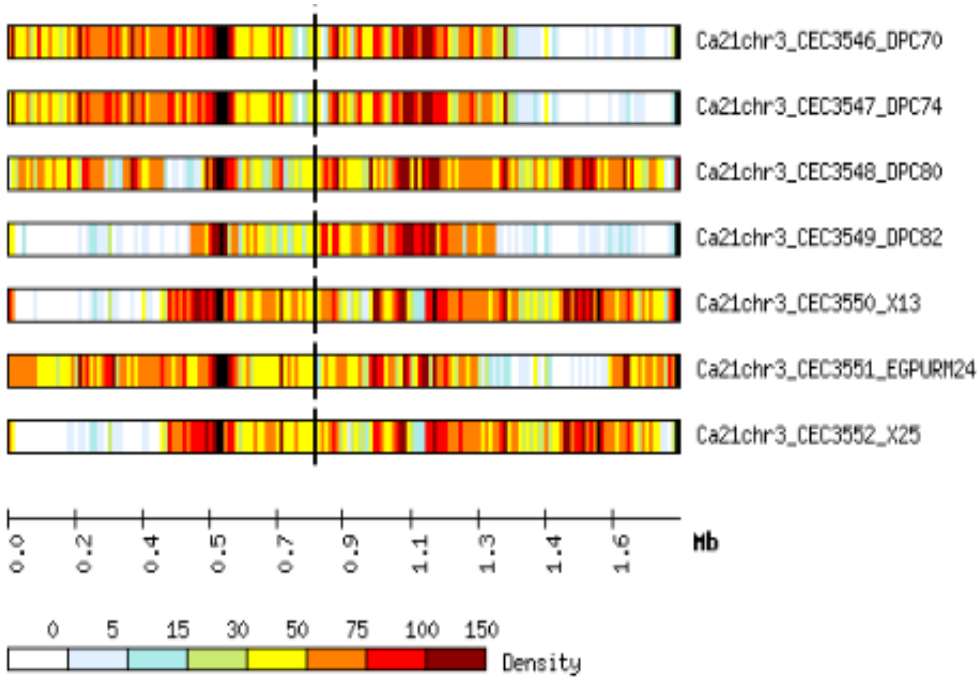

Figure 13: Density map of heterozygous single nucleotide polymorphisms (SNPs) across Chr3 of seven C. albicans strains.
The Chr3 was divided into 10 kb long frames and each frame was assigned a color based on the number of heterozygous SNPs found between Chr3A and Chr3B. Colors range from white (absence of heterozygous SNPs) to deep red where roughly 100 to 150 heterozygous SNPs were identified. The centromere of Chr3 is identified by the black vertical line.

DNA-DSBs have previously been shown to result in LOH events by means of various DNA repair mechanisms (Malkova *et al.* 2000). Cells are efficient at repairing DNA-DSBs, with a panel of homologous recombination-dependent mechanisms. Homologous recombination is a mechanism of genetic shuffling between two similar molecules. During DNA repair, the intact chromosome copy will be utilized as a template in order to patch up the break. The extent of sequence homozygosis will depend on the molecular mechanism involved in DNA repair. For example, when homozygosis is localized, it is the result of a gene conversion (GC). When this region extends from the DNA break site to the telomere of the same chromosomal arm, it is likely to result either from (i) break-induced replication (BIR), (ii) mitotic crossover (MCO) or (iii) GC with crossover (GC w CO) if the break is induced during mitosis. In contrast, if left unrepaired, a DNA-DSB can result in a complete loss of the broken section (CT) or the entire chromosome (CL) (Figure 14). Certainly, CL and CT were previously presented as reductional aneuploidies (2 copies to 1 copy) but are also examples of LOH events. Additionally, chromosome segregation is another cellular process whose defects might result in LOH, leading to a long-tract LOH spanning an entire chromosome.

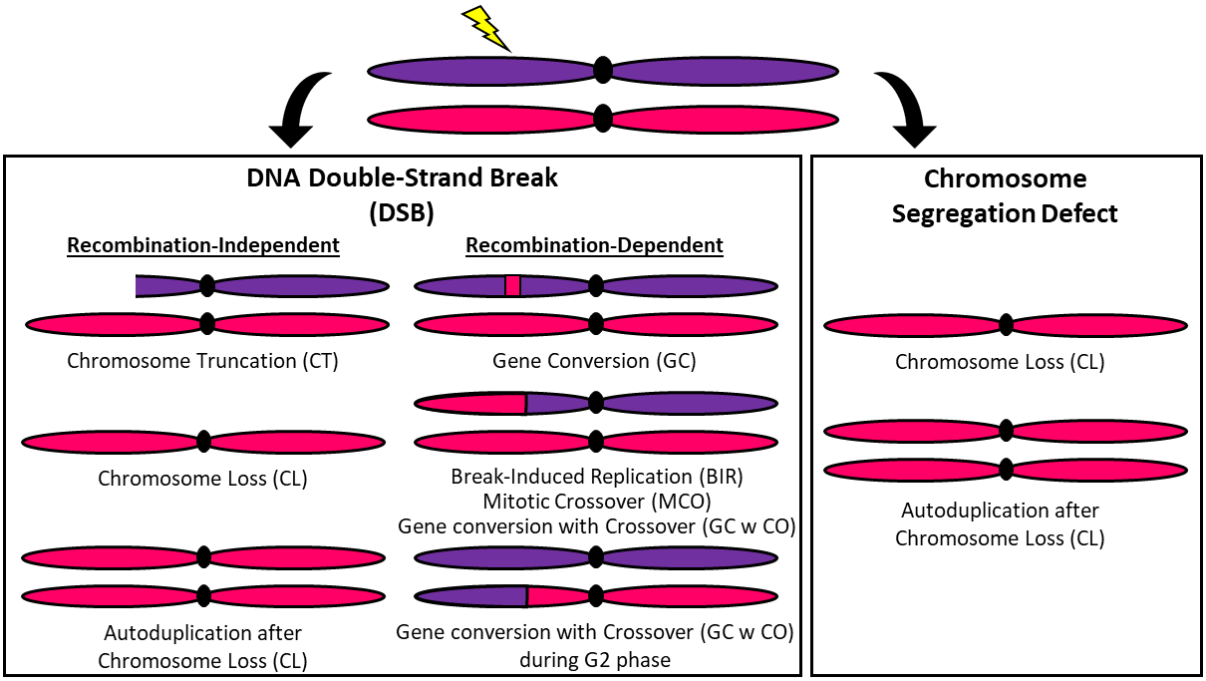

Figure 14: Molecular mechanisms resulting in LOH events upon a DNA-DSB.
A DNA-DSB may be repaired by recombination-dependent mechanisms such as break-induced replication (BIR), mitotic crossover (MCO) or gene conversion with crossover (GC w CO) resulting in a long-tract LOH. While gene conversion (GC) is at the origin of short-tract LOH events. If the DNA-DSB is left unrepaired, the damaged chromosome can either be truncated or lost (often followed by its autoduplication). Chromosome segregation defects also result in long-tract LOH.

LOH events are also associated with several biological origins which assault the integrity of the DNA, notably oxidative stress, antifungal treatments, and elevated temperature. Thus, similarly to other GCRs, LOH arise upon environmental shifts. In a study by Forche *et al.*, it was demonstrated that LOH nature and frequencies vary upon exposure to physiologically relevant stresses at the *GAL1* locus on Chr1 (Forche *et al.* 2011). Indeed, LOH rates augment in presence of stress, where fluconazole induces the strongest LOH response resulting in predominantly large LOH of Chr1. Exposure of *C. albicans* to either the antifungal agent fluconazole or to an elevated temperature (39°C) principally leads to long-tract LOH events while DNA lesions from oxidative stress (H_2O_2) are predominantly repaired by GC, resulting in short LOH events (Figure 14). Additionally, the intensity of a stress seems to be correlated with the augmentation of LOH rates and a shift in LOH types (Forche *et al.* 2011). Adding to that the fact that overexpression of various *C. albicans* genes involved in DNA repair influence LOH nature (Loll-Krippleber *et al.* 2015), one can propose that different environmental cues could trigger different signaling pathways resulting in different LOH outcomes. This is coherent with the hypothesis that LOH events are triggered by stresses and thus may facilitate adaptation of cells when encountering unfavorable conditions. Beyond that, this type of GCR is also primordial in *C. albicans* biology since LOH is (i) required to fulfil the condition of *MTL* homozygosis before undergoing the parasexual life cycle and (ii) essential

during the ploidy reduction step (Bennett and Johnson 2003), as discussed in section C. (Parasexual) Life-cycle of *C. albicans* (p. 17).

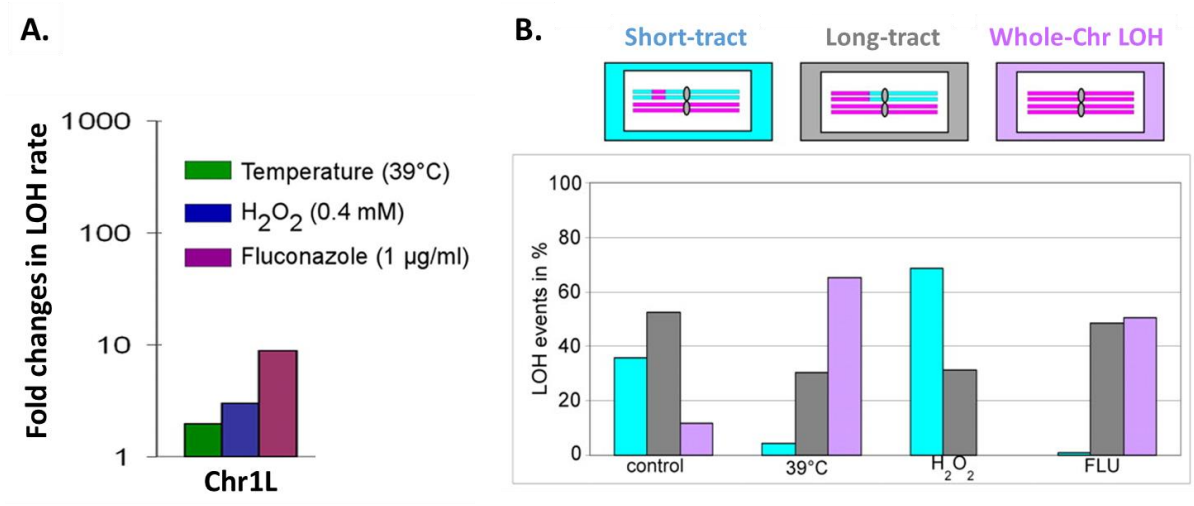

Figure 15: Rate and nature of LOH events in C. albicans vary between stresses.
A. Fold changes in LOH rates at the *GAL1* locus on Chr1 between control and stress growth conditions of *C. albicans*. **B.** Characterization of LOH events at *GAL1* locus on Chr1 upon exposure to diverse stressors reveals that stresses result in different LOH lengths. Figure adapted from (Forche *et al.* 2011)

The impact of LOH events on the biology of *C. albicans* is likely to be linked to the length of the affected area. LOH events encompassing entire chromosomes or chromosomal arms impact a greater number of ORFs than short-tract LOHs. As mentioned before, changes between alleles at the nucleotide level can result in changes in gene expression, protein production and protein function (Staib *et al.* 2002; Muzzey *et al.* 2013, 2014) though they do not always lead to a phenotypic difference in the heterozygous state. Phenotypes associated with recessive alleles can only be observed when found in the homozygous state. Anterior studies conducting haplotype characterization of (i) progenies from the parasexual life-cycle (Forche *et al.* 2008), (ii) homozygous diploid isolates derived from *RAD52* double knock-out mutants (Andaluz *et al.* 2011), and (iii) haploid strains of *C. albicans* (Hickman *et al.* 2013) showed a chromosome haplotype bias in the *C. albicans* laboratory strain SC5314. Indeed, the homozygosis state of some chromosomes was only observed for a given homolog, while recurrently absent for the other homolog. This suggests the presence of detrimental recessive alleles, where lethality or poor viability phenotype of the individuals can only be observed in the homozygous state. Indeed, recessive lethal alleles (RLA) and recessive deleterious alleles have been identified in laboratory and clinical isolates of *C. albicans* (Feri *et al.* 2016).

Additionally, LOH events leading to the complete loss of one copy can also affect gene dosage and unveil haploinsufficiency. LOH encompassing the *MBP1* gene, which regulates G1/S genes involved in DNA repair, on Chr3 in the *C. albicans* strain CAI4 results in methyl methanesulfonate (MMS) sensitivity (Ciudad *et al.* 2016). Indeed, at times both variants of a

gene are required to generate a sufficient quantity of transcripts thus, a single variant allele is insufficient for the cell to produce the same phenotype. Contrarily, an increase in expression may also perturb a phenotype by destabilizing the stoichiometry of interacting proteins, especially in protein complexes. By screening a collection of 5639 *C. albicans* heterozygous deletion mutants, a total of 300 dispensable genes and 87 essential genes were shown to be haploinsufficient for cell size, amongst which some are also required for fungal pathogenicity (Chaillot *et al.* 2017). Indeed, hemizygosis, when only one copy of a gene is present in a diploid organism, participates in generating phenotypic variability under different growth conditions (Ko *et al.* 2013) potentially driven by GCRs, modulating the genetic variants. The impact of gene dosage in *C. albicans* has recently been discussed in a review by Liang and Bennett (Liang and Bennett 2019).

LOH events also facilitate the passage of beneficial recessive mutations through Haldane's sieve (Gerstein *et al.* 2014). Certainly, mutations are not restricted to having negative outcomes, as mutations acquired in regulatory and coding sequences may have beneficial consequences. Hence, LOH can uncover disused beneficial alleles by revealing a new advantageous phenotype. Gain-of-function (GOF) mutations are another type of beneficial mutations, leading to acquisition of a new or enhanced function. Indeed, the sheer presence of a heterozygous GOF mutation improves function, when it is rarely the case for recessive alleles. When associated with an LOH event, the beneficial effect of a GOF mutation is yet again augmented, potentially aiding the fixation of the beneficial allele within the population. The hallmark example of this scenarios in *C. albicans* involves the *TAC1* gene, a transcription factor of drug efflux pumps. Coste *et al.* described several GOF mutations identified within clinical strains leading to hyperactivity of *TAC1*, as measured by the upregulation of its regulated genes *CDR1* and *CDR2*, encoding drug efflux pumps (Coste *et al.* 2006, 2007). Cells heterozygous for the GOF mutant *tac1* have an increased resistance to azoles though the full beneficial potential of GOF *tac1* mutant is obtained in the homozygous state (Figure 16).

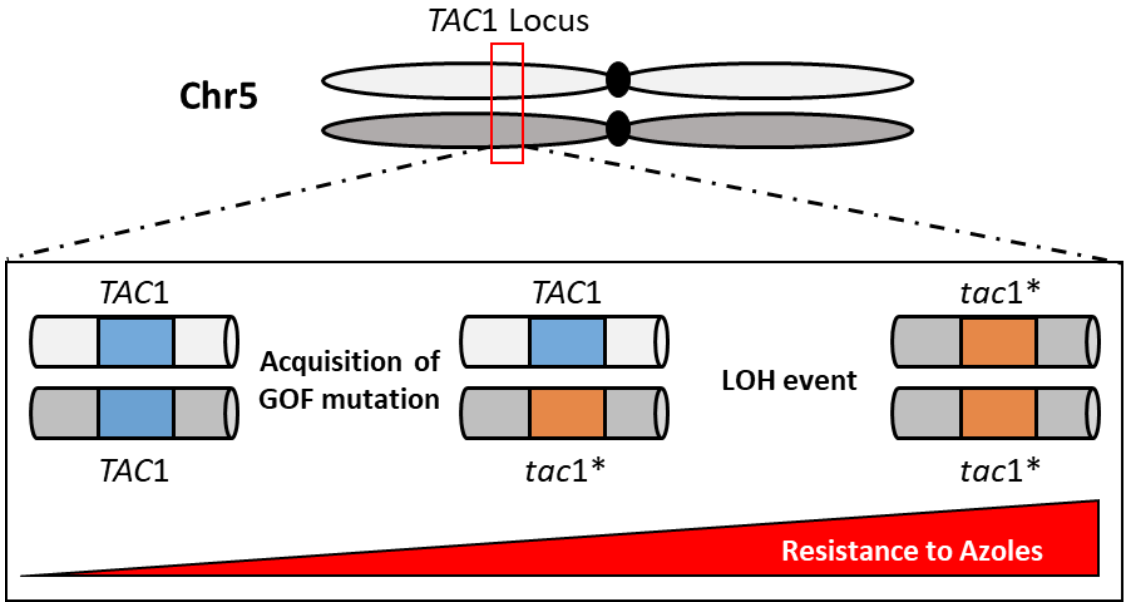

Figure 16: Involvement of LOH in azole resistance.
Acquisition of a gain of function (GOF) mutation leads to hyperactivity of *TAC1* and upregulation of drug efflux pumps (*CDR*1 and *CDR*2), increasing cell resistance to azoles. The full beneficial potential of this mutation is obtained in the homozygous state, following an LOH event. (Coste *et al.* 2007)

LOH events permit genetic shuffling which contribute to the uncovering of alleles bringing selective advantages. Contrarily to GCRs which result in an increased number of allelic copies, LOH are irreversible GCR events having prolonged consequences. Indeed, reacquisition of heterozygosity can be a lengthy process as, in a clonal population, it is only permitted by the accumulation of point mutations. Nevertheless, the parasexual life cycle of *C. albicans* can participate to the regain of heterozygosity.

Overall, *C. albicans* possess a highly plastic genome as several types of genomic rearrangements altering gene synteny and allele ratios and compositions are tolerated. Certainly, this genomic particularity may have evolved due to the predominant clonal population, which limits genetic variations. Thus, through reshuffling of existing alleles, often associated to DNA-DSB and facilitated by numerous repeat sequences, new selective advantageous traits can emerge. Though the acquisition of genetic variation is important, an equilibrium between genomic plasticity and genomic stability is crucial for genetic integrity and for assuring the longevity of the lineage. However, the mechanisms, parameters and regulation of genetic plasticity in *C. albicans* are still poorly understood.

F. Tools to study genome stability in *C. albicans*

To facilitate the study of genome stability in *C. albicans*, several tools have been developed in order to induce, monitor, quantify, and characterize GCRs. The following section will predominantly be focused on the major tools enabling the study of LOH events, especially those which have been used throughout my thesis projects.

1. Tools to induce DNA lesions

i. Genotoxic compounds and physical stress

The frequency of genetic rearrangements, consequences of genomic instability, is enhanced in stress conditions as highlighted in previous sections of the introduction. Indeed, livelihood of cells is permanently threatened by stress, causing DNA damages which must be repaired. Thus, studies aimed to understanding genome stability often use genotoxic agents, mimicking stress conditions, to generate DNA breaks and lesions. A variety of genotoxic agents, inducing various types of DNA damages, exist though, some are more commonly used within the *C. albicans* community, such as hydroxyurea (HU), antibiotics (bleomycin and camptothecin), oxidizing agents (hydrogen peroxide, menadione and tert-Butyl hydroperoxide (tBHP)), alkylating agents (Methyl methanesulfonate (MMS) and Ethyl methanesulfonate (EMS)), ultraviolet radiation (UV) and heat.

The latter compounds assault DNA in various ways. HU is an inhibitor of the ribonucleotide reductase thus, hindering dNTP synthesis and DNA replication (Koc *et al.* 2004). Bleomycin and camptothecin are highly mutagenic antibiotics used for cancer treatment that are also cytotoxic for unicellular eukaryotic organisms, e.g. yeasts. Bleomycin is a cytotoxic drug capable of inducing both double- and single-strand DNA breaks by inhibiting DNA synthesis and formation of reactive oxygen species (ROS) (Chen *et al.* 2008). Mutagenic properties of camptothecin are caused by stabilizing the interaction between the DNA topoisomerase Top1 and DNA, hence promoting DNA breaks (Sloan *et al.* 2017). The metabolization process of oxidizing agents produce ROS which are well known to cause damage in lipids, proteins and DNA. Indeed, oxidative damages generated by hydrogen peroxide (H_2O_2) in yeast include DNA single-strand breaks (SSB), DSB and base modifications (Girard and Boiteux 1997; Zhang *et al.* 2019b). In contrast, alkylating agents provoke damages through addition of ethyl or methyl groups predominantly on guanine and adenosine residues. While EMS is primarily associated with base substitutions, MMS is more associated to DNA breaks (Beranek 1990; Myung and Kolodner 2003). In terms of physical

stresses, exposure to UV promotes formation of pyrimidine dimers causing torsion in DNA molecule (Boyce and Howard-Flanders 1964). As presented in previous sections, heat shock is a very efficient way to induce genomic instability in *C. albicans*, as it inhibits the activity of DNA repair mechanisms, thus promoting DNA damages within the cell (Kantidze *et al.* 2016). In addition, agents targeting the cellular machinery involved in cytokinesis can also influence genomic stability, especially the appearance of polyploid cells and aneuploidy events as highlighted in previous sections. Indeed, Harrison *et al.* (Harrison *et al.* 2014) demonstrated that fluconazole has a potency of inducing aneuploidy by mitosis collapse and delay in budding. Overall, because stress from molecular agents and environmental conditions can act directly upon the DNA molecules, or cellular components in direct interaction with the DNA molecules, they can be used to elevate stochastic DNA damages within a cell and investigate genome stability.

ii. Inducible DNA-DSB system

The use of genotoxic compounds allows to augment the levels of stochastic DNA breaks and lesions throughout the genome. Pioneer studies relied on the appearance of spontaneous LOH events, using a combination of the latter compounds, to better understand the underlying molecular mechanisms of DNA repair. However, the engineering of controlled and directed DNA-DSBs propelled the knowledge and understanding of DNA repair mechanisms, by allowing investigation of factors influencing the choice of repair pathways and the effects of mutation burdens. The baker's yeast *S. cerevisiae* has long been an important model organism in this domain, where numerous molecular tools have been put in place using various nucleases including the homothallic switching endonuclease (HO), transcription activator-like effector nucleases (TALEN), the intron-encoded endonuclease I-*Sce*I and the CRISPR-*Cas*9 (section F.1.ii-iii p.48) (Plessis *et al.* 1992; Sugawara and Haber 2012; Mosbach *et al.* 2018; Lemos *et al.* 2018). In *C. albicans*, a few years back, Feri *et al.* have developed an inducible locus-specific DNA-DSB system utilizing the *S. cerevisiae* mega-endonuclease I-*Sce*I. Under the control of the Tet-On promoter, I-*Sce*I is expressed in presence of any derivatives of tetracycline and generates a DNA-DSB at the 18 bp-long I-*Sce*I target site (ATTACCCTGTTATCCCTA). Locus specific DNA-DSBs can be achieved because this target sequence is not natively found in the genome of *C. albicans* (Figure 17A). The coupling of this system with and LOH reporter system (BFP/GFP LOH reporter system presented below in section F.2.ii p.53) allowed to analyze and better understand how *C. albicans* deals with DNA-DSB, a topic that is further discussed in the DNA repair section (section G p.59). Overall, the implementation of such a system allows to enrich rare genomic rearrangements (mainly LOH) (Figure 17B), permitting to study the effects of various growth

conditions and genetic backgrounds (including KO/OE (overexpression) mutants) on genome stability in *C. albicans*.

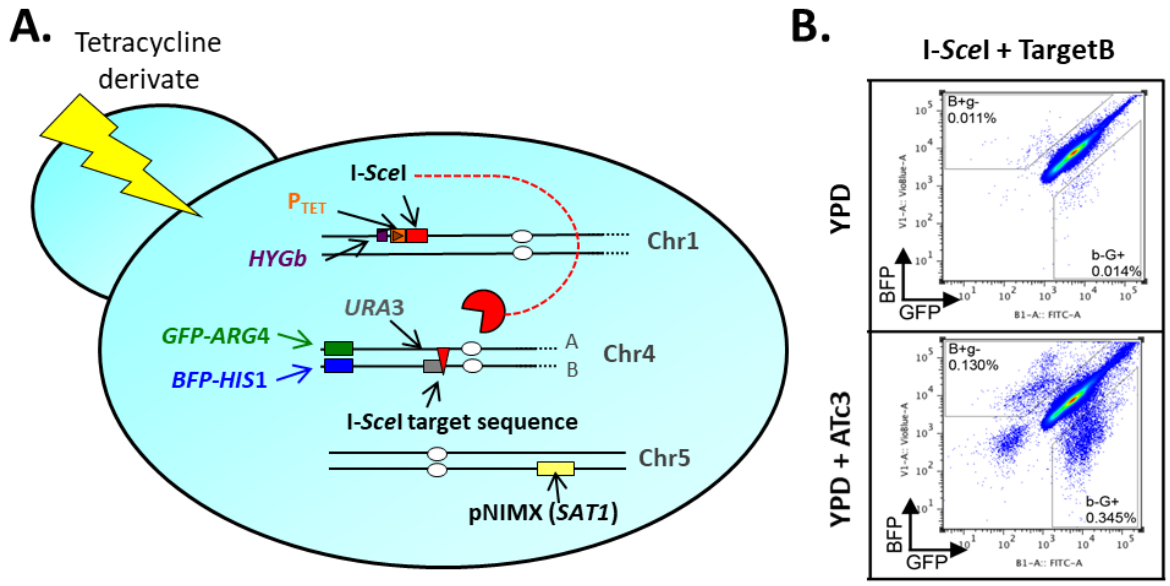

Figure 17: An inducible DNA-DSB system coupled to the BFP/GFP LOH reporter system.
A. Addition of a tetracycline derivative activates the Tet-O promoter (P_{TET}) resulting in expression of the I-*Sce*I gene. Upon I-*Sce*I expression, a DNA-DSB on the left arm of Chr4B is achieved at I-*Sce*I target sequence. Coupling of the inducible DNA-DSB system to the BFP/GFP LOH reporter system allows to monitor appearance of long-tract LOH by flow cytometry. **B.** Flow cytometry profiles obtained in non-inducible (YPD) and DNA-DSB inducible (YPD + ATc3) conditions. Efficiency of the system is illustrated by a significant increase in mono-fluorescent cells upon I-*Sce*I induction. Figure interpreted and adapted from Feri *et al.* (Feri *et al.* 2016).

iii. CRISPR-Cas9

CRISPR-Cas9 is more commonly used to generate genetic modifications and has revolutionized the genome editing process in a multitude of organisms. Initially described in *Streptococcus pyogenes*, the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and its endonuclease Cas9 act as a bacterial defense system against non-self DNA (Barrangou *et al.* 2007). A single guide RNA (sgRNA) directs the cleaving activity of Cas9 to a specific genomic site. This sgRNA is composed of a recognition motif of 20 bp directly followed by a protospacer-adjacent motif sequence (PAM), with the NGG sequence (Figure 18) (Jinek *et al.* 2012). Bacterial protection is achieved by recognition and induction of DNA-DSB in the non-self DNA, which is then degraded (Barrangou *et al.* 2007). In *C. albicans,* the CRISPR-Cas9 system has been exploited to serve as a genome editing tool, facilitating the strain engineering process. The ability to target site-specific DSBs during genetic manipulations favors the occurrence of the desired genetic modification by inducing the DNA damage-response and DNA repair pathways at the target locus, rather than relying on

stochastic DNA-DSBs occurring genome-wide during the classical transformation process (non-CRISPR) .

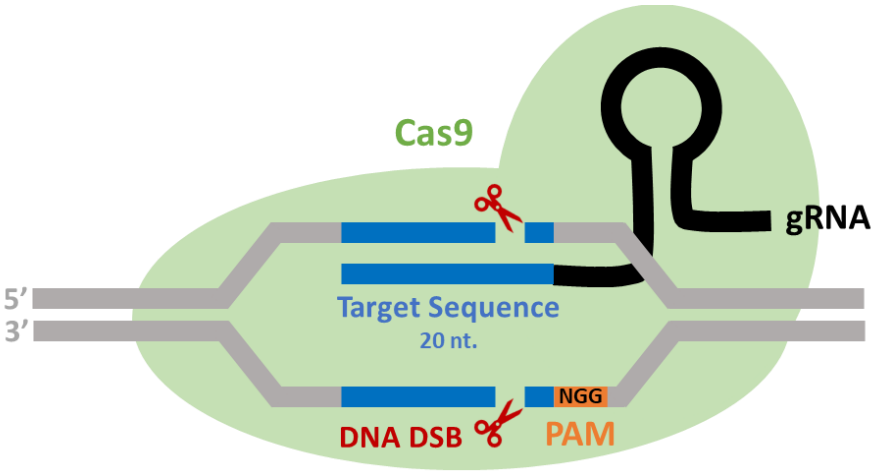

Figure 18: CRISPR-Cas9-mediated DNA double-strand break (DSB).
The Cas9 nuclease (green) is targeted to the DNA (gray) at the locus of interest by the guide RNA (gRNA, black), containing a 20 nucleotide (nt.) homologous sequence (target sequence, blue). The gRNA pairs to the target sequence directly adjacent to the PAM motif (NGG) and bluntly cleaves the DNA (DNA-DBS) roughly 3 bp upstream of the PAM, illustrated by red scissors.

CRISPR-Cas9 technology has been implemented in mammalian cells, plants, bacteria, and fungi, including *C. albicans,* and continues to be adapted to an increasing number of organisms. In diploid organisms, such as *C. albicans,* this technique has facilitated the construction of mutants, where both alleles may be modified at once (Vyas *et al.* 2015). Due to its novelty, we do not yet know all the repercussions of the CRISPR-Cas9 system in terms of its impact on the cell biology and genome stability in *C. albicans*. In *S. cerevisiae*, the integration and constitutive expression of Cas9 has been associated to cell toxicity as illustrated by lowered fitness of the cells and slower growth. Additionally, toxicity of constitutive Cas9 and sgRNA coexpression has also been observed by means of a lower transformant yield (DiCarlo *et al.* 2013). To limit these effects, a transient system has been developed (Min *et al.* 2016) while others have managed to isolate and directly use RNA/Cas9 protein complexes (Ramakrishna *et al.* 2014; Grahl *et al.* 2017) to circumvent constitutive Cas9 and sgRNA expression. To facilitate the engineering of clinical strains and to limit the amount of exogenous DNA integrated within the genome, recyclable systems have also been put in place, relying on recombination between flanking repeat sequences to excise the CRISPR-Cas9 components (Huang and Mitchell 2017; Nguyen *et al.* 2017). Furthermore, off-target DNA-DSBs by Cas9 have been reported (Fu *et al.* 2013; Tsai *et al.* 2015; Banakar *et al.* 2019), which may lead to serious biological repercussions, highlighting the importance of precautions.

Apart from cell transformation, other applications of the CRISPR-Cas9 system have been implemented. As mentioned in the previous section, CRISPR-Cas9 has also been used to induce locus-specific DNA-DSBs in cells in order to study and compare the molecular mechanisms involved in DNA repair between organisms (Vyas *et al.* 2018). Additionally, dead Cas9 (dCas9) proteins have also been engineered which, upon pairing with the appropriate sgRNA, efficiently identify and bind target sequences but are incapable of inducing DNA-DSBs. When targeting promoter regions, this allows to bind and block transcription, which results in lowered expression of a gene of interest. Indeed, somewhat comparable to knockdown systems, different expression levels of a specific gene can be obtained depending on the sgRNA paired with the dCas9 (Wensing *et al.* 2019). Moreover, though not yet implemented in *C. albicans,* the highly versatile CRISPR-Cas9 system can be adapted to numerous applications, i.e. DNA nicks (SSB), labelling of DNA, various nuclease activities, recombinases inducing genomic rearrangements or scaffolding leading to multiprotein and nucleic acid complexes (Figure 19) (Mali *et al.* 2013).

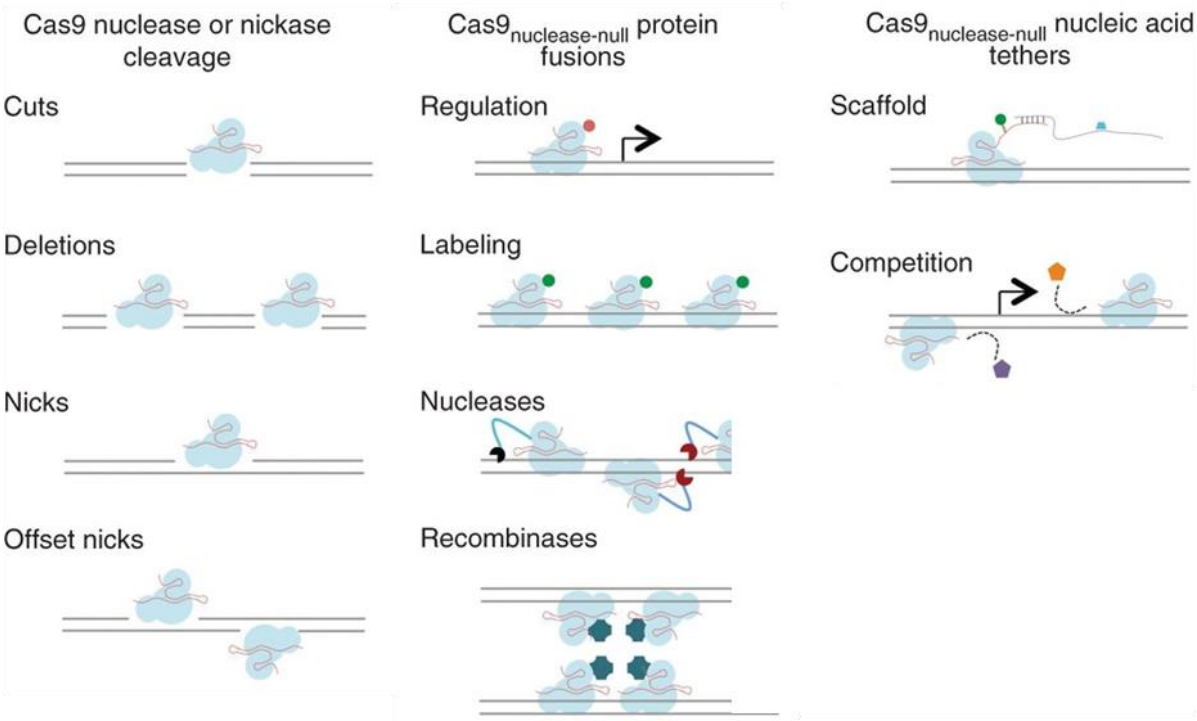

Figure 19: Application of the CRISPR-Cas9 system.
Diverse potential applications of Cas9 resulting from biological engineering. Implementation of genome editing tools, permitting locus specific single- and double-strand DNA breaks. Various Cas9-protein fusions allowing gene regulation, DNA labelling using fluorophores, directing various nuclease activities, or programming of genomic rearrangements. The Cas9/RNA complex may also be modified and adapted to facilitate/promote scaffolding, leading to multiprotein and nucleic acid complexes, or used in competition assays e.g. affinity assay between DNA and RNA. Figure adapted from (Mali *et al.* 2013).

2. Tools for monitoring LOH

i. *GAL1*/*URA3* system

The *GAL1*/*URA3* system relies on the principle of auxotrophic marker counter-selection. This system utilizes both the native *C. albicans GAL1* gene, encoding for a galactokinase implicated in galactose metabolism (located on the left arm of Chr1), and the commonly used *URA3* auxotrophic marker, encoding for a orotidine-5'-phosphate decarboxylase involved in the biosynthesis of uridine nucleosides. In an *ura3*$^{-/-}$ strain, a heterozygous locus is generated at the native *GAL1* locus by the replacement of a *GAL1* copy with the *URA3* gene, with the cells still capable of metabolizing galactose and synthesizing uridine. In order to identify cells which have undergone an LOH event at the *GAL1* locus within the population of interest, serial dilutions are plated on media containing 2-deoxygalactose (2-DG) or 5-fluoroorotic acid (5-FOA), respectively (Figure 20A). Indeed, cells possessing at least one copy of the *URA3* gene are able to initiate uridine synthesis and transform 5-FOA in the 5-fluorouracil toxic metabolite leading to cell death (Boeke *et al.* 1984). Thus, the 5-FOA resistant (5-FOA^{R}) colonies are indicative of individuals which have undergone an LOH and have lost the *URA3* marker. Using the same principal, LOH events involving the loss of *GAL1* can be identified using 2-DG, as GAL^{+} cells metabolize 2-DG into toxic 2-deoxyglucose (Gorman *et al.* 1992) (Figure 20B). The *GAL1*/*URA3* system is frequently used in *C. albicans* to assess effects of KO mutants or diverse growth conditions on genomic stability (Legrand *et al.* 2007; Forche *et al.* 2011; Ciudad *et al.* 2020). In other studies, LOH frequencies are assessed by integrating *URA3* on a single homolog at different genomic loci and counter-selecting for unidirectional events of LOH, limiting the study of LOH events/frequencies to a single haplotype (Forche *et al.* 2011). Additionally, because cells lacking a functional *URA3* gene have been shown to be avirulent, *in vivo* unidirectional LOH frequencies may still be calculated by using only the *GAL1* marker and monitoring the appearance of 2-DG resistant (2-DG^{R}) cells, upon the replacement of a native *GAL1* allele by the non-counter-selectable *ARG4* or *HIS1* auxotrophic markers (Forche *et al.* 2003).

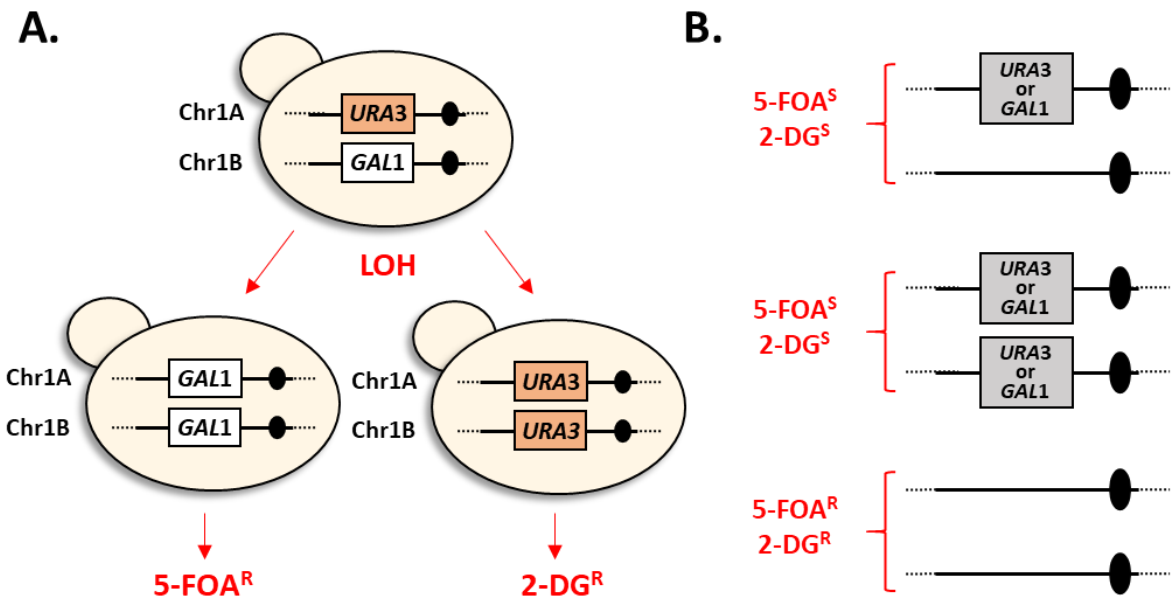

Figure 20: Monitoring LOH by counterselection.
A. Counterselection of auxotrophic markers *URA3* or *GAL1* on 5-fluorrotic acid (5-FOA) or 2-deoxygalactose (2-DG)- containing medium, respectively. The presence of *URA3* or *GAL1* allows metabolization of 5-FOA or 2-DG, respectively, into cell toxic compounds leading to cell death. **B.** *GAL1*/*URA3* system used to monitor genome stability, by measurement of LOH events on Chr1 in *C. albicans* strains. Black ovals: centromere, ^{S}: sensitive and ^{R}: resistant

ii. BFP/GFP LOH reporter system

Counter-selection of auxotrophic markers is quite simple to use with a small number of strains and efficient in terms of identifying and quantifying cells that have undergone LOH at the target locus. Nevertheless, plating and counting of colony forming units (CFUs) can quickly become a cumbersome task when studying multiple strains and/or genomic loci. Thus, another LOH reporter system has been developed to circumvent serial dilutions, plating, the incubation period and CFU count, by taking advantage of flow cytometry (Loll-Krippleber *et al.* 2015). The BFP/GFP LOH reporter system involves the engineering of an artificial heterozygous locus harboring fluorescent proteins. Independently of the locus, the blue fluorescent protein-encoding gene (*BFP*) is integrated on one homolog whereas the green fluorescent protein-encoding gene (*GFP*) is integrated on the second homolog, rendering cells double fluorescent. The fluorescent proteins can be associated to either auxotrophic markers or drug resistance markers to allow selection of cells that have integrated the system. Upon LOH, the cells will become mono-fluorescent due to the loss of either the BFP- or GFP-encoding genes (Figure 21). Using flow cytometry, LOH arisen within cell populations can be precisely quantified in the matter of minutes, with thousands of cells being screened per second for fluorescence status. Quantification of LOH events can be routinely assessed with a benchtop cytometer, while populations of interest may be sorted and further analyzed using

a flow cytometer allowing physical cell sorting. The principal advantages of the BFP/GFP LOH reporter system are (i) the fast analysis and data recovery, (ii) no locus restrictions, (iii) evaluation of LOH events in absence of selection and (iv) flexibility of auxotrophic markers used allowing LOH events to be studied also in clinical prototroph strains. Additionally, this system can be exploited to enrich for long tract LOH events by placing the BFP/GFP reporter system in a telomere proximal locus, therefore maximizing the number of detectable LOH events for the given chromosome arm (Loll-Krippleber *et al.* 2015). This tool rendered high-throughput genome stability screens possible. The BFP/GFP LOH reporter system was used to screen a collection composed of 124 *C. albicans* inducible OE strains, each one overexpressing an orthologous ORF of a gene involved in DNA repair, replication or recombination in *S. cerevisiae.* Such screens permit identification of genes whose OE results in an increase in LOH events, key player in *C. albicans* genome stability (Loll-Krippleber *et al.* 2015).

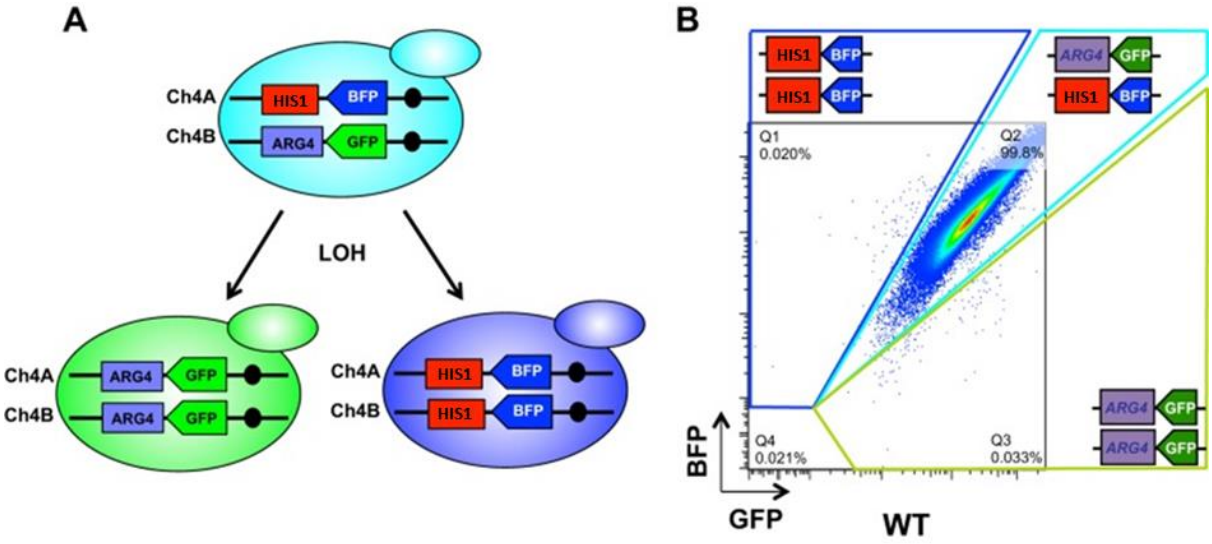

Figure 21: The BFP/GFP LOH reporter system.
A. Schematic representation of the BFP/GFP LOH reporter system integrated on the left arm of Chr4, using BFP associated to the histidine auxotrophic marker and GFP associated to the arginine auxotrophic marker. Heterozygous cells are: BFP and GFP positive and, histidine prototroph and arginine prototroph. Mono-GFP cells have undergone loss of the BFP-*HIS1* bearing homolog, and are GFP-positive and BFP-negative, histidine auxotroph and arginine prototroph. Mono-BFP cells, have undergone loss of the GFP-*ARG4* bearing homolog, and are GFP-negative and BFP-positive, histidine prototroph and arginine auxotroph. **B.** Flow cytometry output of 10^6 cells with appropriately drawn gates delimiting each cell population; double-fluorescent (heterozygous), mono-BFP and mono-GFP. Figure adapted from (Loll-Krippleber *et al.* 2015)

3. Tools used to characterize gross chromosomal rearrangement events

Most of the tools presented above allow to promote genomic instability and to select and quantify LOH appearance. Further characterization of these GCRs relies on several techniques that are routinely used to evaluate the extent of the genome changes. In this section, I present the general approaches used to confirm and characterize the occurrence of GCR events throughout the chapters of my thesis.

i. Whole-genome sequencing

The advent of sequencing and the affordability of conducting large scale sequencing projects have vastly changed the way we conduct research nowadays. Whole-genome sequencing (WGS) is a commonly used approach to identify, characterize, and compare microorganisms of interest, with regards to various biological aspects including genomic rearrangements. WGS is a highly informative approach because we can detect both single nucleotide variations (SNVs) and copy number variations (CNV). SNVs represent variations involving a single nucleotide (SNP or indels) while CNV are structural variation, duplication or deletion, encompassing numerous bases within the genome. A CNV can be defined as a region of the genome which is normally found in multiple copies and the copy number varies between genomes of different strains. In a diploid heterozygous microorganism like *C. albicans*, CNV would be identified by either the loss of a copy of the genome (deletion), resulting in a single copy, or the gain of additional copies, resulting in 3 or 4 copies of the region of interest.

Indeed, CNV analysis permits the unveiling of aneuploidy events across the genome. By investigating the sequencing depth across the genome, or genome coverage, aneuploidies may be identified as genomic regions possessing 0.5x (reductional aneuploidy), 1.5x (amplifying aneuploidy) or 2x (amplifying aneuploidy) the average amount of reads throughout the genome (Figure 22A). Because *C. albicans* possesses predominantly a diploid heterozygous genome, the analysis of sequencing data is conducted on both haplotypes independently. Indeed, aneuploidy events may also be identified using the allele balance at heterozygous sites (ABHet) with a value between 0 and 1. The ABHet is calculated as the number of reference reads from individuals with heterozygous genotypes divided by the total number of reads from such individuals. An ABHet value of 0.5 is the expected value for a diploid genome. A triploid strain may contain either three identical alleles (allelic frequency of 1) or two identical alleles and one different allele (frequency 0.66 and 0.33) and a tetraploid strain may have allelic frequencies of 0.5 (2x2 identical alleles) or 1 (4 identical alleles) or 0.25

and 0.75 (3 identical alleles and 1 different allele). Certainly, LOH events may also be identified using this method, as they are defined by an absence of an ABHet value (Figure 22B).

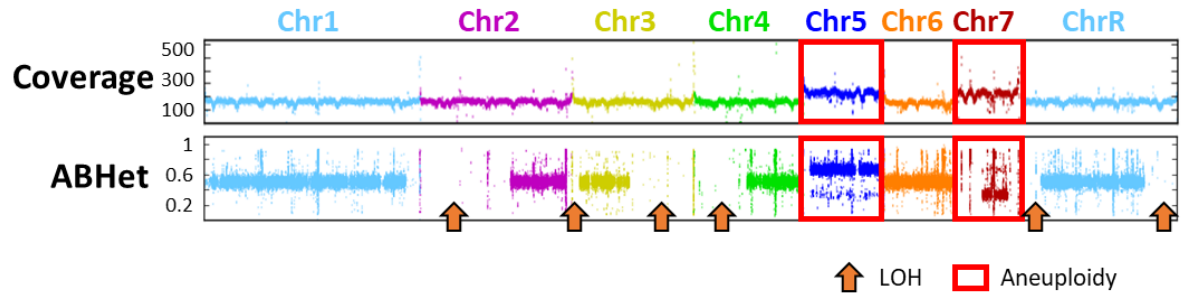

Figure 22: Identification of gross chromosomal rearrangements using sequencing data. Coverage graph representing the sequencing depth across the genome of a *C. albicans* strain. Copy number variations (CNV), aneuploidy, can be observed on Chr5 and Chr7 as they possess an increased coverage. Allele balance at heterozygous sites (ABHet) is the number of reference reads from individuals with heterozygous genotypes divided by the total number of reads from such individuals. Both Chr5 and Chr7 possess ABHet values of 0.66 and 0.33, respectively, indicative of trisomy (3 copies) whereas the other chromosomes possess ABHet values of 0.5, a diploid state. Major LOH events are illustrated by an absence of ABHet value and highlighted by the orange arrows. Aneuploidies are highlighted by red squared.

SNP analyses are also useful for identifying and delimiting LOH events. Mapping of heterozygous SNPs density within a defined bin across the genome also permits to identify large region of LOH, as illustrated in Figure 13 from the heterozygosity section (D.4 p.33). Nevertheless, short LOH events can be identified between strains by comparison of SNP calling data, where we can identify the lost haplotype. Such analyses, based on CNV analysis and SNP calling of various sequenced strains followed by strain to strain comparisons, have been very informative with regards to the identification of newly appeared GCRs, their localization and size distributions in several strains (clinical and laboratory) and growth conditions including, *in vitro* growth conditions, various murine colonization (oropharyngeal and GIT colonization) and infection models (Hirakawa *et al.* 2015; Ropars *et al.* 2018; Ene *et al.* 2018; Wang *et al.* 2018; Todd *et al.* 2019; Sitterle *et al.* 2019; Forche *et al.* 2019).

ii. Characterization through auxotrophic spotting and fluorescent confirmation

Most genetic rearrangement reporter systems are based on the presence/absence of auxotrophic or fluorescence markers (sections F.2 p.52). Thus, confirmation of auxotrophic status is primordial in order to confirm the occurrence of a genetic modification at the locus of interest before further characterization. This is conducted by means of spot assays on various minimal media with appropriate dropout amino acids, where presence or absence of colony proliferation is assessed. For instance, the BFP and GFP fluorescent markers, in the BFP/GFP LOH reporter system, are be associated with the histidine (*HIS1*) and arginine (*ARG4*)

auxotrophic markers (Figure 21) thus, their presence/absence may be tested by the latter laboratory experiment. Theoretically, upon LOH the mono-GFP sorted cells should proliferate on arginine-deprived medium and not grow on histidine-deprived medium, as they have lost the BFP-*HIS1* construct. Contrarily, mono-BFP cells should proliferate on histidine-deprived medium and not grow on arginine-deprived medium, as they have lost the GFP-*ARG4* construct (Figure 21). Certainly, these cells may also undergo a fluorescence check by cytometry, to confirm the presence or absence of the BFP and/or GFP, theoretically reiterating the conclusions of the spot assay. This double confirmation of marker loss permits to confidently conclude the occurrence of an LOH event, by revoking the possibility of the accumulation of a loss of function mutation within the components of the LOH reporter system. The presence of an LOH event which had been counter-selected with the *GAL1*/*URA3* LOH reporter system can also be confirmed by spot assay. Colony forming units (CFUs) obtained on 5-FOA or 2-DG containing media can be spotted on uridine-deprived and galactose-containing media in order to confirm the loss of either *URA3* or *GAL1*, respectively. Indeed, these verification steps also solidify the quantification of LOH events by assessing false positive frequency, i.e. LOH-independent loss of marker incidence.

iii. SNP-RFLP

As the size of LOH is thought to be correlated with its biological impact, it is of interest to study the length of LOH tracts. To do so the allelic status is often evaluated at different loci along the chromosome of interest using the SNP-RFLP (single nucleotide polymorphism – restriction fragment length polymorphism) technique. SNP-RFLP is a quick and efficient tool often used in *C. albicans,* first described by Forche *et al.* (Forche *et al.* 2009b). This tool relies on the presence of heterozygous SNPs affecting a restriction site on only one haplotype. The experimental procedure consists of an initial polymerase chain reaction (PCR) amplification step followed by an enzymatic digestion and relieving of profiles by gel electrophoresis. Regardless of SNP location there are three possible digestion profiles: (i) one non-digested band, meaning both alleles correspond to the enzyme-resistant haplotype, (ii) a profile composed of two smaller bands representing an enzyme-sensitive homozygous profile and (iii) the presence of three bands where one band represents a non-digested profile and the second and third bands represent a digested profile, hence a heterozygous profile (Figure 23). Among the numerous existing SNP-RFLP candidate loci found throughout the SC5314 *C. albicans* genome, Forche and collaborators (Forche *et al.* 2009b) optimized 32 SNP-RFLP loci spread amongst the 16 chromosomal arms for straightforward use by the *C. albicans* community. To be noted, not all *C. albicans* strains share the same SNP-RFLP loci due to the numerous nucleotide differences. However, if genome sequencing data is available for the

clinical strains of interest, the technique can easily be implemented as SNP-RFLP sites can be identified. To do so, a nucleotide multiple sequence alignment is first required while candidate heterozygous SNPs are selected using the following criteria: (i) interrupts a commonly known restriction site on one haplotype, and (ii) selected enzyme does not cut again within a range of roughly 1 kb. Ideally, selected heterozygous SNPs should also be identified in closely related *C. albicans* strains. If the heterozygous SNP is found in multiple strains, it is most likely to be a true SNP and not a sequencing artefact. This can be easily investigated due to the numerous publicly available *C. albicans* genome sequences. Then, primer pairs may be designed and tested permitting to properly visualize the three SNP-RFLP profiles (Figure 23) upon PCR amplification and digestion. By monitoring of the allelic status in multiple loci across a chromosome, LOH size and consequently the molecular mechanism associated to the LOH can be identified.

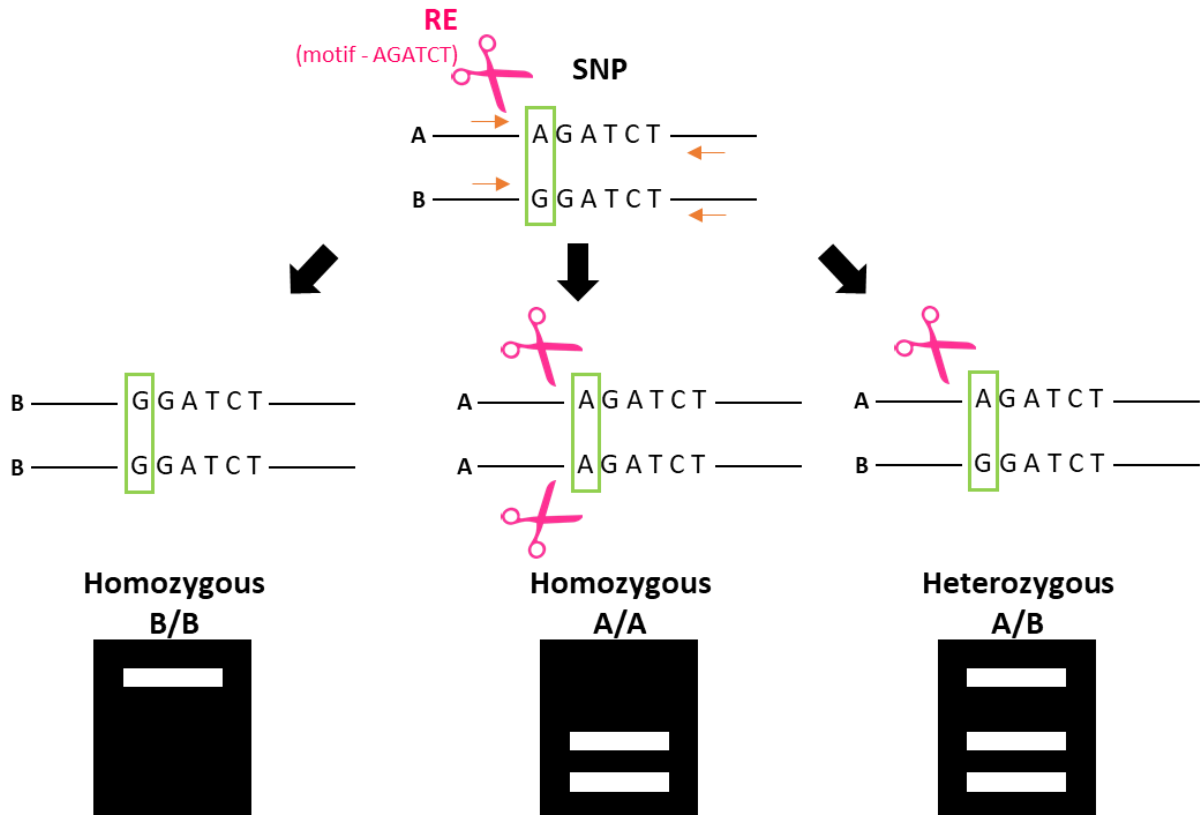

Figure 23: Identification of allelic status by SNP-RFLP.
This tool relies on the presence of heterozygous SNP (green box) affecting a restriction site on only one haplotype. An initial PCR step allows the amplification of the locus of interest (primers, orange arrows) which is then digested using the appropriate restriction enzyme (RE, prink) recognizing a specific restriction motif. The digestion products are separated using gel electrophoresis with three possible outcomes, (i) non-digested profile (one band) indicating a homozygous profile of the RE-resistant allele, (ii) a digested profile (two bands) indicating a homozygous profile of the RE-sensitive allele and (iii) a mixed non-digested and digested profile (3 bands) indicating a heterozygous allelic profile composed of both RE-resistant and -sensitive alleles.

G. DNA breaks and DNA repair mechanisms: Sources of genomic rearrangements

As previously highlighted throughout the introduction, DNA damages are inherent to a cell and cells must efficiently identify, assess, and repair these attacks in order to maintain genetic integrity and ensure their livelihood. These DNA lesions may originate from exogenous factors though cells can also intentionally generate breaks during processes of DNA unpackaging and DNA repair. Although DNA repair should result ideally in little or no alteration of the genetic material, DNA breaks are often sources of GCRs (as presented in section E). It is still unclear if the resulting GCRs are intentional or unintentional outcomes of the repair. This brings forth the question whether the high genomic plasticity of certain species, including *C. albicans,* is a desired evolved feature or an evolutionary consequence burdening/challenging this species existence. In the following section, we will discuss why and how DNA lesions are major sources of GCRs via various DNA repair mechanisms.

Repair mechanisms of both single- and double-strand DNA lesions will be addressed primarily based on knowledge obtained from the model organism *S. cerevisiae*. The impact of each molecular mechanism on generating genetic variation and the current standing in *C. albicans* will be discussed. This section will first tackle how DNA irregularities impacting only one or few nucleotides are repaired. Of importance, these aberrations can occasionally physically influence the DNA helix by causing helix-distortion, leading to single-strand DNA breaks. Independently, DNA lesions can also sever both DNA strands causing severe damages to DNA integrity and cell viability as they often involve large fragments of DNA. Thus, we will also address the numerous DNA-DSB repair mechanisms.

1. Excision and mismatch repairs

i. Base excision repair (BER)

The base excision repair (BER) mechanism detects and repairs non-helix-distorting damaged bases, the majority of DNA damages occurring within a cell. These damages include oxidized, alkylated and deaminated bases, often arising from exposure to environmental factors such as ROS and UVs. BER is initiated by the recognition and hydrolysis of a damaged base by a DNA *N*-glycosylase, resulting in an abasic site. Various DNA *N*-glycosylase exist, each one recognizing a specific type of base damage (e.g. in *S. cerevisiae*: Ntg1, Ntg2, Ogg1…). The repair of the apurinic/apyrimidinic (AP) site is conducted using one of the two BER subpathways, short-patch BER (SP-BER) or long-patch BER (LP-BER), involving excision of either a single or several nucleotide/s, respectively (Bauer *et al.* 2015). The SP-BER subpathway requires the sequential action of an AP lyase and an AP endonuclease to cleave the DNA backbone on the 3' and 5' sides of the lesion, respectively, leading to a clean 1-nt gap. This gap is then filled by DNA polymerase and DNA ligase activities (Dianov *et al.* 2001) (Figure 24). In contrast, the LP-BER subpathway processes the AP site with the 5' cleavage activity of an AP-endonuclease, encoded by the *APN1* and *APN2* genes in *S. cerevisiae*, generating substrate for DNA polymerase (Doetsch and Cunningham 1990). Subsequently, the DNA polymerase, Pol4 in *S. cerevisiae,* fills the gap and its synthesis activity continues several bases downstream (over a total distance of 2-8 nt.) resulting in the formation of a DNA flap by displacement of DNA located 3' of the abasic cleavage site. The DNA flap is removed by the activity of the Rad27 flap endonuclease and a DNA ligase seals the remaining nick (Fortini *et al.* 2003) (Figure 24).

In *C. albicans*, only the effect of homozygous deletion of *S. cerevisiae* orthologs *NTG1* and *OGG1* (DNA *N*-glycosylases) and *APN1* (AP-endonuclease) on BER was investigated by Legrand *et al.* (Legrand *et al.* 2008). Upon characterization of mutant strains, the authors reported no sensitivity to oxidizing agents (major cause of BER substrates) and no altered responses to macrophages or various antifungal compounds. Additionally, KO of previously mentioned genes did not affect mutation levels in *C. albicans* as tested on Chr1 using the *GAL1*/*URA3* system (Legrand *et al.* 2008). These results suggest genetic redundancy within BER and/or, between BER and other excision repair mechanisms in *C. albicans*.

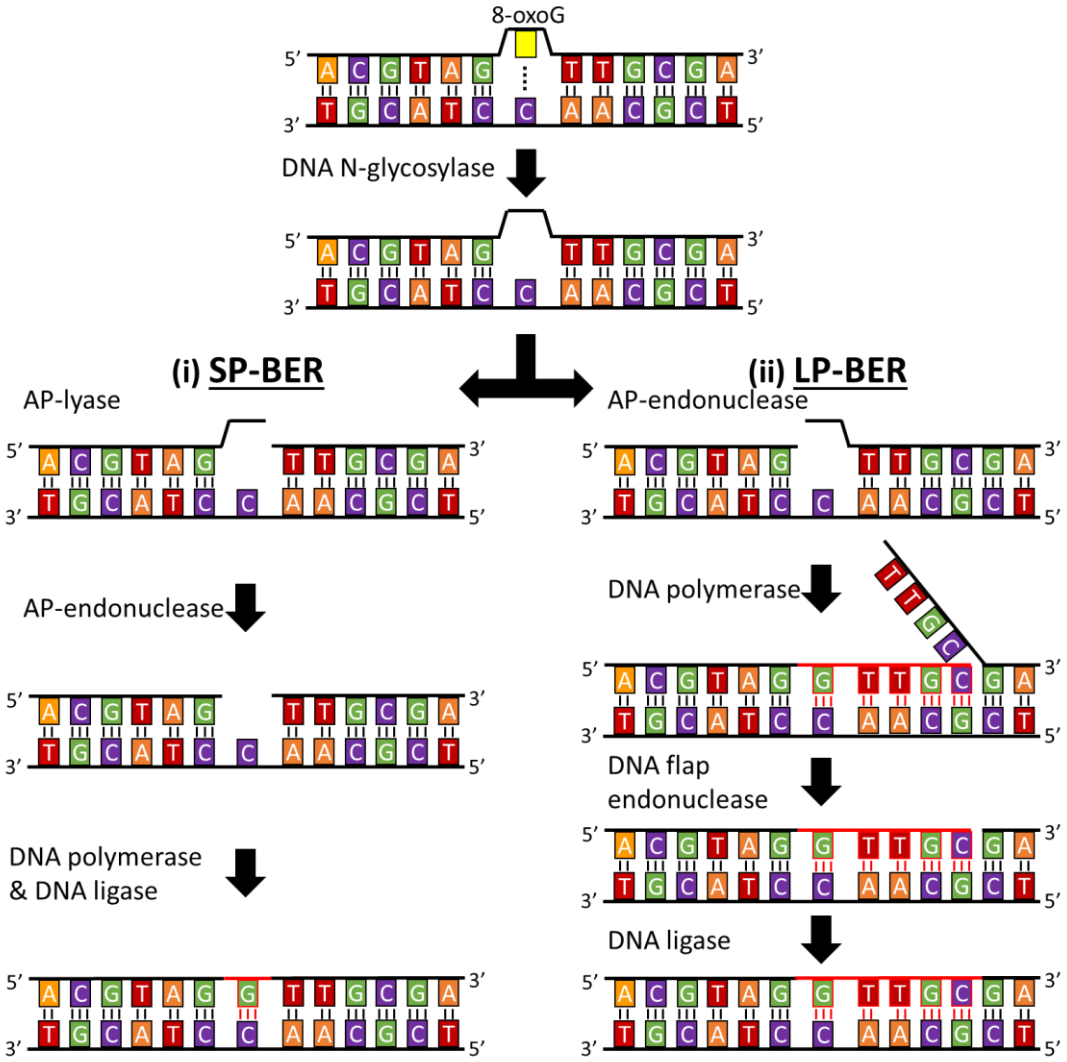

Figure 24: Base excision repair (BER).
Repair of an oxidized guanine base (8-oxoguanine) using the BER molecular mechanism. Initiation of BER and hydrolysis of damaged base is conducted by DNA *N*-glycosylase. The resulting abasic site can be either repaired through (i) short-patch BER (SP-BER) or (ii) long-patch BER (LP-BER) subpathways. (i) In SP-BER, AP-lyase and AP-endonuclease cleave the DNA backbone 3' and 5' of the abasic site, respectively. The generated gap within the backbone is filled by DNA polymerase and DNA ligase. (ii) During LP-BER, the AP-endonuclease cleaves the DNA backbone 5' of the abasic site allowing DNA synthesis by DNA polymerase. Polymerase activity (encompassing 2-8 nt.) fills in the gap at the abasic site and generates a DNA flap by displacement of DNA 3' of gap. The DNA flap is removed by DNA flap endonuclease and DNA ligase restores integrity of DNA backbone. Repaired and newly synthesized DNA is indicated in red.

ii. Nucleotide excision repair (NER)

The nucleotide excision repair (NER) mechanism is implicated in the removal of base damages leading to distortion of the DNA molecule, intra- or inter-strand crosslinks, often induced by UV light (e.g. pyrimidine dimers). This type of damage is removed by generating incisions on both sides of the DNA lesion resulting in the removal of an oligonucleotide (25-30 nt. long) containing the lesion (Huang *et al.* 1992). Studies in *S. cerevisiae* have described

that transcription-coupled NER activity is faster than NER occurring in non-transcribing regions (Hanawalt 1991) thus, two types of NER have been described (i) transcription-coupled (TC) and (ii) global genome (GG) (Li 2015). Indeed, actively transcribing genomic loci adopt a permissive DNA structure, facilitating the access of the DNA repair molecules to the lesion site. Though UV also leads to changes in DNA conformation/packaging in non-actively transcribed genomic regions, DNA remains rather condensed and less accessible to the DNA repair machinery (Smerdon and Lieberman 1978). Therefore, damage repair in latter regions via NER is less efficient and often requires chromatin remodeling (though the exact mechanism is not yet fully comprehended (reviewed in (Mao and Wyrick 2019))). Differences in chromatin structure certainly can explain why identification of DNA lesions is divergent between the two NER types. In *S. cerevisiae,* GG-NER recognizes helix-distorting damages via the Rad4-Rad23-Rad33 complex leading to recruitment of the TFIIH transcription factor complex (usually required for the initiation of transcription at RNA polymerase II promoters), promoting open structure and bubble formation at the DNA lesion site. In contrast, initiation of TC-NER is achieved via blocked RNA polymerase due to a helix-distorting DNA lesion while Rad16 or Rpd6 proteins (TC-NER specific) or the Rad4-Rad23-Rad33 complex recognize DNA lesions (Boiteux and Jinks-Robertson 2013). Although initial recognition of DNA lesion differs between TC-NER and GG-NER, the recruitment of pre-incision complex, removal of DNA damage and repair of gap are very similar between the two pathways. In both cases, DNA lesion recognition triggers recruitment of the remaining pre-incision complex components, RPA stabilizing proteins (stabilizing single-strand DNA) and the Rad14 protein (implicated in UV damage recognition). The pre-incision complex then recruits the Rad1-Rad10 complex and Rad2 in order to generate incisions in the DNA backbone, 5’ and 3’ of the DNA damage site, respectively (Habraken *et al.* 1993; Davies *et al.* 1995). Upon removal of the excised oligonucleotide, the gap is mended by a DNA polymerase while the DNA backbone nicks are sealed by the Cdc9 ligase (Boiteux and Jinks-Robertson 2013; Mu *et al.* 2018) (Figure 25).

In *C. albicans*, NER is vastly understudied, Legrand and collaborators investigated the role of single-strand DNA endonucleases Rad2 and Rad10 (member of the Rad1-Rad10 DNA flap removal complex). The authors found that both homozygous deletion mutants are, as expected, highly sensitive to UV light exposure. Additionally, it seems that both endonucleases do not play an important role in genomic stability, as their deletion does not affect stability of Chr1 (Legrand *et al.* 2007). Similar to BER, genetic redundancy within the NER pathway or/and, between NER and other excision repair mechanisms is likely to happen in *C. albicans.* Indeed, dysfunction of NER is often compensated by BER, and vice versa (Swanson *et al.* 1999; Boiteux and Jinks-Robertson 2013). Recently, Feng *et al.* (Feng *et al.* 2020)

investigated the role of NER-involved proteins Rad4 and Rad23 in *C. albicans* and demonstrated that both KO and OE leads to a UV sensitivity phenotype in strains. Additionally, they showed that *RAD*23 deletion is associated to *C. albicans* hypovirulence in systemic mouse model infection and in a macrophage interaction assay. However, Rad23 does not seem to play a direct role on virulence, as gene expression data revealed that *RAD23* may be regulating transcription of other virulence factors (Feng *et al.* 2020).

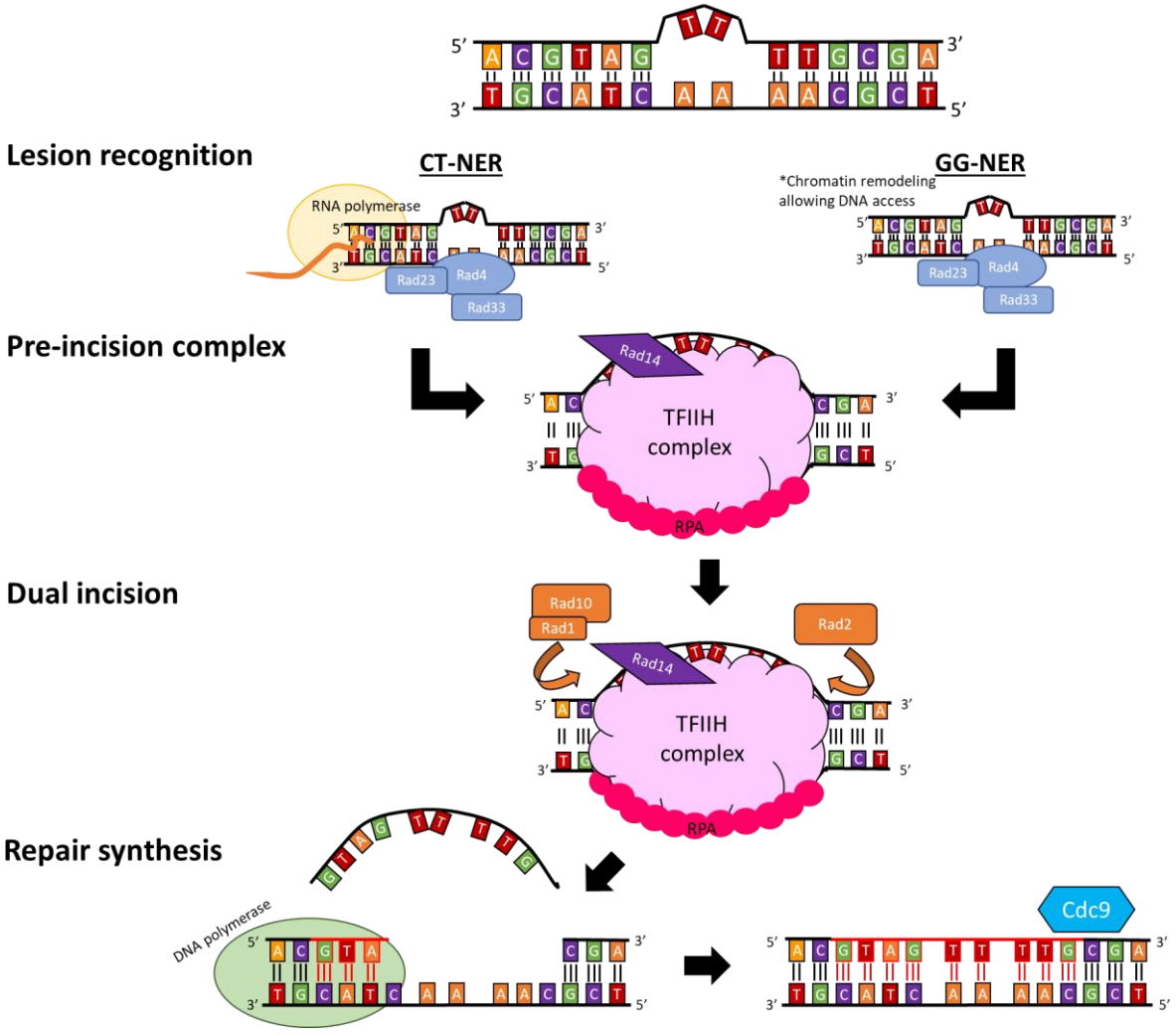

Figure 25: Nucleotide excision repair (NER).
Repair of an intra-strand crosslinked thymine dimer using the NER molecular mechanism, in *S. cerevisiae*. Recognition of DNA lesion by Rad4-Rad23-Rad33 complex is either permitted via blocked RNA polymerase by helix-distorting DNA damage (CT-NER) or chromatin remodeling (GG-NER). Once recognized, the recognition complex allows the recruitment and assembly of the pre-incision complex, composed of: TFIIH transcription factor complex (promoting open structure DNA), RPA stabilizing proteins (intact stabilizing single-strand DNA) and Rad14 protein (recognizing UV damage). The pre-incision complex will then recruit the Rad1-Rad10 complex and Rad2 protein, responsible for generating a dual incision flanking the DNA damage site. The dual incision will result in removal of a 25-30 nucleotide-long oligonucleotide. This gap is then repaired by DNA polymerase and nick in the DNA backbone is repaired by the Cdc9 ligase. Repaired and newly synthesized DNA is indicated in red.

iii. Mismatch repair (MMR)

Mismatched nucleotides within otherwise complementary paired DNA can arise from polymerase errors occurring during DNA replication or recombination and at times from certain base modifications. The mismatch repair (MMR) pathway plays an important role in maintaining the integrity of a cell's genetic material by augmenting replication fidelity and genome stability through correction of polymerase errors. Thus, the MMR machinery is spatially and temporally coupled to DNA replication and is active behind proofreading replicative DNA polymerases Polδ and Polε resulting in a 100-1,000-fold reduction in polymerase activity-associated mutations (Drotschmann *et al.* 1999; Harfe and Jinks-Robertson 2000; Healey *et al.* 2016; Boyce *et al.* 2017). Indeed, defect in the MMR pathway leads to mutation accumulation in cells (Drotschmann *et al.* 1999; Shcherbakova and Kunkel 1999; Goellner 2020) and has been associated to tumor development in humans (Peltomäki 2003). Mismatches are recognized by the MutS complex, consisting of either the Msh2-Msh6 or Msh2-Msh3 complexes in *S. cerevisiae*, perceiving base-base mismatches and 1-2 nucleotide long insertion/deletion loops by the Msh2-Msh6 complex or small to larger insertion/deletion loops by the Msh2-Msh3 complex (Hombauer *et al.* 2011). Upon damage recognition, MutS undergoes an ATP-dependent conformation change, forming a sliding clamp around the DNA (Gradia *et al.* 1999; Mendillo *et al.* 2005). Subsequently, the heterodimeric MutL complex (Mlh1-Pms1 or Mlh1-Mlh3) is recruited to the mismatch site along with the Mlh1-Mlh2 accessory complex. The DNA replication processivity of the PCNA clamp activates the endonuclease activity of MutL to generate a nick within the newly synthesized DNA backbone. This nick permits the entry of the Exo1 protein which excises mismatch-containing DNA in a 5' to 3' manner, while DNA polymerases Polδ or Polε are recruited to fix the gap (Liu *et al.* 2017; Goellner 2020) (Figure 26). Shell and collaborators have shown the intimate relationship between DNA replication and the MMR machinery, as Msh6 and the PCNA clamp are physically bound by a tether, suggesting that PCNA may be recruiting the MMR machinery in order to promote post-replication mismatch recognition and repair (Shell *et al.* 2007). Additionally, the authors showed that loss of this interaction results in a 10-15% reduction in MMR efficiency in *S. cerevisiae* (Shell *et al.* 2007). It is also likely that the MMR activity competes with the nucleosome deposition machinery, reviewed in (Goellner 2020). The deposition of nucleosomes and packaging of DNA into higher chromatin structure is also situated behind the replication fork. This activity is in part associated with the histone chaperone Caf1, chromatin assembly factor 1, which is also bound to a replication component, namely the PCNA clamp. It has been shown that MMR is partially inhibited by nucleosome deposition, especially when a mismatch is incorporated into a nucleosome, rendering MMR less efficient (Li *et al.* 2009). Furthermore, activity of Caf1 can also block MMR's sliding clamp,

inhibiting mismatch recognition and consequently its repair (Kadyrova *et al.* 2011). It has been suggested that MMR does not only occur in naked DNA environments but can also take place in environments that possess partial or completely reconstructed nucleosomes. Overall, we are still lacking a full understanding of how chromatin remodeling and MMR pathways (and potentially other excision repair pathways) interconnect to assure genetic integrity. Recent studies have focused on the latter subject including a study that used bioinformatic analyses to predict protein-protein interactions through motif analysis (Goellner *et al.* 2018). Notably, the authors identified and further confirmed, through yeast two-hybrid assay, the interaction between Fun30, a Snf2p family member with ATP-dependent chromatin remodeling activity, and Msh2. However, deletion of *FUN30* in *S. cerevisiae* results in only a slight increased mutation rate suggesting that other chromatin remodeling factors may be implicated or may assure function upon absence of Fun30, through genetic redundancy (Goellner *et al.* 2018).

In *C. albicans,* the MMR pathway has only been investigated through characterization of *msh2*$^{-/-}$ and *pms1*$^{-/-}$ strains. Deletion of either of these genes results in an increased mutation rate and is associated with an increased appearance of antifungal resistant colonies (Legrand *et al.* 2007). More recently, a study also highlighted a link between mutations found within the *MSH*2 gene and antifungal resistance in *C. glabrata* clinical strains (Healey *et al.* 2016), though these observations have been challenged by other authors (Dellière *et al.* 2016; Bordallo-Cardona *et al.* 2019). Nevertheless, mutator phenotypes associated with disruption of MMR pathways may be beneficial for pathogenic fungi through microevolution and generation of adaptive phenotypes, as supported by the work conducted in *Cryptococcus neoformans* (Boyce *et al.* 2017).

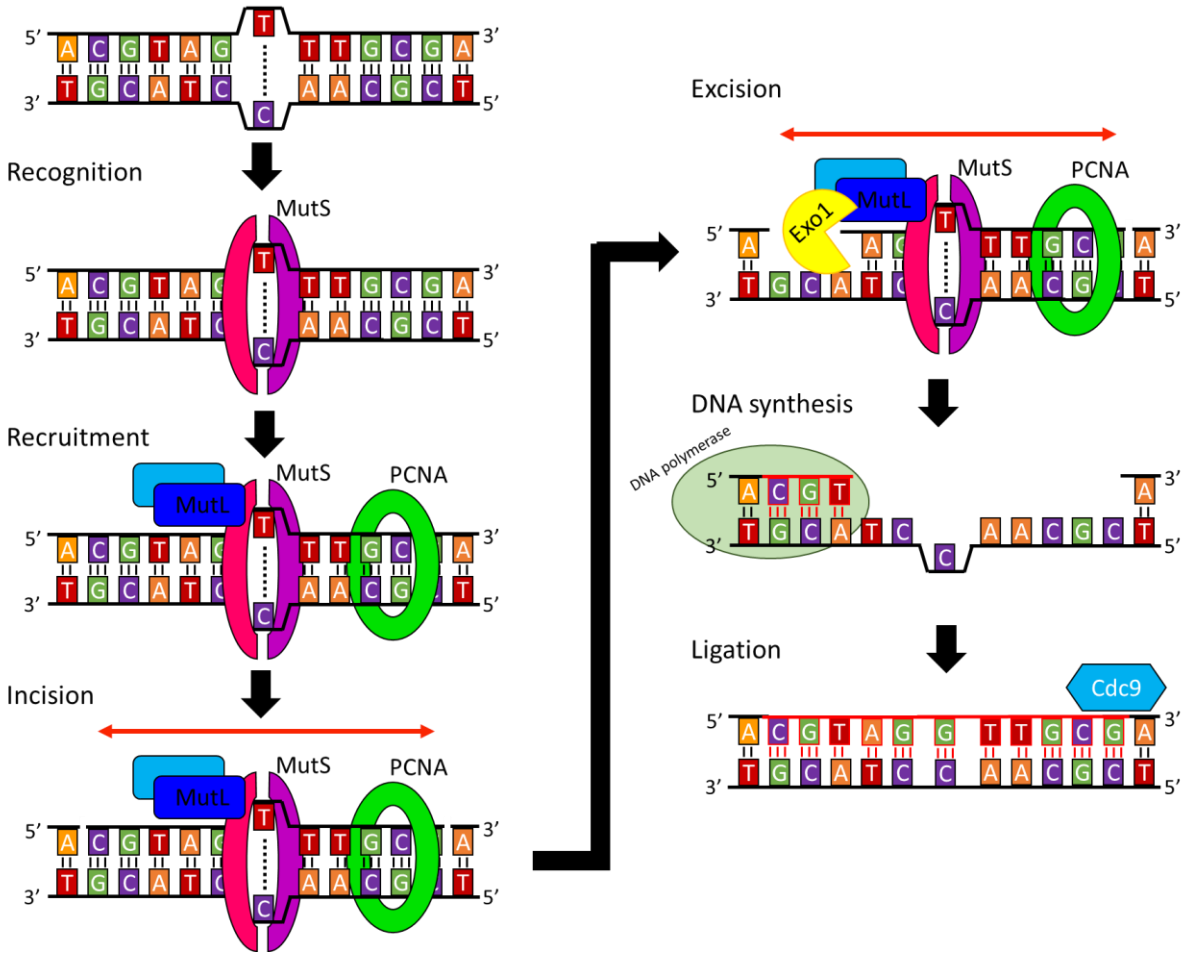

Figure 26: Mismatch repair (MMR).
Repair of mismatched bases using the MMR molecular mechanism, in *S. cerevisiae*. Recognition of the DNA mismatch by MutS (Msh2-Msh3 or Msh2-Msh6) complex which undergoes an ATP-dependent confirmation change forming a sliding clamp around the DNA. MutS complex then recruits the heterodimeric MutL complex (Mlh1-Pms1 or Mlh1-Mlh3) to the mismatch site along with Mlh1-Mlh2 accessory complex. Incision of the DNA backbone is achieved by endonuclease activity of MutL via sliding clamp properties of MutS, mediated by PCNA clamp (replicative clamp loader), in order to generate a nick within the actively replicating DNA strand (top strand). This nick permits the entry of the Exo1 protein leading to excision of DNA containing the mismatch. The gap is then filled by DNA polymerase Polδ or Polε and the remaining nick in the DNA backbone is repaired by the Cdc9 DNA ligase. Repaired and newly synthesized DNA is indicated in red.

2. DNA-DSB repair

i. Non-homologous end-joining (NHEJ)

The non-homologous end-joining (NHEJ) repair mechanism is thought to be a rapid and efficient way of repairing DNA-DSB, as, it involves the identification and direct ligation of the two DNA ends without extended homology. However, NHEJ is characterized as an error prone repair mechanism often generating small insertions and deletions (indels). Indeed, this may lead to important consequences via frameshifts within ORFs. However, some assays suggest that the mutagenic effect of NHEJ was initially overestimated, as it is often very

difficult to evaluate the exact rate of NHEJ because it is frequently genetically silent (leaving no identifiable repair scares) (Bétermier *et al.* 2014). Indeed, HR-mediated repair generates roughly 1,000 times more mutations than NHEJ (Deem *et al.* 2011).

Two major types of NHEJ exist: classical NHEJ and alternative NHEJ (alt-NHEJ). Classical NHEJ is a “cut and paste” repair mechanism which requires the binding of the Ku heterodimer, composed of Ku70 and Ku80 proteins, to the broken ends of the DNA. The Ku proteins ensure DNA protection from extensive resection and unwinding. Subsequently, the broken ends are ligated by the Lig4 complex (Figure 27). Other proteins, such as nucleases and polymerases may also be implicated in NHEJ, however this usually only occurs when the DNA ends are not cohesive or upon complex DNA breaks, as DNA ends need some processing before the ligation step (Rodgers and McVey 2016). In contrast, alt-NHEJ usually involves short resection and small homology between the two DNA ends thus, this pathway is also known as microhomology end joining (MMEJ). This DNA-DSB repair pathway is initiated by short resection or unwinding of single-strand DNA (ssDNA) of the broken DNA ends in order to expose a ssDNA microhomology region (6-20 nt. long) (McVey and Lee 2008). This microhomology permits the annealing of the two ends followed by filling of the gap, excision of DNA flaps formed (usually mediated by Rad1-Rad10) and ligation of the DNA backbone (Cdc9 ligase) (Ma *et al.* 2003) (Figure 27). MMEJ often results in deletions of various sizes between the microhomology regions because their annealing often generates DNA flaps which are removed. If extensive resection occurs due to absence of microhomology exposure, the DNA-DSB is rather repaired through single-strand annealing (section G.2.ii.b.b1, P.71).

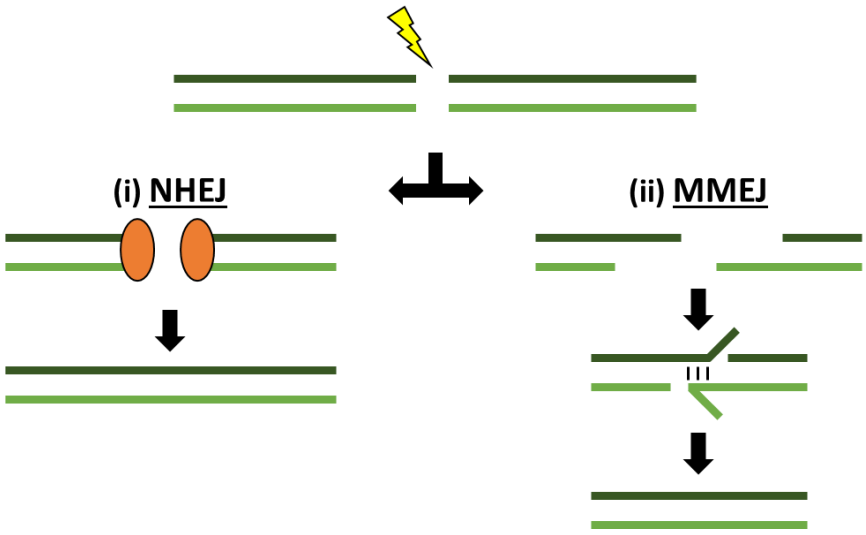

Figure 27: Non-homologous end joining (NHEJ).
There are two types of NHEJ which can be used to repair a DNA double-strand break (DSB) (yellow bolt) in *S. cerevisiae*, (i) classical NHEJ and (ii) alternative NHEJ mediated by microhomology (MMEJ). (i) NHEJ mechanism is initiated by the binding of Ku heterodimer (orange ovals), composed of Ku70 and Ku80 proteins, to the broken ends of the DNA. They ensure DNA protection from extensive resection and unwinding. Then, if needed, ends are processed by nucleases and polymerases before

being ligated by the Lig4 complex. Classical NHEJ result in error-free repair or minimal deletion/insertion of DNA at DNA-DSB locus. (ii) MMEJ involves short 5' to 3' resection of DNA ends or unwinding of single-strand DNA leading to unveiling of microhomology regions between two DNA strands. This microhomology permits the annealing of the two ends followed by filling of gap, excision of DNA flaps formed (usually mediated by the Rad1-Rad10 complex) and ligation of the DNA backbone (the Cdc9 ligase). Repair by alt-NHEJ is error-prone, often leading to deletion/insertion of DNA at DNA-DSB locus. The 5' to 3' strand is illustrated in dark green while lighter green illustrates the 3' to 5' DNA strand.

In *C. albicans,* the *CaKU70* gene shares little sequence identify with the other *KU70* homologs though it has retained the shared motifs and overall organization (Chico *et al.* 2011). Deletion of the *KU70* or *KU80* genes in *C. albicans* has no effect on the strains phenotypes in terms of sensitivity to DNA damaging agents, doubling times, mutation rate and antifungal sensitivity, suggesting that NHEJ is a dispensable DNA repair mechanism in *C. albicans* (Legrand *et al.* 2007; Chico *et al.* 2011). However, *KU70* seems to play a role in telomere length regulation in a gene dosage manner (Chico *et al.* 2011). Additionally, upon directed DNA-DSB, using I-*Sce*I or Cas9 nucleases, and subsequently analyzing molecular mechanisms used to mend the damaged DNA, two studies showed that *C. albicans* preferentially repairs DNA-DSB through HR-mediated repair pathways (Feri *et al.* 2016), unlike relatively closely related *C. glabrata* which prefers NHEJ (Vyas *et al.* 2018). Dispensability of NHEJ in *C. albicans* is once more reinforced by an absence of genomic instability upon *LIG4* deletion (Andaluz *et al.* 2002).

ii. Homologous recombination-dependent repair

a. Initiation of homologous recombination

DNA-DSBs can be fixed via numerous mechanisms of HR repair which all share a common initiation step. HR repair of DNA-DSB damages is based on the capacity of the cellular machinery to find and access redundant DNA sequences (sister chromatid or chromosome homolog) in order to use them as repair templates to regain genome integrity. Commencement of HR repair relies on identification of the broken DNA ends, followed by DNA resection, with end resection being a major determining factor for the NHEJ to HR choice in cells. Though we do not yet fully comprehend how two homologous DNA molecules find each other within the nucleus (homology search), our rich knowledge concerning factors involved in eukaryote end resection is based on numerous genetic studies conducted in the model organism *S. cerevisiae.* The broken DNA ends resulting from DNA-DSB are recognized and bound by the MRX DNA end processing complex, composed of the Mre11/Rad50/Xrs2 protein complex, and its co-factor Sae2 (Symington 2014). Processing of DNA ends results in 5' end resection and formation of a 3' ssDNA fragment. Extensive 5' resection is performed by the

5'-3' exonuclease Exo1 or by the helicase/nuclease activities of the Sgs1/Top3/Rmi1 complex in conjunction with the Dna2 nuclease (Mimitou and Symington 2008; Zhu *et al.* 2008). The MRX complex then recruits replication protein A (RPA) which stabilizes the generated 3' ssDNA. Certainly, the MRX complex competes with Ku protein heterodimers (involved in NHEJ, see section G.2.i, p.66) which also bind to DNA-DSB ends. The binding of the Ku complex to broken ends likely blocks the resection process. However, MRX-Sae2 may clip Ku proteins from the ends, similarly to the clipping of Spo11 protein (meiosis specific nuclease), in order to initiate HR repair (Zhang *et al.* 2007; Garcia *et al.* 2011). Understandably, disruption of the NHEJ pathway results in an increased frequency of HR-mediated repair of DNA-DSB (Rodgers and McVey 2016). Additionally, chromatin remodeling is often necessary in order to allow for long range 5' DNA resection and requires chromatin remodeling complexes such as the Fun30 chromatin remodeling factor. As a safeguard, the DNA damage-response protein Rad9 is recruited in the vicinity of DSB sites through binding to specific histone marks, to form a barrier and negatively regulate long range resection, in order to limit accumulation of long tracts of ssDNA in the genome (Chen *et al.* 2012; Symington 2014).

Displacement of the RPA proteins from ssDNA, which is required and favors homology search, is assured by several mediator proteins, including Rad52 which enables Rad51 to overcome RPA inhibition and promotes Rad51 binding to the 3' ssDNA. Rad51 is a DNA-dependent ATPase which forms a nucleoprotein filament with DNA around the ssDNA. This recombinase is involved in the homology search followed by the strand invasion of ssDNA into the intact duplex DNA, which will serve as the repair template. Additionally, Rad51 stabilizes this interaction between the ssDNA and homolog duplex DNA, which is essential for HR processing (San Filippo *et al.* 2008; Wright *et al.* 2018). The initiation of HR repair in yeast (according to *S. cerevisiae* data) is summarized in Figure 28.

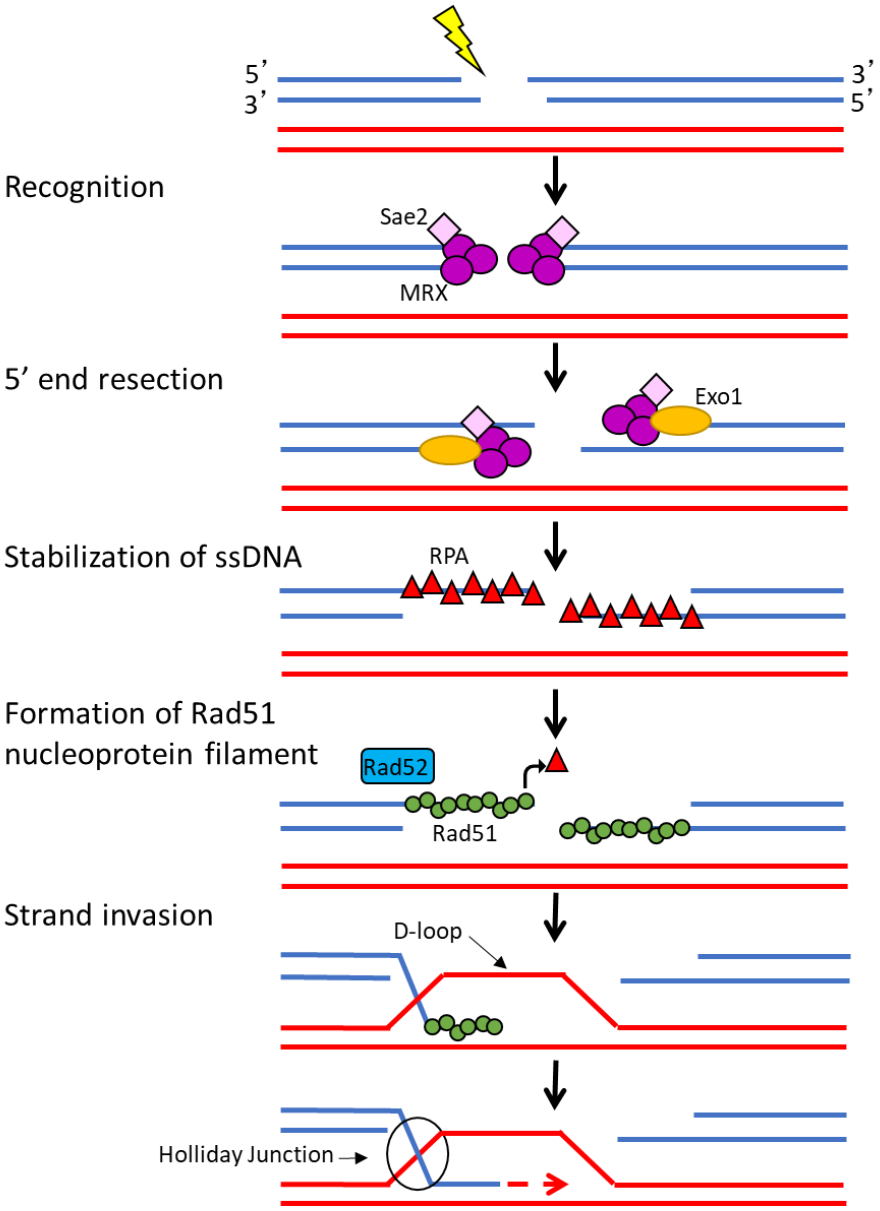

Figure 28: Initiation of homologous recombination during DNA-DSB repair.
Schematic representation of the initiation of homologous recombination (HR) during DNA-DSB repair. Upon DNA-DSB, MRX-Sae2, composed of the MRX protein complex (Mre11/Rad50/Xrs2) and cofactor Sae2, detect and bind to broken DNA ends. DNA ends are processed during a step called 5' resection by MRX nucleolytic activity and other recruited nucleases such as Exo1. The 5' resection results in the formation of 3' single-strand DNA (ssDNA). Additionally, MRX recruits RPA proteins which stabilize the 3' ssDNA. Displacement of RPA proteins is mediated by Rad52 which favors Rad51 nucleoprotein filament formation. The Rad51 nucleoprotein filament is a key player in the homology search and can displace the helix DNA leading to strand invasion of the ssDNA into the homolog duplex DNA, forming a displacement or D-loop. The structure formed at the intercept of ssDNA and donor DNA is named a Holliday junction. The progression of the HR repair involves D-loop extension and DNA synthesis, resulting in a multitude of outcomes

b. Homologous recombination completion

During search for homology, ssDNA can laterally interact with duplex DNA through dsDNA-binding properties of bound Rad51 proteins (reviewed in (Wright *et al.* 2018)). The

nucleoprotein filament is also able to displace the DNA double helix leading to strand invasion of the ssDNA into the homolog duplex DNA, called displacement or D-loop (promoted by Rad54 (Petukhova *et al.* 1999)). The latter leads to the formation of heteroduplex DNA (hDNA), describing a region within the D-loop where the invading ssDNA has intertwined with the complementary strand of the (invaded) homologous double-strand DNA (dsDNA), also referred to as donor DNA (Figure 28). The next step in the progression of HR repair is D-loop extension and DNA synthesis, resulting in a multitude of outcomes presented in the following subsections, i.e. single-strand annealing, synthesis-dependent strand annealing, gene conversion, mitotic crossover and break-induced replication (Figure 29). Though HR repair is referred to as error-free, the repair of DNA-DSB damages often result in DNA deletions or in LOH of different sizes. The impact of each mechanism upon genome structure will also be tackled below.

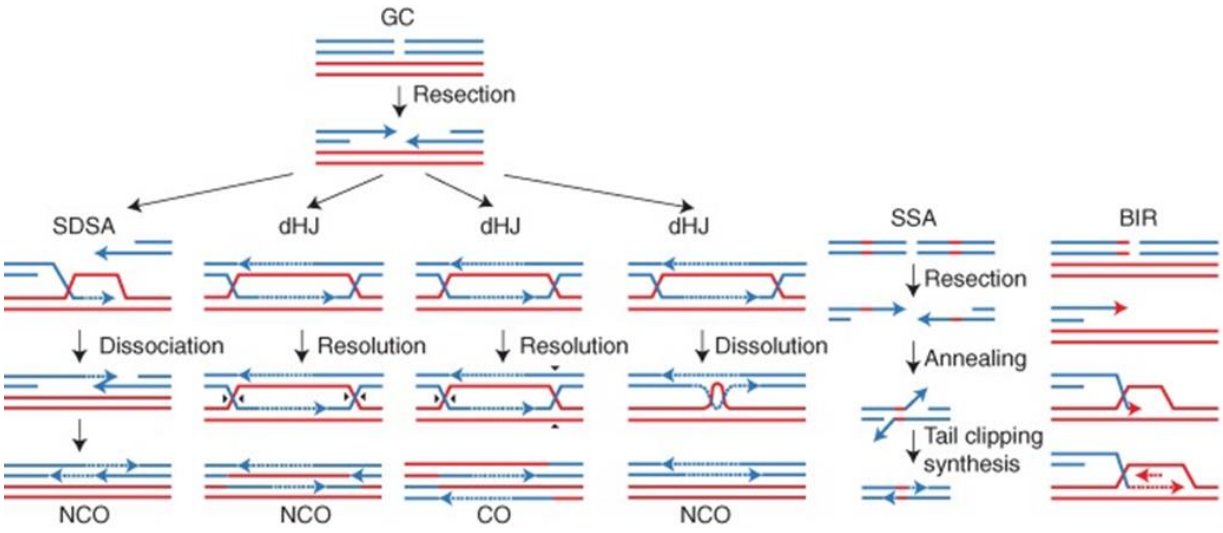

Figure 29: Homologous recombination-mediated DNA-DSB repair.
Summary of the major homologous recombination (HR)-mediated molecular mechanisms of DNA-DSB repair. Following DNA-DSB identification and 5' resection, several outcomes are possible; SDSA: synthesis-dependent strand annealing, GC: gene conversion, SSA: single-strand annealing and BIR: break-induced replication. Resolution of the double Holliday junction (dHJ) formed during GC can either be associated to crossover (CO) or non-crossover (NCO) depending on the location of resolvase cleavage activity. Dissolution of dHJ is independent of enzymatic activity and always leads to NCO. While disruption of D-loop results in SDSA-mediated repair, NCO event. If DNA end sectioning leads to the exposure of complimentary repeat sequences, theses sequences may anneal together and DNA-DSB is repaired via SSA. Lastly, the broken chromosome fragment, which does not contain the centromere, can be completely lost necessitating DNA synthesized from the break point until the telomere. This repair is named BIR. Figure from (Anand *et al.* 2013)

b1. Single-strand annealing (SSA)

Repair of DNA-DSB through single-strand annealing (SSA) occurs when 5' end resection exposes repeat DNA into ssDNA. The complementary ssDNA repeat sequences

permit the annealing of the two 3' ssDNA ends. Indeed, SSA is a Rad51-independent HR repair mechanism, as it does not require extended homology search and duplex DNA invasion. However, SSA is promoted by the Sae1 co-factor of the MRX complex and by Rad52 which favors repeat annealing. Once the repeats are annealed, excess DNA flaps are cleaved by Rad1-Rad10 excision activity and gaps are filled by DNA polymerase and DNA ligase (Prado and Aguilera 1995; Bhargava *et al.* 2016) (Figure 29). Certainly, SSA-mediated repair leads to DNA loss of various sizes because the nucleotides located between the two repeats are deleted during the repair process. Recent studies in *S. cerevisiae* and *C. albicans* highlight the importance of recombinogenic properties of inverted repeats (IR), demonstrating how different IR can lead to DNA deletions or genomic rearrangements, including CNV and LOH (Ramakrishnan *et al.* 2018; Todd *et al.* 2019). Indeed, these genomic alterations are often the result of SSA-mediated DNA-DSB repair, intervening between IR.

b2. Synthesis-dependent strand annealing (SDSA)

Synthesis-dependent strand annealing (SDSA) involves both homology search and strand annealing. Initiation of SDSA involves recognition and 5' resection of damaged DNA ends, homology search and ssDNA strand invasion leading to formation of a D-loop (as described in detail in section G.2.ii.a, p.68). However, the D-loop quickly collapses and both 3' ssDNA ends are annealed together. The remaining gaps are filled by DNA polymerase and the nicks of the DNA backbone are sealed by DNA ligase. Sometimes, upon annealing of the two 3' ends, DNA flap may form and be removed by a DNA flap endonuclease (Rad1-Rad10), though this occurs only rarely (Morrical 2015) (Figure 29). And because none or little DNA synthesis occurs during D-loop extension in SDSA, SDSA-mediated repair rarely results in LOH (Ira *et al.* 2003).

b3. Gene conversion (GC), with or without crossover

HR-mediated repair leading to "gap filling" or "patching repair" of a DNA-DSB is referred to as gene conversion (GC), as the two DNA ends are mended together by a newly synthesized path of DNA. GC results in short-tract LOH surrounding the locus which underwent DNA-DSB. Upon initiation of HR repair leading to D-loop formation, the D-loop is extended by the activity of DNA polymerase (Polδ or Polε) which synthesizes the nascent strand. D-loop disruption is unavoidable and is required for chromosome resolution. Indeed, different types of D-loop dissolution are possible, leading to various patterns of LOH. Certainly, D-loop may collapse and nascent strand may be joined to the second 3' ssDNA end in order to mend the gap (in a similar manner to SDSA mechanism) (Morrical 2015). Alternatively, the

second 3' end can anneal with the D-loop forming a double Holliday junction (dHJ) which can then be solved with or without a crossover event. Indeed, dHJs may be dissolved, leading to non-crossover (NCO) events, or be resolved via endonucleolytic enzymatic activities, leading to NCO or crossover (CO) events (Figure 29). In *S. cerevisiae*, at least three proteins (complexes) are involved in D-loop disruption: the Sgs1-Top3-Rmi1 complex, Srs2 (helicase) and Mph1 (helicase). Sgs1, Srs2 and Mph1 are ssDNA translocases which eject the nascent strand from the hDNA, resulting in D-loop dissolution (NCO) (Sakofsky and Malkova 2017; Wright *et al.* 2018; Piazza *et al.* 2019). These helicases (Mph1 and Sgs1) also participate in execution of DNA checkpoint, favoring fast DNA repair mechanisms such as short GC patches (Jain *et al.* 2016). Furthermore, several resolvases, Mus81-Mms4, Slx1-Slx4 and Yen1, have also been identified and studied in the model yeast though, their mechanism of function and regulation are not yet completely comprehended. Resolvases induce DNA backbone nicks at each HJ permitting untangling of the two DNA strands and direct ligation. Depending in which plane the resolvases induce the nicks, resolution of dHJ can result in a CO event, where the two homologs exchange their genetic material. Thus, the latter molecular mechanism is named GC with CO (GC w CO) (Yin and Petes 2015; Sakofsky and Malkova 2017; Pardo *et al.* 2020).

Ideally, during dHJ processing, the NCO events are favored as they limit genomic rearrangements, and LOH in the vicinity of the DNA-DSB locus. Events of GC with NCO will here on be referred to as GC. In contrast, the processing of dHJ leading to CO events can lead to extensive LOH spanning from the break site until the telomere. As illustrated in Figure 30, long-tract LOH mediated by GC w CO often result from DNA-DSB which occurred in G2 phase of the cell cycle, when twice the genetic material is present within the nucleus.

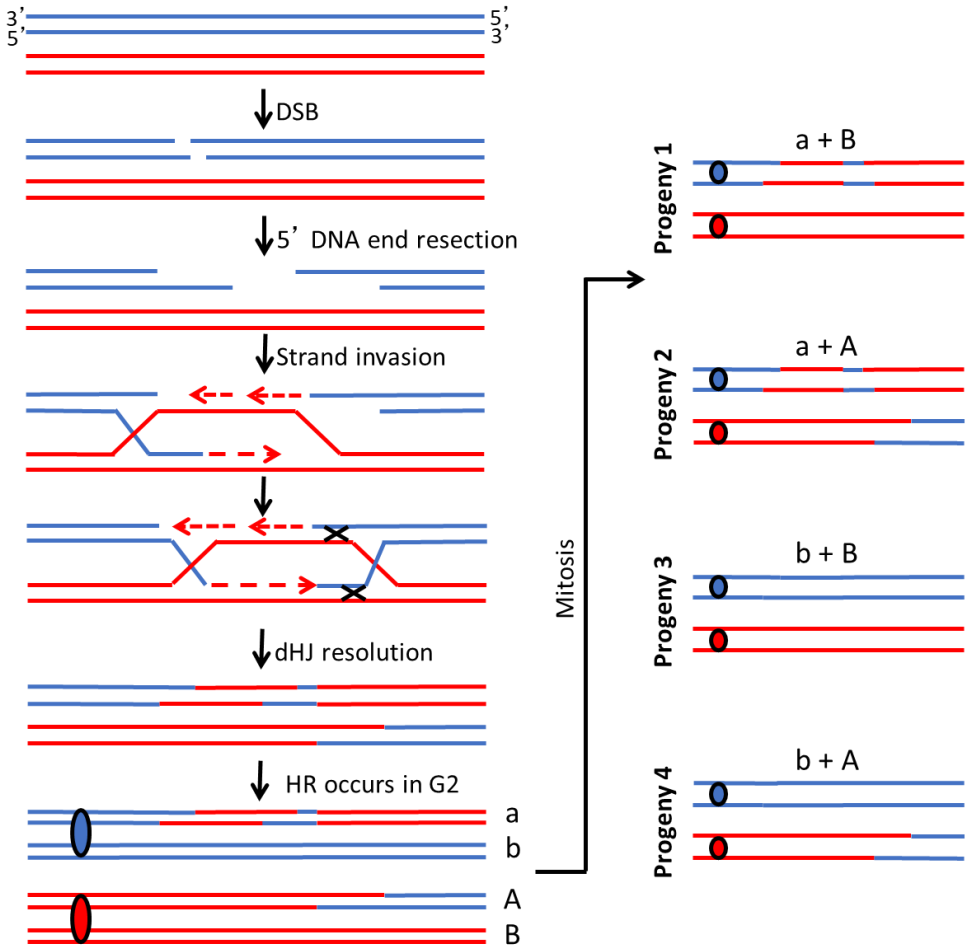

Figure 30: Gene conversion with crossover during G2 phase of cell cycle can lead to long tract LOH events.
Upon DNA-DSB, DNA broken ends are identified and undergo 5' resection leading to the formation of 3' single-strand DNA. The latter then invades intact duplex DNA of the homolog in order to form a D-loop followed by a double Holliday junction (dHJ). Resolution of dHJ can lead to a crossover event (black X's) between the two homologs involved in DNA repair. When the DSB and following repair occurs during G2 phase of the cell cycle, mitosis results in possible progenies. Progenies 2 and 3 display short-tract LOH events localized around the DSB locus. While progenies 1 and 4 possess long-tract LOH, spanning from the DNA-DSB locus until the telomere. The centromeres are represented by ovals, while blue lines represent haplotype A and red ones represent haplotype B. Figure adapted from (Feri *et al.* 2016).

In *S. cerevisiae*, two sizes of spontaneous GC have been described, short (900 bp – 25 kb) and long (25 – 121 kb) tracts (Yim *et al.* 2014). In *C. albicans,* GC has been observed both in *in vitro* and *in vivo* passaged (Forche *et al.* 2018; Ene *et al.* 2018; Sitterle *et al.* 2019) and clinical isolates (Ropars *et al.* 2018; Wang *et al.* 2018). Additionally, exposure to H_2O_2 (Forche *et al.* 2011) and overexpression of *RAD23* and *BIM1* (Loll-Krippleber *et al.* 2015) seem to enhance GC. However, in *C. albicans*, distribution of GC (internal LOH tracts, surrounded by heterozygous regions) has not been thoroughly investigated though, whole genome analyses have revealed the presence of numerous internal LOH regions, LOH tracts flanked by heterozygous regions, spanning several kilobases which could have arisen from

GC (Hirakawa *et al.* 2015; Ropars *et al.* 2018; Wang *et al.* 2018). Furthermore, recent studies have highlighted the frequent occurrence of micro-LOH events. These micro-LOH can span 1 - 80 bp (Ene *et al.* 2018; Sitterle *et al.* 2019) indeed, suggesting that different sizes of GC tracts may also exist in *C. albicans*. Even though we cannot refute the idea that these small LOH events may result from other repair mechanisms such as LP-BER, NER or MMR. Additionally, GC also is implicated in maintaining native centromere function (Thakur and Sanyal 2013), as previously discussed in section D.1.i. Furthermore, GC seems to be the predominant molecular mechanisms used upon I-*Sce*I and Cas9 DNA-DSB repair *in vitro* in *C. albicans* (Feri *et al.* 2016; Vyas *et al.* 2018). And while repair via GC w CO remains rare, such events are still detectable (Feri *et al.* 2016). Because NHEJ seems to be minimal (or even absent) in *C. albicans*, and given the positive correlation between heterozygosity and fitness, it appears logical to hypothesize that GC is the preferred mechanism of DNA-DSB repair in this pathogenic yeast, enabling to maximize maintenance of heterozygosity.

b4. Break-induced replication (BIR)

GC and break-induced replication (BIR) share the same initiation steps leading to strand invasion into homologous duplex DNA. However, GC occurs when both broken DNA ends are engaged in the repair process while BIR takes place when only one end successfully finds a homologous template (Figure 29). BIR has been extensively studied in *S. cerevisiae,* permitting detailed descriptions of both types of BIR, i.e. Rad51-dependent and Rad51-independent mechanisms. The Rad51 protein, involved in the formation of the Rad51 nucleoprotein filament and in the process of homology search, is dispensable during BIR repair, although participation of Rad51 renders BIR most efficient (Malkova *et al.* 1996, 2005; Davis and Symington 2004). Furthermore, BIR initiation was shown to be delayed as compared to GC as it necessitates extensive 5’ resection (> 1 kb), generating long 3’ ssDNA ends (Malkova *et al.* 2005; Chung *et al.* 2010). The Rad51-independent BIR pathway is not efficient and recombination events in diploid Rad51$^{-/-}$ strain almost never occurs at the break site (Downing *et al.* 2008). In absence of Rad51, the strand invasion process is mediated by both Rad52 and the large 3’ ssDNA homology regions exposed by 5’ resection, potentially facilitating pairing of homologous DNA. Additionally, this pairing is favored by the exposure of repeat sequences during 5’ resection, such as transposable elements and inverted repeats (Malkova *et al.* 2001; Downing *et al.* 2008). While, repair by BIR independently of Rad51 but dependent on short homology or homeology repeat sequence annealing forms another branch of BIR repair named microhomology mediated BIR (MMBIR) (Hastings *et al.* 2009). Indeed, MMBIR is a mix between Rad51-independent BIR and SSA repair. Undoubtedly, extensive 5’ resection plays a crucial role in BIR. The data of Chung and collaborators propose that

resection occurs before, during and after homology search. Moreover, they observed that the majority of BIR events required > 2.5 kb of 5' resection, of which 50% of these events showed ≥ 15.5 kb resection stretches and a few events even spanned more than 27.5 kb (Chung *et al.* 2010). DNA resection also pays a role in stabilizing and maintaining DNA checkpoint arrest until repair is completed (Jain *et al.* 2009), which is especially important in the case of BIR as several thousand or hundreds of thousands of bases must be synthesized to repair the DNA damage (in a conserved DNA synthesis manner (Saini *et al.* 2013)).

BIR often generates GCRs (Sakofsky *et al.* 2012) including, but not restricted to, long tract LOH events which span from the DNA-DSB locus until the telomere in diploid heterozygous strains. Multiple cycles of strand invasion and D-loop extension, through DNA synthesis, have been described during BIR (Stafa *et al.* 2014; Mehta *et al.* 2017). These cycles participate to the mutagenic effect of BIR as they can promote GCRs such as translocations (Anand *et al.* 2014; Piazza *et al.* 2019). Additionally, the multiple cycles of strand invasion can lead to template switching, which occurs at the start of BIR followed by a stabilization after roughly 10 kb of DNA synthesis (Smith *et al.* 2007). Indeed, BIR is an error-prone mechanism of DNA-DSB repair. Mutagenic effects of BIR are also due to DNA polymerase δ which induced many frame shifts, with a 1,000-fold increase in mutagenic rate as compared to normal replication (Deem *et al.* 2011). Moreover, as this HR mechanism involves the unraveling and synthesis of large DNA stretches, the MMR machinery is also active at the locus of repair. Deem *et al.* demonstrated the importance of MMR during BIR, as MMR can compensate the mutagenic properties of Polδ (Deem *et al.* 2011). Please refer to section G.1.iii for in depth understanding of MMR (p.64). Point mutation clusters (4 - 115 kb) can be generated during BIR-mediated DNA-DSB repair (~50% of tested strains), notoriously upon exposure to DNA-damaging agents (such as MMS or APOBEC (deaminase activity)) (Saini *et al.* 2013; Sakofsky *et al.* 2014; Elango *et al.* 2019). This can be explained by lengthy DNA resection generating long ssDNAs, which are extremely sensitive to DNA damages. Yeast cells which endure multiple simultaneous DNA-DSB repairs have a higher risk of generating such mutation clusters and other genomic rearrangement events, signatures of cancerous cells (Sakofsky *et al.* 2019).

BIR in eukaryotes is thoroughly reviewed by Sakofsky and Malkova in (Sakofsky and Malkova 2017).

BIR is defined as the appearance of long-tract LOH events spanning from a given locus until the telomere and are frequently observed in the genome of *C. albicans* though, they cannot always be fully distinguished from GC w CO or MCO events (Forche *et al.* 2011; Loll-

Krippleber *et al.* 2015; Feri *et al.* 2016). Indeed, numerous studies have illustrated an important prevalence of BIR in *C. albicans*, occurring *in vitro* spontaneously, in OE and/or KO strains of HR genes (*RAD51*, *RAD53*, *RAD50*, *MRE11*), and upon exposure to high temperature, fluconazole and oxidative stress (Legrand *et al.* 2007; Forche *et al.* 2011; Loll-Krippleber *et al.* 2014, 2015; Feri *et al.* 2016; Ene *et al.* 2018). Furthermore, BIR events also arise *in vivo* in commensal and infection models (Forche *et al.* 2009a, 2018; Ene *et al.* 2018; Sitterle *et al.* 2019). In fact, evidence of BIR events are frequently observed in clinical strains using WGS data analysis (Ropars *et al.* 2018; Wang *et al.* 2018). More recently, Todd and collaborators have shown that numerous BIR break-sites within laboratory and clinical isolates coincide with IR (Todd *et al.* 2019).

iii. No repair

As opposed to DNA break impacting only one DNA strand, DSB cut through both strands completely severing the DNA helix into two separate fragments. If not repaired, the DNA fragment, which does not possess the centromere, and all its genetic content will be permanently lost for the cell and is defined as chromosome truncation (CT). Depending on the severity of the DNA-DSB, the whole damaged chromosome may also be lost, resulting in chromosome loss (CL). Thus, it is crucial for the cell to quickly identify and reconnect the two DNA fragments in order to avoid major DNA losses.

As mentioned in previous sections, both CT and CL can be observed or induced in *C. albicans*. For instance, certain prolonged growth conditions may induce CL such as growth on sorbose which promotes loss of Chr5 (Janbon *et al.* 1998), an advantageous tool with regards to *in vitro* mating as it permits to generate *MTL* homozygous individuals which could undergo the parasexual life cycle. Indeed, DNA-DSB that are not repaired result in long tract LOH events and CNV, monosomy (Figures 11 and 14). Exposure of *C. albicans* cells to heat and fluconazole, seems to lead to the appearance of CL events (Hilton *et al.* 1985; Bouchonville *et al.* 2009; Forche *et al.* 2011; Loll-Krippleber *et al.* 2015). Additionally, the size of MRS has been shown to influence the frequency of CL (Lephart *et al.* 2005). It is worth noting that monosomic chromosomes are often reduplicated as the haploid state is not very stable in *C. albicans*. From a mechanistic point of view, deletion of *RAD51* and *RAD52* result in an increase in CT and CL, suggesting that DNA-DSBs fail to be properly repaired in these genetic backgrounds (Andaluz *et al.* 2011; Bellido *et al.* 2019; Ciudad *et al.* 2020). According to Andaluz and collaborators, all LOH events occurring in rad52$^{-/-}$ strains result from CL (66%) and CT (34%) (Andaluz *et al.* 2011). Consequently to CT, the impacted chromosome also loses telomeric sequences thus, continues to be vulnerable to DNA degradation during DNA

replication. To avoid further DNA loss and to stabilize the damaged chromosome, new telomere-like sequences are added at the damaged end. CTs are capped with a 9 nt.-long repeat sequence possessing a pattern similar to telomeric repeat sequences (Selmecki *et al.* 2005). However, overall the “no repair” scenario rarely occurs as illustrated by the analysis of DNA-DSB repair mechanisms used upon I-*Sce*I cleavage (Feri *et al.* 2016).

H. Biological implications of genomic instability

Thus far we have explored various features influencing acquisition of genetic variation in *C. albicans*. Notably, (i) population structure, (ii) life cycle, (iii) various genomic features promoting recombination, (iv) the relatively elevated heterozygosity level amplifying variation, (v) the tolerated high genomic plasticity resulting in many genetic rearrangements and (vi) the participation of DNA repair mechanisms in the appearance of GCRs, specifically LOH. Overall, this illustrates that an important balance is required between heterozygosity and genomic plasticity, two hallmarks of the genome of *C. albicans*. However, we have not yet discussed the biological implications of the latter genomic characteristics and their potential association to adaptation in this opportunistic human fungal pathogen. Thus, this section aims to concretize the importance of genomic rearrangements on the acquisition or selection of genetic variants, a strong survival strategy potentially permitting and/or promoting continuity of the species.

1. Genomic rearrangements as a response to environmental factors

If elevated genomic plasticity in *C. albicans* acts as a survival strategy for this species, we expect that these yeast cells respond to various environmental factors through the acquisition of genomic rearrangements. In the laboratory environment, numerous studies have shown that exposure of *C. albicans* to various conditions induce GCRs, as presented throughout the introduction. Indeed, exposure to fluconazole is often associated with the acquisition of aneuploidy events in both laboratory and clinical isolates (Selmecki *et al.* 2005; Todd *et al.* 2017). In brief, antifungal exposure can lead to gain of aneuploidy through cytokinesis failure leading to a tetraploid progeny which then undergoes unequal nuclear segregation (Harrison *et al.* 2014). If focusing specifically on LOH events, many *in vitro* factors influence the appearance of these genomic rearrangements such as, KO and OE mutant strains, oxidative stress, heat, genotoxic agents, and fluconazole ((Gusa and Jinks-Robertson 2019) and other references cited throughout the introduction). The work of Forche *et al.*

provided ample evidence that environmental factors influence LOH in *C. albicans,* as LOH rates were shown to augment upon exposure to various conditions (H_2O_2, fluconazole and heat) at five genomic loci. With the use of the *GAL1/URA3* LOH reporter system, their work also demonstrated that the nature of LOH on Chr1 varies between growth conditions, as exposure to H_2O_2 predominantly results in short-tract LOH while fluconazole and growth at 39°C favor the appearance of long-tact LOH (Figure 15) (Forche *et al.* 2011). This suggests that different types of DNA breaks caused by stresses can potentially influence the molecular mechanisms used to mend the break, hence dictate the LOH size (section G, p.59). However, a clear identification of stress responses that might control LOH-promoting mechanisms is still missing in *C. albicans*. Additionally, there is little or no knowledge regarding coordination of LOH events, as LOH dynamics at a genome-wide level is poorly studied in *C. albicans*. A single study assessed the rate of loss of one homolog at different genomic loci in standard growth conditions. Authors concluded that no obvious differences in LOH rates were observed in rich media and that these rates seemed to be correlated with the distance to the CEN (Forche *et al.* 2011). Nevertheless, the genome-wide reactivity to stress by means of LOH appearance has yet to be thoroughly explored.

Studying the effects of various stresses on *C. albicans* is of high importance especially regarding its natural environment. Indeed, being an opportunistic fungal pathogen of humans, *C. albicans* is constantly confronted to fluctuating environmental conditions. Indeed, as a commensal yeast of the human GIT, yeast cells are exposed to a changing environment in terms of pH, oxygen, oxidative stress levels and temperature variations. Not to mention, *C. albicans* is in close contact with thousands of other commensal microorganisms residing within the host and characterizing the human microbiota. Certainly, the naturally fluctuating environment constitutes a source of signals that are likely to promote genomic rearrangements. *C. albicans* commensal strains and strains passaged in an infectious mouse model have been shown to acquire GCRs, namely LOH (Diogo *et al.* 2009; Forche *et al.* 2009a). A more recent study conducted a series of evolution experiments (*in vitro* and *in vivo*) and compared mutations landscapes in a total of 28 evolved *C. albicans* isolates (Ene *et al.* 2018). Using whole-genome sequencing data, the authors identified and compared the genomic variations acquired by *C. albicans* during passaging in standard laboratory conditions versus in murine systemic infection model and two GIT colonization models, mediated by antibiotics or by diet. Because *C. albicans* is not a natural colonizer of the murine GIT, stable GIT colonization necessitates the modification of the mouse GIT microbiota using antibiotics or a glucose rich diet. The complete spectrum of mutations acquired during microevolution of *C. albicans* includes numerous point mutations and micro-LOH events and some longer-tract LOH and aneuploidy events (Ene *et al.* 2018). Authors estimated an *in vitro* base substitution

frequency of ~$1.71x10^{-7}$ base pairs per generation and a frequency of GC tracts (small LOH) of ~$1.61x10^{-10}$ base pairs per generation, though both types of mutational events are equally important during micro-evolution as they impact comparable number of positions. Importantly, this study revealed that the same types of mutational events observed *in vitro* were also occurring *in vivo*, but were 11x more frequent in murine *in vivo* environment. Additionally, aneuploidy events were described in 10.7% (3/28) of evolved lineages, only associated to GIT micro-evolution (Ene *et al.* 2018). Additionally, other studies have also described recurrent trisomy of Chr5, Chr6 or Chr7 in murine commensal models, oropharyngeal and GIT (Forche *et al.* 2018, 2019; Tso *et al.* 2018). Evolution patterns similar to those described above in mice have also been observed within-host in healthy human *C. albicans* carriers (Sitterle *et al.* 2019). Furthermore, even in an *in vivo* scenario, regions surrounding repetitive sequences are often rich in new genetic variants (Forche *et al.* 2018; Ene *et al.* 2018; Sitterle *et al.* 2019). Indeed, these recent publications greatly strengthen the idea that genomic plasticity can participate to the adaptation of *C. albicans* to its perpetually evolving natural environment.

2. Genomic rearrangements and adaptation

Adaptation is the process through which an organism becomes better fitted to its environment. For example, it can involve the acquisition of a single nucleotide modification which leads to a gain in fitness in a given situation. In the context of this thesis, the following section will only tackle examples of adaptation associated with genomic rearrangement events. As discussed throughout section E of the introduction focusing on genome plasticity events, different types of GCRs can contribute to the acquisition of genetic variation and consequently lead to adaptation. Potentially the most well know example of adaptation in *C. albicans,* involves the acquisition and association of multiple genomic modifications leading to antifungal resistance. Numerous research teams are dedicated to better understand the emergence of drug resistance in *C. albicans* and therefore a large panel of mutations leading to antifungal drug resistance has been described in the literature (Sitterlé *et al.* 2020). Often, these mutations modify or disrupt the drug target(s) or impact favorably key regulators involved in drug efflux or neutralization. Because *C. albicans* is mostly found in the diploid state, acquisition of a mutation usually occurs in only one allele and can have various effects; (i) it can abolish the WT phenotype (dominant mutation), (ii) it can lead to an intermediate phenotype (codominance) or (iii) it can be completely buffered by the second WT allele (recessive mutation). Indeed, by coupling such mutation(s) to genomic rearrangements, the mutation's effect may be uncovered or enhanced. A perfect example was presented in section D.4 Heterozygosity, p.33, where the effects of a GOF mutation in the *TAC1* gene is amplified by the acquisitions of an LOH event leading to a higher resistance to azoles, as the drug is

more efficiently pumped out of the cells (Figure 16) (Coste *et al.* 2006, 2007). In *C. albicans*, both *TAC1* and *ERG11* (encoding lanosterol-14α-demethylase) genes are located on the left arm of Chr5. The *TAC1* gene encodes a transcription factor regulating the expression of drug efflux pumps while *ERG11* is involved in ergosterol biosynthesis and the target of azole drugs (Figure 1). Mutations in both genes lead to an additive contribution to fluconazole resistance in a gene copy number-dependent manner as demonstrated by the gain of the isochromosome i(5L) (Selmecki *et al.* 2008).

In the recent years, there has been growing evidence of genetic modifications which lead to adaptation *in vivo* through GCRs. Evidently, adaptation is not a direct synonym of increased virulence of *C. albicans* but may also lead to a better fitted strain with a reduced virulence. As mentioned above, *C. albicans* being a poor colonizer of the murine GIT, Tso *et al.* benefitted from this characteristic to evolve *C. albicans* isolates into a better fitted commensal strain characterized by a reduced virulence potential in a systemic infection model. Upon characterization of their evolved strains, the authors observed defects in yeast-to-hyphae transition, a key virulence factor of *C. albicans* pathogenesis. The described phenotype was associated to loss-of-function mutations located in the *FLO8* gene; a transcription factor involved the progression of the yeast-to-hyphae morphological switch. Theses mutations were also often associated to an LOH event which exposed the mutated *flo*8 recessive phenotype (Tso *et al.* 2018). Indeed, LOH participated to the rapid adaptation of evolved strains towards commensalism. Likewise, Liang *et al.* demonstrated how heterozygosity, GCRs and morphological switches in *C. albicans* participate to its adaptive potential. Numerous *C. albicans* clinical strains are hemizygous for *EFG1*, a pleiotropic factor regulating filamentation, metabolism, biofilm formation, virulence and the white-opaque switch. Additionally, *EFG1* is also implicated in the gray (or GUT) *C. albicans* morphology phenotype which has been described in cells that have been passaged through the gut (Pande *et al.* 2013). Only *EFG1* hemizygous strains, often possessing a premature STOP codon disrupting functionality of one of the two allele, can undergo the white-gray switch. Interestingly, high frequency of switch is often associated with an LOH event leading to two *EFG1* nonfunctional alleles, resulting in complete inactivation of *EFG*1. Finally, *in vivo* passaging of hemizygous strains through the murine GIT promotes inactivation of *EFG1* through LOH, resulting in gray strains with an increased virulence that outcompete WT *EFG1*$^{+/+}$ strains in GIT colonization and systemic infection models (Liang *et al.* 2019).

C. albicans strains acquiring GCRs may also display niche adaptation phenotypes. As mentioned in the previous subsection, passaging of *C. albicans* strains through the oropharyngeal colonization murine model promotes the recurrent and independent

appearance of Chr5 and Chr6 trisomy events (Forche *et al.* 2018). Subsequently, authors demonstrated that Chr6 trisomy, resulting in one Hap A and two Hap B copies (ABB), appears twice as frequently as Chr6 trisomy resulting in two copies of HapA and one copy of HapB (AAB). Additionally, Chr6 ABB strains display a commensal phenotype, as they generate a reduced inflammatory response and, are less adherent and invasive of oral epithelium (Forche *et al.* 2019). This suggests that Chr6 ABB trisomy is better adapted to the oral niche as it leads to a better colonization of the oral cavity potentially associated with its commensal-like phenotype. Furthermore, *C. africana* strains from Clade 13 (Figure 4) seem to be restricted to the vaginal niche as they have only been reported to cause genital infections and display a reduced fitness and virulence as compared to other more heterozygous *C. albicans* strains (Borman *et al.* 2013; Ropars *et al.* 2018). Though further studies are needed to validate this hypothesis, the numerous LOH events which potentially lead to the appearance of Clade 13 could have participated to the restriction of *C. africana* to the vaginal niche. Overall, genomic rearrangement events seem to impact the appearance of adaptive phenotypes in *C. albicans*.

I. Overview of the aims and achievements of the PhD thesis

As discussed throughout this literature review, genomic rearrangements are intricately involved in many aspects of *C. albicans* biology and pathogenesis, notably through the occurrence of LOH. At the beginning of my thesis there was no studies which particularly described the landscape of LOH events in *C. albicans* (*in vitro* or *in vivo*) or the potential effect of genomic features on LOH occurrence. Additionally, although some evidence suggested a link between LOH and *C. albicans* pathogenesis, no *in vivo* example of adaptation mediated by genomic rearrangements was available. Studies conducted in the recent years have greatly advanced our knowledge to this regard, including my thesis work. During the course of my thesis, I was able to investigate several aspects pertaining to LOH and to better understand how LOH are implicated in generating and shaping genetic variations in *C. albicans* as a result of the many questions that have been raised throughout my thesis. The result section of my thesis will be divided into five chapters.

1. The majority of articles investigating genetic diversity between several sequenced *C. albicans* strains are based on analysis of SNPs and CNVs. However, despite LOH being ubiquitously found within clinical isolates, there is little descriptive data relating to LOH in terms of frequency and distribution. In my first chapter, I took advantage of our laboratory's collection of 192 sequenced clinical *C. albicans* isolates in order to conduct a descriptive analysis of the LOH landscape found within natural strains of Cluster 1. We developed an

analysis pipeline allowing the identification of > 7,000 LOH events within 54 *C. albicans* isolates of Cluster 1. This LOH data set revealed that each strain possesses numerous concomitant LOH events, short- and long-tract, affecting on average 9.4% of their genome. Similarly to microevolution studies conducted in the laboratory, our data suggested that short-tract LOH are equally important in natural clinical isolates, as the majority of LOH are under 10 kb. Indeed, this descriptive study gives an overview of LOH in natural *C. albicans* strains and highlights the important presence of LOH events in the genome of this opportunistic pathogenic yeast.

2. Although LOH are pervasive, previous studies reported chromosome homozygosis biases, where a given haplotype is never found in the homozygous state. Theses biases are usually associated with the uncovering of recessive lethal alleles upon LOH. My second chapter focused on the identification of recessive lethal allele(s) on Chr7, responsible for the absence of Chr7B homozygous individuals. In the genetic background of the laboratory reference strain SC5314, 13 candidate recessive lethal alleles were identified amongst which two were confirmed as a recessive lethal allele and a recessive deleterious allele. This project also allowed me to investigate the potential role of repetitive sequences, such as MRS, in buffering the deleterious impact of these alleles in *C. albicans*.

3. As mentioned above, LOH often results from the repair of DNA-DSB and most of our knowledge involving these molecular mechanisms is based on studies in *S. cerevisiae*. During the establishment of the inducible I-*Sce*I DNA-DSB system in *C. albicans* by a previous PhD student from the BPF unit, Adeline FERI, we observed that homozygosis tracts resulting from break-induced replication (BIR) extended from the break site to the telomere but also unexpectedly extended from the break site towards the centromere. In the third chapter, I intended to identify the molecular mechanism(s) that could be responsible for the extended homozygosis tracts by characterizing several deletion mutants of key genes involved in various aspects of DNA repair. This mechanistic study of LOH permitted to characterize the role of three new genes (*POL32, MPH1* and *MUS81*) on genome stability in *C. albicans* and to highlight certain similarities and potential divergences between *C. albicans* and *S. cerevisiae.* Additionally, this chapter revealed that ascending LOH towards centromeres seem to be associated with intrinsic features of BIR and the potential involvement of the MMR machinery which “corrects” natural heterozygous positions leading to LOH.

4. When I entered the world of *C. albicans*, I was warned that the transformation protocol in *C. albicans* was highly mutagenic. As I am mostly interested in genome stability, I naturally

wondered if there were any differences with regards to transformation-induced mutagenesis between protocols. With the advent of the CRISPR-Cas9 technology and as I envisioned the construction of numerous strains, I was able to compare the mutagenic landscapes between a classical transformation protocol used in the laboratory and the new transformation protocol mediated by CRISPR-Cas9. This study emphasized the mutagenic effects of transformation in highly plastic *C. albicans,* in terms of LOH and aneuploidy, and showed that final constructed strains using CRISPR-Cas9-mediated transformation possessed a decreased amount of aneuploidy events.

5. All previous chapters illustrate the importance of LOH in *C. albicans* although, at the start of my PhD, the biological importance of LOH was largely based on the phenotypic characterization of strains. Few studies sought out to understand how LOH were regulated upon stress at a genome-wide level. Thus, I pursued to investigate how chromosomes react to stress by elucidating genome-wide LOH dynamics. I built a collection of strains allowing to robustly examine LOH frequencies in various genomic loci. Several strains also permitted me to assess the heterogeneity of genome stability in *C. albicans* with regards to LOH frequency, in standard laboratory conditions and upon stress. This investigation demonstrated that heterogeneity is not restricted to phenotype in *C. albicans,* as our results revealed elevated levels of LOH frequency heterogeneity inter-strain, intra-strain and inter-chromosome in basal and stress conditions. Heterogeneity of genome stability could act as an important adaptive strategy for predominantly clonal *C. albicans* permitting to rapidly generate a wide-spectrum of allele combinations.

Overall, the work conducted over these past years has led to deepen the comprehension of LOH events in *C. albicans* from multiple facets, i.e. mechanistically but also in terms of scales, from different laboratory conditions (intra- and inter-strain) to within a natural population. Each of the following chapters contains, in article format, a detailed description of each project conducted throughout my thesis. Overall outcomes, the resulting questions and additional work conducted during my thesis, which could not necessarily be included in one of the five chapters, will be discussed within the perspectives section of this thesis.

.

Results and Discussion

Chapter 1: Patterns of loss-of-heterozygosity within the natural population of *Candida albicans*

Contains: Article in preparation

Patterns of Loss-of-heterozygosity within the natural population of *Candida albicans*

Timea MARTON[1,2], Corinne MAUFRAIS[1,3], Marie-Elisabeth BOUGNOUX[1,4], Christophe D'ENFERT[1], Melanie LEGRAND[1]

[1] Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, France
[2] Université Paris Diderot, Sorbonne Paris Cité, Paris, France
[3] Hub de Bioinformatique et Biostatistique, Département de Biologie Computationnelle, USR 3756 IP CNRS, Institut Pasteur, Paris, France
[4] University of Paris, Medical School, APHP, Necker-Enfants Malades Hospital, Parasitology Mycology Unit, Department of Microbiology, Paris, France

Abstract

The majority of studies investigating genetic diversity between several *C. albicans* strains are based on analysis of single nucleotide polymorphisms and copy number variation data. Despite LOH being ubiquitously found within clinical isolates, little descriptive data relating to LOH is available. Thus, we took advantage of the laboratory's collection of sequenced clinical *C. albicans* isolates to conduct a descriptive study of LOH landscape found within 54 natural strains from the genetic cluster 1, one of the major *C. albicans* population cluster. Our analysis pipeline permitted to identify a total of 7,025 LOH events within the 54 *C. albicans* strains allowing us to characterize the distributions of LOH throughout the genome. In cluster 1, the median LOH size is around 1 kb and the majority of these events are short-tract (<10 kb), limiting the loss of heterozygous SNPs and maintaining a relatively high overall heterozygosity level within the strain. The number of LOH correlates with chromosome size, where large chromosomes are most affected although all chromosomes often possess numerous events. Additionally, 275 distal LOH events, spanning from a given genomic locus until the telomere, were identified and are much larger in size (median size of 100 kb). Overall, we estimated a median of 128 LOH events per strain translating into approximately 9.4% of the genome impacted by LOH in cluster 1 of the *C. albicans* population. This type of genomic rearrangement is quite evenly distributed across the 8 chromosomes of the *C. albicans* genome, as no obvious LOH hotspots were identified. However, the distribution of LOH events revealed that genomic regions surrounding centromeres are poor in LOH while certain Major Repeat Sequences coincide with an increase in LOH frequency, suggesting that they could favor recombination. Altogether, this descriptive study allowed to depict LOH patterns within Cluster 1 of the *C. albicans* population, emphasizing the presence and potential importance of short-tract LOH as a long-term evolution strategy within a collection of natural isolates. The appearance of short-tract LOH is a powerful evolution strategy generating genetic variation while limiting the loss of heterozygosity in predominantly clonal *C. albicans.* We envision expanding this study to the other genetic clusters of the *C. albicans* population, in hopes to eventually identify and characterize cluster- or niche-specific LOH events potentially associated with adaptive phenotypes in *C. albicans*.

Introduction

Candida albicans is mostly found as a commensal of the human gastrointestinal tract although it can become pathogenic and be associated with a wide spectrum of infections, ranging from inoffensive lesions to life threatening systemic infections, notably a concern for immunocompromised individuals. The genetic diversity within the natural *C. albicans* population was assessed, using multilocus sequence typing (MLST) or single nucleotide polymorphism (SNP) calling data from whole-genome sequencing, to identify 18 genetic clusters (Bougnoux *et al.* 2002; Odds *et al.* 2007; Ropars *et al.* 2018). No obvious geographical or sampling origin associations were identified between strains of the latter genetic clusters. However, the intra-clade diversification of strains seems to be largely linked to the accumulation of SNP or loss-of-heterozygosity events (LOH) (Ropars *et al.* 2018). The overall population structure represents a clonal population although, evidences of admixture between genetic clusters have been observed (Ropars *et al.* 2018; Wang *et al.* 2018). Indeed, *C. albicans* does possess a parasexual life cycle permitting alleles shuffling, though meiosis has not been described (Bennett and Johnson 2003; Forche *et al.* 2008; Anderson *et al.* 2019). Additionally, several research teams have focused on analysis and comparison of whole-genome sequencing data, predominantly based on analysis of the diversity of SNPs and/or copy number variations (CNVs), in hopes to describe and better understand the biology of *C. albicans*. These studies revealed a relatively high average number of heterozygous SNPs across the diploid *C. albicans* genome, with one heterozygous position every ~250 bp (Jones *et al.* 2004; Hirakawa *et al.* 2015; Ropars *et al.* 2018). This elevated level of heterozygosity has been correlated to a strain's fitness (Hickman *et al.* 2013; Hirakawa *et al.* 2015). However, while mapping these heterozygous positions across the eight chromosomes, numerous regions of homozygosis (poor in heterozygous SNPs) interspersed between heterozygous genomic regions are observed, revealing that LOH are omnipresent within the genome of *C. albicans*. Occurrence of aneuploidy events, corresponding to an abnormal number of chromosomes, was also described in roughly 9% of strains within a collection of 182 clinical *C. albicans* isolates (Ropars *et al.* 2018). Overall, two hallmarks of the genome of *C. albicans* are (i) its elevated heterozygosity and (ii) its important tolerance for genomic rearrangements, including aneuploidy, isochromosome formation and LOH (reviewed in (Todd *et al.* 2017; Liang and Bennett 2019)). These traits are thought to facilitate the acquisition of adaptive phenotypes, potentially participating in the success of *C. albicans* as both commensal and pathogen (reviewed in (Beekman and Ene 2020)).

The description of LOH events found within natural strains of *C. albicans* has mostly been focused on their general distribution, highlighting that LOH are pervasive and can affect

all chromosomes (Hirakawa *et al.* 2015; Ropars *et al.* 2018; Wang *et al.* 2018). The distribution analyses were conducted by analyzing the frequency of genomic regions poor in heterozygous SNPs and led to the conclusions that regions surrounding centromeres appear to be refractory to LOH while telomeres appear to be the most affected by LOH events (Ropars *et al.* 2018; Wang *et al.* 2018). Indeed, repeat sequences also promote the appearance of LOH events, such as major repeat sequences (MRS) or inverted repeats (IR) (Ene *et al.* 2018; Todd *et al.* 2019). It was also noted that the majority of long LOH reach telomeres and, can be referred to as distal LOH events. However, these chromosomal rearrangements are more extensively described based on microevolution experiments conducted *in vitro* or *in vivo*, which follows the appearance of genetic variations over time using whole-genome sequencing data (Forche *et al.* 2018; Ene *et al.* 2018; Sitterle *et al.* 2019). Such studies revealed frequent appearance of base substitutions and small LOH events (1 bp – 3 kb), while long-tract LOH are scarcer, and notably highlighted the importance of short-tract LOH events which were overlooked by antecedent studies (Ene *et al.* 2018; Sitterle *et al.* 2019). Although SNPs are acquired at a higher rate than LOH events over the course of time, the overall number of impacted bases is comparable between mutation types (Ene *et al.* 2018). Similar in-depth characterization of all types of LOH found within natural isolates of *C. albicans* has not been conducted in order to investigate if similar LOH patterns exist within the natural population.

We took advantage of the whole-genome sequencing data resources available within the laboratory to conduct a small study aiming at characterizing the patterns of LOH events within a natural population of *C. albicans* strains. We decided to focus on a single genetic cluster from the *C. albicans* population, Cluster 1, because i) it contains the broadly utilized laboratory reference strain (SC5314), ii) it seems to possess a high nucleotide diversity index according to the population genomics study of Ropars *et al.* (Ropars *et al.* 2018) and iii) many strains corresponding to this genetic cluster have been sequenced. We selected 54 strains from diverse geographic locations and various sampling origins to characterize LOH and study genomic distribution of LOH. To robustly identify LOH events within this collection of Cluster 1 strains, we constructed a highly heterozygous reference genome, named the “Root” genome, deprived of all LOH events likely to play a role in Cluster 1 diversification. We identified a total of 7,025 LOH within the 54 *C. albicans* strains, resulting in a median of 128.5 LOH events per strain, covering roughly 9.4% of the genome. The majority of LOH are short-tract (≤10kb) while distal LOH are larger and represent 3% of LOH events identified within the collection of cluster 1 strains. Additionally, we observed that centromeric regions are poor in LOH, while LOH frequency seems to increase along the chromosome arms towards telomeres. Certain MRS may also be highly recombinogenic, as abrupt increases of LOH

frequency coincided with their location. Overall, this study provides a holistic snapshot of LOH in *C. albicans* isolates from the genetic cluster 1.

Results

A collection of 54 *C. albicans* strains for sampling genetic diversity within Cluster 1

A total of 54 *C. albicans* strains were selected from the laboratory's collection of sequenced natural isolates in order to characterize LOH and their distribution within genetic cluster 1 of the *C. albicans* population. These strains originate from five continents although Europe and Africa are most represented (Table S1). The strains were also isolated from different sampling sites such as, spoiled food, patients with superficial and invasive infections or healthy individuals (Figure 1). To observe the relative distance between DNA sequences of theses 54 *C. albicans* strains, we constructed a distance phylogenetic tree based on previously generated SNP calling data available in the laboratory. As illustrated in Figure 1, this collection of strains is relatively diverse although certain strains display high similarities at the nucleic acid level e.g. strains CEC4257 and CEC4256. Furthermore, we do not observe any obvious strong regrouping of strains associated to sampling origin (Figure 1).

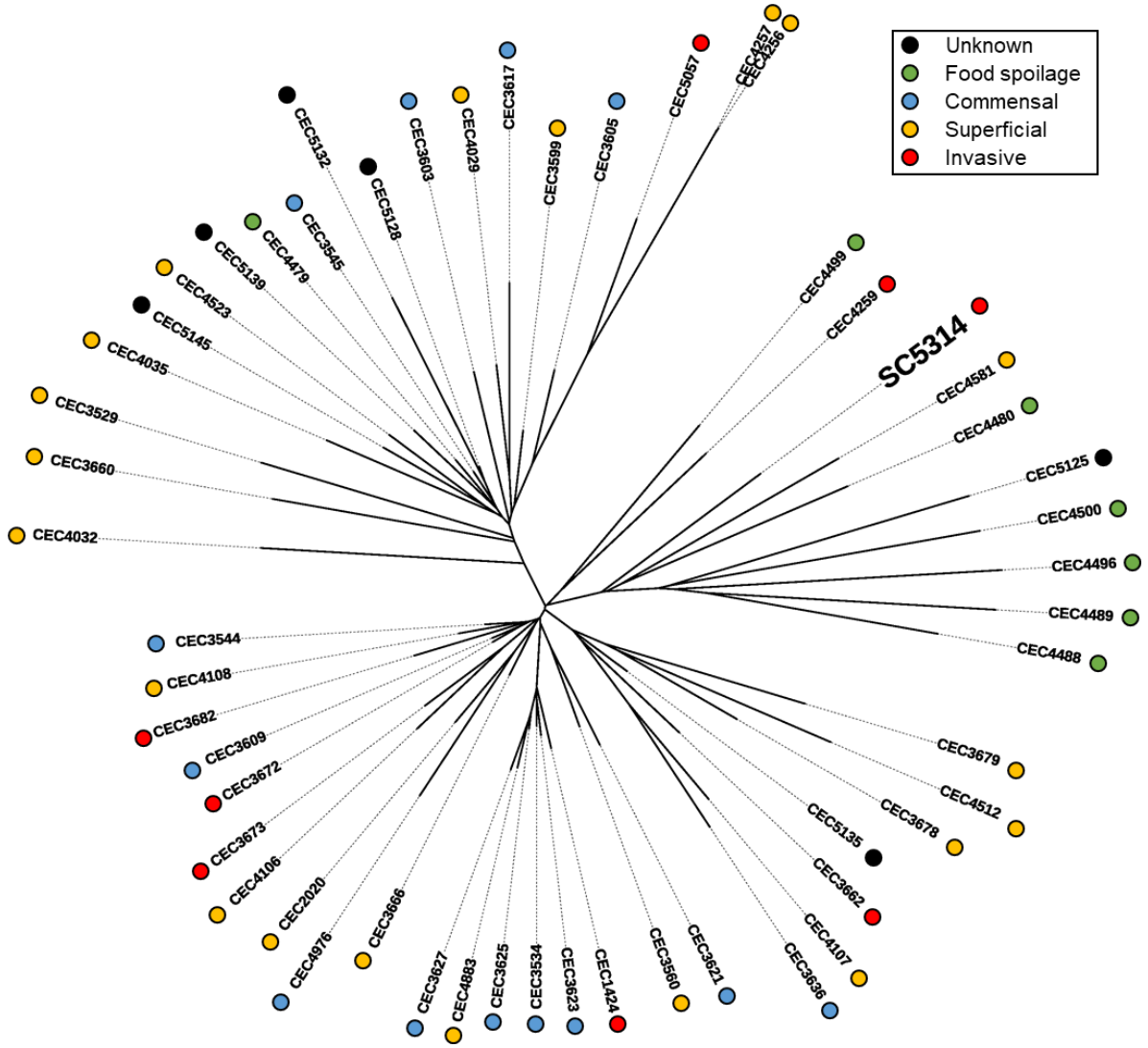

Figure 1: Phylogeny representing distance of nucleic acid sequences between the 54 genetic cluster 1 C. albicans strains.
Distance phylogeny depicting extent of differences at the nucleic acid level between the 54 cluster 1 *C. albicans* strains included in this study. Sampling site origins are indicated by different colored circles: black, unknown origin; green, spoiled food; blue, commensal; yellow, superficial infection and red, invasive infection. The *C. albicans* laboratory reference strain, SC5314, is indicated in bigger font.

Reconstruction of a root ancestral genome to identify cluster 1 LOH events

Because of pre-existing LOH events within each one of the 54 cluster 1 strains, we would bias our analysis by assigning one specific strain from the 54 as a reference for the identification of LOH events. Thus, we constructed an artificial reference genome which we named root genome. This root genome represents the most heterozygous strain in cluster 1, somewhat a type of ancestral genome deprived of the LOH events which are observable within this genetic cluster. We initially constructed several root genomes by varying the percentage of strains required to judge a given position to be heterozygous in the root genome. By plotting the differences in the number of heterozygous positions obtained between root genomes

(Figure 2A), we tried to discriminate between heterozygous SNPs which are associated with the random appearance of point mutations during cluster expansion and those which could be considered as “natural” heterozygous positions present in an ancestral strain of cluster 1. Indeed, as our stringency increases, we lose heterozygous positions within the root genome (Figure 2A). Our interpretation of Figure 2A proposes that additional heterozygous SNPs in Roots10 to 25 (where 10 to 25% of the 54 strains must be heterozygous in order for a given position to be assigned as heterozygous in the root) could be explained by the independent accumulation of point mutations in our collection of genomes. Building root genomes using a >30% cut-off, we observed a somewhat exponential decrease of heterozygous positions most likely associated to the appearances of LOH events (Figure 2A). Thus, we selected Root30 as our reference genome for further analysis, in which a heterozygous position was assigned in the Root genome if ≥ 30% of strains were heterozygous at this given position. Our selected reference genome Root30, possesses 78,144 heterozygous positions across the 8 chromosomes (Chr), representing a heterozygous SNP every ~ 182 bp. This heterozygosity level is a little above the previously described levels of heterozygosity within the overall population of *C. albicans* (1 heterozygous SNP every 200 bp) (Hirakawa *et al.* 2015; Ropars *et al.* 2018). As illustrated by the heterozygous SNP density heatmap (Figure 2B), heterozygous SNPs in Root30 are evenly distributed throughout the genome except for the right arms of Chr3, Chr6 and ChrR which remain highly homozygous.

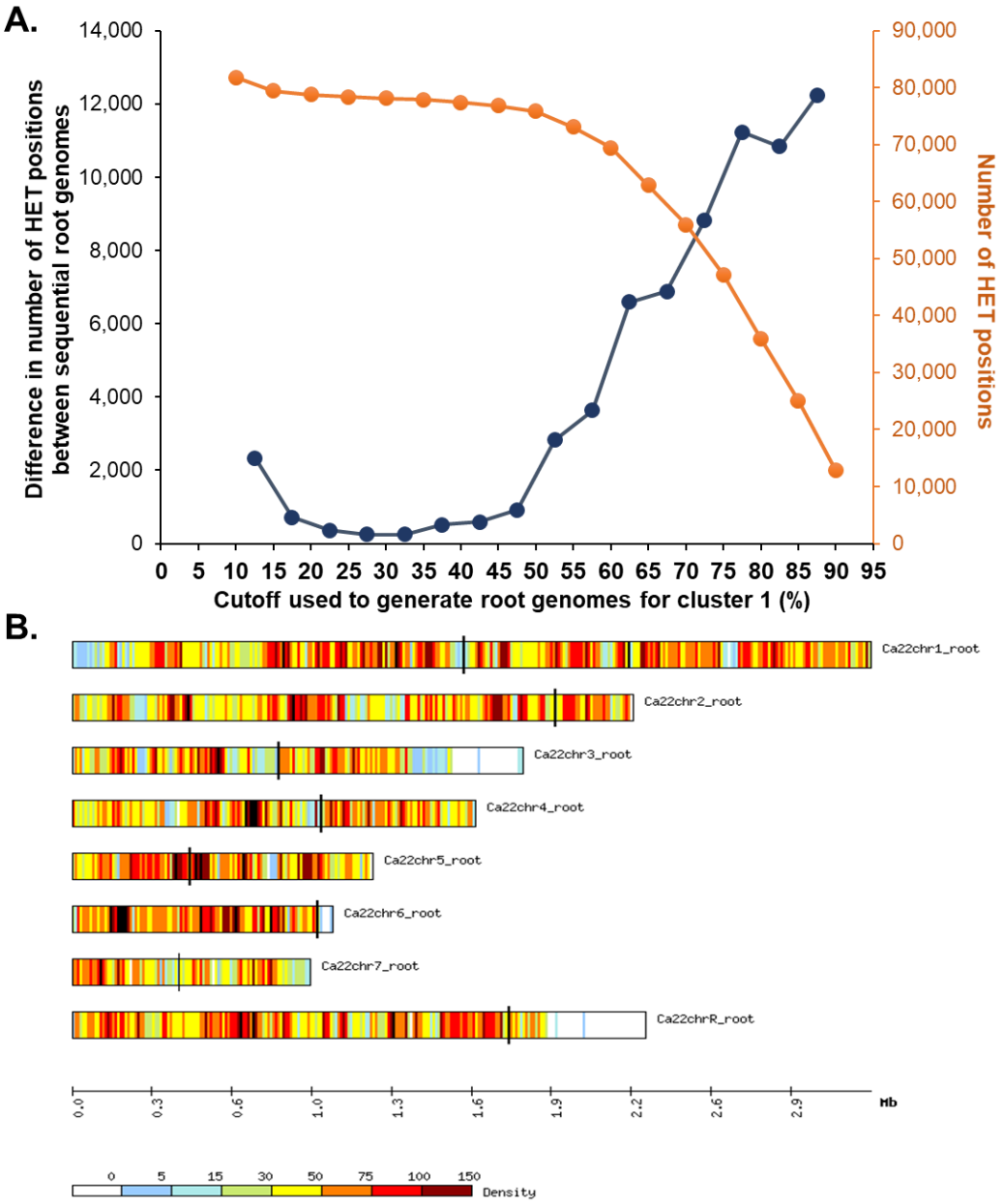

Figure 2: The root genome, an artificial reference genome built for analysis of LOH events in genetic cluster 1 of the C. albicans population.
A. Comparison of several Root genomes built using different cut-offs in terms of percentages (10.01 – 90%) of strains required to judge a given position to be heterozygous in the Root genome. Note that we always considered a position to be homozygous if 90% of strains were homozygous for the given positions, while only varying our heterozygosity cut-off between all root genomes. Plot represents the number of heterozygous (HET) positions (right orange Y axis) or the difference in number of HET positions between sequential root genomes (left black Y axis) within/between each Root genome (X axis). **B.** Heatmap representing the distribution of heterozygous SNPs within the Root30 genome (78,144 heterozygous positions), subsequently used as a reference genome for the identification of LOH events across the eight chromosomes. Vertical black bars represent centromeres.

Using our artificial reference genome, we then sought out to identify LOH events within our collection of 54 *C. albicans* strains from genetic cluster 1. We chose to base our identification of LOH only on high confidence heterozygous positions within all 54 strains, i.e. a total of 33,137 positions across the eight chromosomes, allowing for a robust analysis where

each strain undergoes the same comparison with the root, avoiding any biases associated with differences in sequencing quality amongst strains. Additionally, we decided to mask annotated repeat sequences found throughout the genome because repeat regions assembly is notoriously poor, which could potentially impact our analysis. The boundaries of LOH events were defined as regions of transitions between heterozygous to homozygous and/or homozygous to heterozygous stretches. These boundaries are identified in each strain, upon individual strain comparison with the Root30 genome, and are composed of three consecutive heterozygous positions (≥ 300bp) (H/H) followed by three consecutive homozygous positions in the strain of interest that are heterozygous in Root30 (≥300bp) (H/h) (Figure 3). According to the genomic coordinates of each most junction proximal SNP, an LOH can be localized within the genome and its size can be calculated. Two types of LOH size can be computed with the latter coordinates LOH_{max} or LOH_{min}, as the difference between right and left positions of either junction delimiting H/H or H/h positions, respectively (Figure 3). For the purpose of this study, we solely focused on LOH_{min} size in order not to overestimate LOH sizes. By requiring three consecutive SNPs within a ≥ 300 bp region for each LOH junction, we ensure that non-annotated repeat sequences do not impact our LOH analysis.

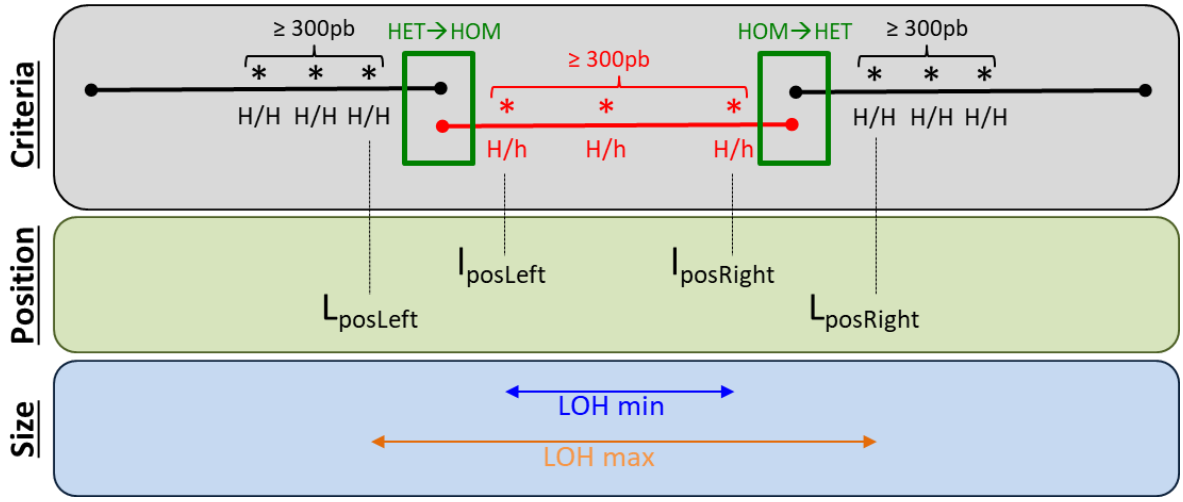

Figure 3: Criteria used to identify LOH events throughout the 54 C. albicans strains of genetic cluster 1.
An LOH event is defined by its boundaries of transition between heterozygous to homozygous (HET → HOM) and/or homozygous to heterozygous (HOM → HET) stretches, illustrated in green. Each boundary consists of three consecutive heterozygous positions (≥ 300bp) (H/H) followed by three consecutive heterozygous positions in Root30 that are homozygous in the strain of interest (≥ 300bp) (H/h). Utilizing the coordinates of LOH boundary proximal h/h or h/o SNPs (section "position"), the LOH_{max} or LOH_{min} sizes can be calculated.

LOH landscape across Cluster 1

Following the above presented criteria, a total of 7,025 LOH events were identified within the 54 *C. albicans* strains, translating into a median of 128.5 LOH per strain (min= 49 LOH in strain CEC3666 and max= 237 LOH in strain CEC 4257) (Figure 4A). These LOH possess various sizes spanning from 300 bp (minimum size defined by the criteria used for LOH identification) to 1,737 kb with an overall median size of 1.2 kb (Figure 4B). The median LOH sizes between the 8 chromosomes vary significantly (Kruskal-Wallis Test, $p<0.0001$), as Chr4 and Chr7 display slightly longer LOH events while median LOH on ChrR are shorter as compared to overall median (Figure S1). However, distributions of LOH sizes per chromosome (Figure 4B & S1) are comparable to the overall LOH size distribution illustrated in Figure 4B. The number of LOH events varies significantly between chromosomes (Kruskal-Wallis Test, $p<0.0001$) and is positively correlated with chromosome size (Pearson's correlation, $R^2=0.9632$, $p<0.0001$) (Figure 4D-E) and consequently the number of ORFs per chromosome (Pearson's correlation, $R^2=0.9685$, $p<0.0001$)(Figure S2). There are roughly 14 LOH events per chromosome (Figure 4D), and 2.7 % of a chromosome is impacted by LOH-associated homozygosis. Within this collection of strains, LOH tracts cover roughly 9.4% of the genome/strain in terms of number of nucleotides, where CEC3627 (0.75%) is least and CEC3617 (28.06%) is the most impacted by LOH (Figure 4C). We observe a negative correlation between the cumulative LOH size per strain and its average number of heterozygous SNPs/kb (Pearson's correlation, $R^2=0.5414$, $p<0.0001$) (Figure S3), suggesting that we efficiently identified LOH events within our collection of strains.

We also investigated the frequency of LOH size (Figure 4F) by dividing LOH sizes into three categories: short-tract (≤10 kb), intermediate-tract (>10kb - ≤100 kb) and long-tract (>100 kb). In *C. albicans*, heterozygosity of a strain is associated to fitness (Hickman *et al.* 2013; Hirakawa *et al.* 2015; Ropars *et al.* 2018), the frequency we observed in terms of LOH size illustrates the fact that heterozygosity tends to be maintained as the vast majority of LOH events are short-tract (95.6%), while the remaining 4.4% comprise intermediate- and long-tract LOH (Figure 4F). These short-tract events are conceivably outcomes of DNA break repair through gene conversion, a homologous recombination dependent repair pathway. Globally, each strain possesses numerous LOH events which are predominantly short and quite evenly distributed across the genome.

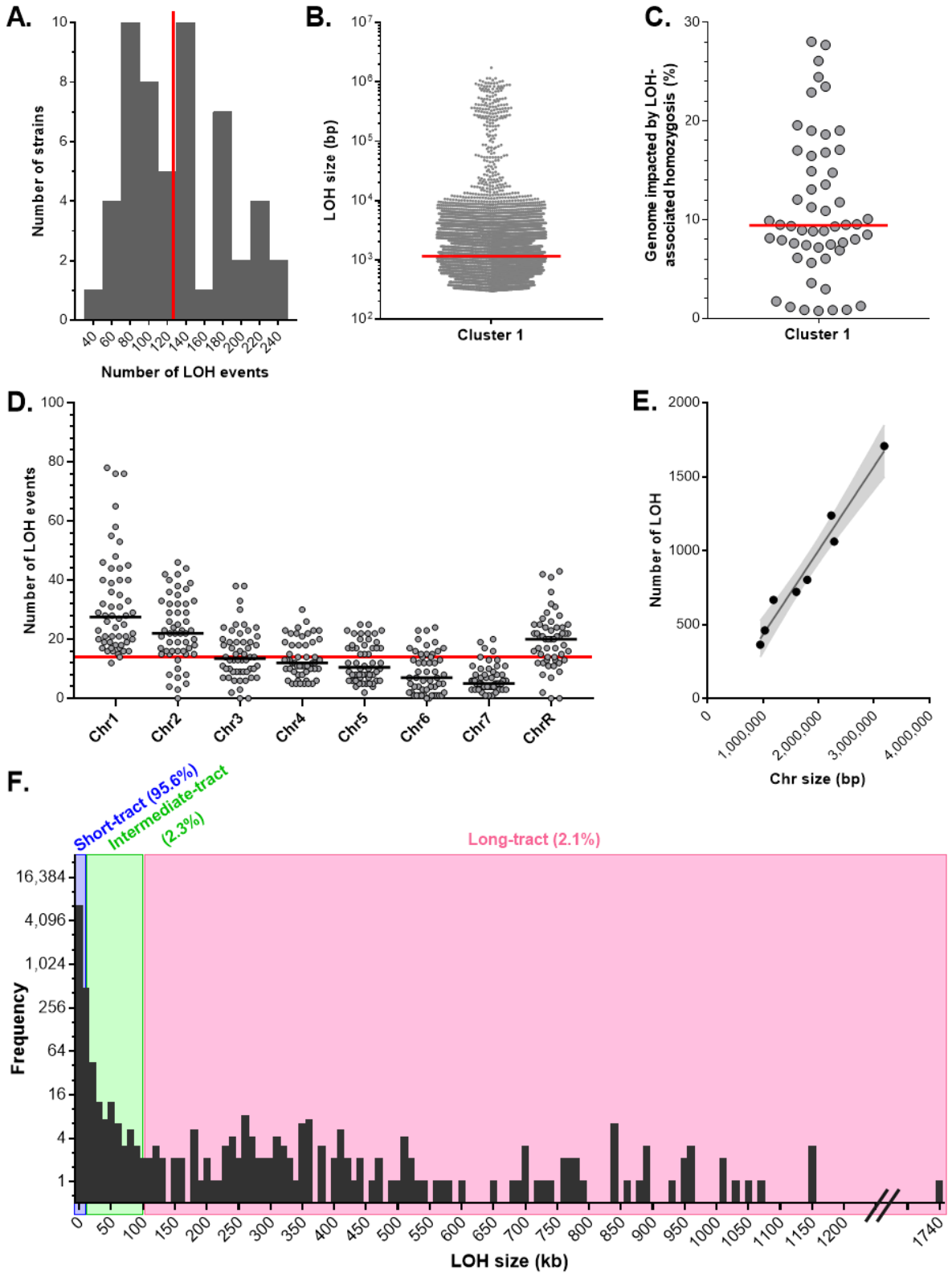

Figure 4: Characterization of LOH events within genetic Cluster 1 of the C. albicans population. **A.** Frequency of the number of LOH events per strain (n=54), vertical red line indicates median number of LOH events per strain. **B.** Distribution of LOH sizes (bp) identified within the collection of strains (n=7,025), median LOH size is illustrated by the red line. **C.** Plot of the percentage of the genome impacted by LOH-associated homozygosis within strains of cluster 1 (n=54), each dot represents a strain while the red line indicates the median percentage of the genome implicated in LOH per strain. **D.** Each circle represents the number of LOH event per chromosome for a given strain. The median

number of LOH per chromosome (n=54) is indicated by the gray horizontal lines while overall median number of LOH per chromosome is marked by the red line. **E.** Relationship between the median number of LOH events per chromosome and the chromosome size (n=8). The black line represents the linear correlation (95% confidence interval, gray shade). **F.** Distribution of LOH size (kb) frequency for the 7,025 LOH events identified. LOH were divided into three categories: short-tract (≤10kb) in blue, intermediate-tract (>10kb - ≤100kb) in green and long-tract (>100kb) in pink. The percentage of LOH events found within each category is indicated above graph.

Distal LOH landscape across Cluster 1

Following this analysis, we were specifically interested by the patterns of distal LOH events within the genetic cluster 1 of *C. albicans*. A distal LOH event is described as a homozygosis tract spanning from a given point of the chromosome reaching until the telomere, thus it defines a single event that can impact thousands of bases and has a greater probability of impacting the cell's biology. We determined that 209 of the 7,025 events were distal LOH events and were encountered within 49/54 strains, resulting in a median of 4 distal LOH events per strain within genetic cluster 1 (Figure 5A). These events are distributed quite evenly across chromosome arms (Figure 5B). Only the left arm of Chr6 does not display any distal LOH event, probably because this chromosomal arm is already homozygous in Root30, the artificial reference genome (Figure 2B). The identified distal LOH events range in size from 774 bp (Chr3R) to 1,737 kb (Chr2L, longest chromosome arm) (Figure 5C). As described in the above section, we analyzed the distribution in distal LOH size by categorizing events into three types. In contrast to overall LOH size distribution, we noticed that the majority of distal LOH events correspond to long-tract (69.9%) and intermediate-tract (23.4%) while, short-tract distal LOH represent only 6.7% of events (Figure 5D). Indeed, distal LOH events often result from DNA repair pathway of break-induced replication or mitotic crossover which lead to homozygosis until the telomere. No obvious correlation between number of distal LOH events and chromosome arm length was obtained however, median distal LOH size per chromosome arm is positively correlated with the chromosome arm length (Figure S4).

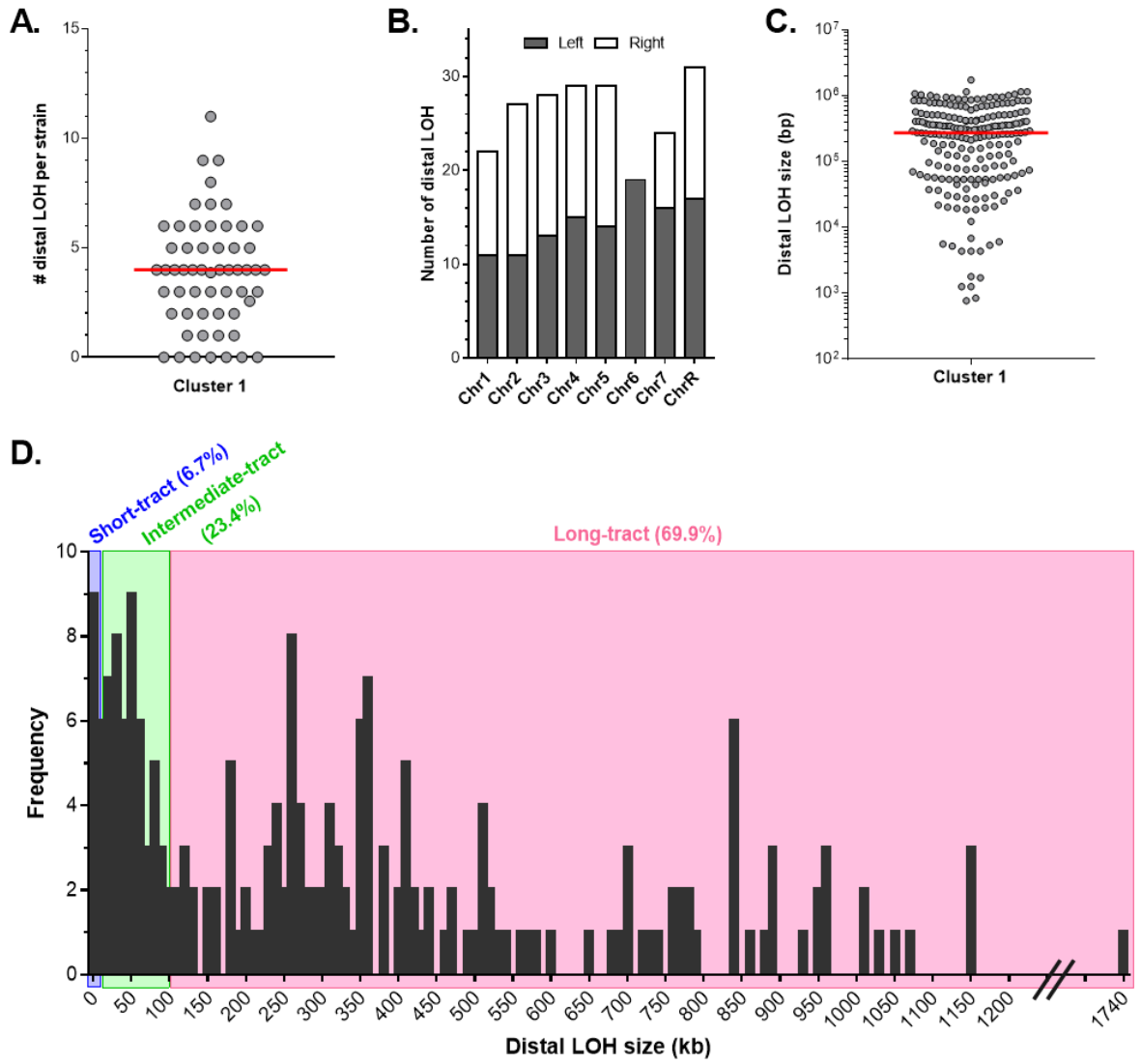

Figure 5: Characterization of distal LOH events within genetic Cluster 1 of the C. albicans population.
A. The number of distal LOH events per strain (n=49), red line indicates median number of distal LOH events per strain. **B.** Graph of distal LOH events (n=209) across chromosome arms, left arm (gray) and right arm (white). **C.** Plot of distal LOH sizes identified, each circle represents a distal LOH (n=209) while the red line indicates the median length of events. **D.** Distribution of distal LOH size (kb) frequency for the 209 distal LOH events identified. LOH where divided into three categories: short-tract (≤10kb) in blue, intermediate-tract (>10kb - ≤100kb) in green and long-tract (>100kb) in pink. The percentage of LOH events found within each category is indicated above graph.

Genomic distribution of LOH

Because LOH events have been associated to the acquisition of adaptive phenotypes, we aimed at exploring the presence of LOH hotspots across the genome in order to identify genomic regions which potentially favor the appearance of LOH. We plotted the frequency of LOH across the genome (by 10 kb windows) within the 54 strains (Figure 6A) and observed that LOH events are somewhat evenly distributed throughout the genome although, certain

outlier regions can be located. We noted that many telomere regions, end of chromosomes, are represented in white (Figure 6A) because these regions were either (i) already homozygous in Root30 (Figure 2B), or (ii) they underwent infrequent LOH. We then identified 21 LOH "hotspots", 10 kb long and distributed across seven chromosomes (absence of Chr7), in which more than 25 LOH events were identified (Table S2). Upon investigation of their genomic content, no striking feature type or biological process was detected, suggesting that these hotspots may either be associated to DNA fragile sites or be random. Although, to be noted that roughly half of the genome features located within these "hotspots" are uncharacterized in *C. albicans I* ORFs, potentially biasing this analysis.

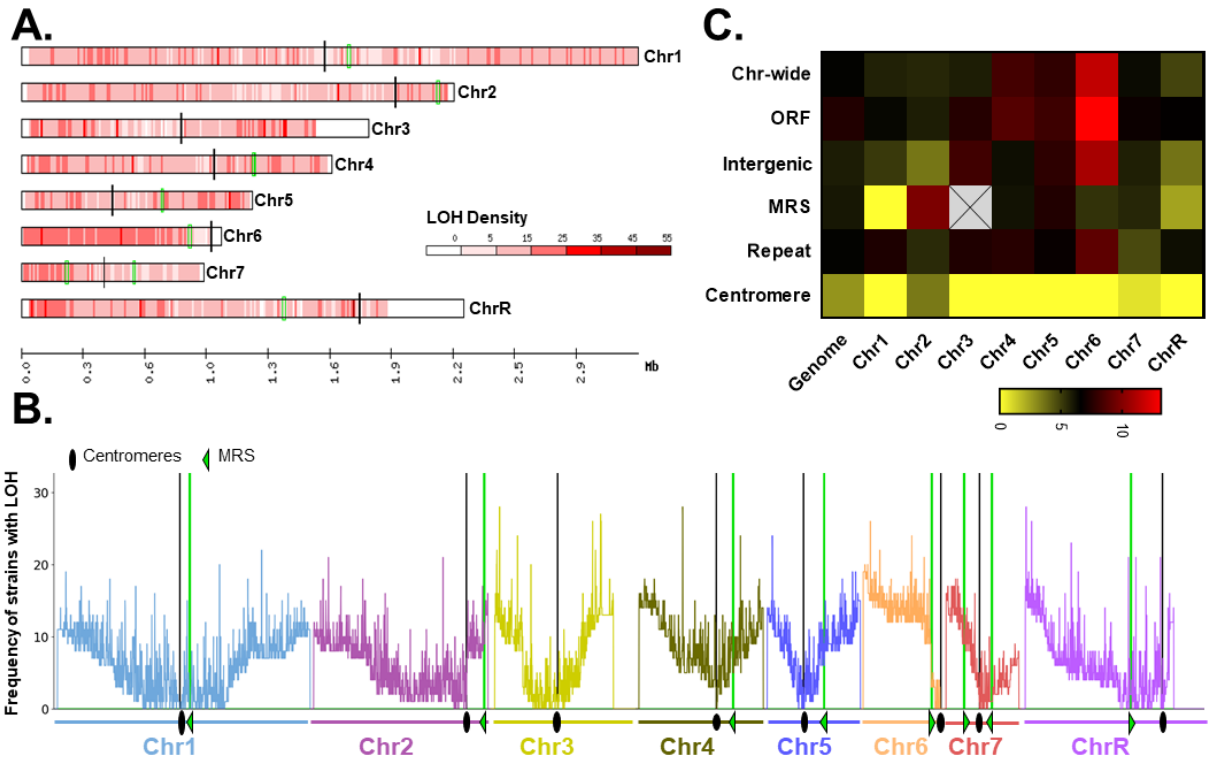

Figure 6: Distribution of LOH in C. albicans genetic cluster 1.
A. Heatmap representing LOH density across 10 kb windows along the eight chromosomes (n=7,025). Centromeres are indicated by black vertical bars while the location of Major Repeat Sequences (MRS) is indicated in green. **B.** Distribution curve of the frequency of strains (n=54) possessing LOH across 2 kb windows along the 8 chromosomes. Each chromosome is illustrated by a given color while black and green vertical lines represent centromere and MRS locations, respectively. **C.** Heatmap of the mean LOH frequency per kilobase calculated genome-wide and across each chromosome, to allow comparison across chromosomes and several feature types. The row entitled Chromosome-wide (Chr-wide) takes into account all LOH covering entire chromosomal DNA during frequency calculation, while mean frequency values for the other rows were calculated based on only LOH in coding regions (ORFs), non-coding regions (intergenic), all features annotated as repeats in Candida Genome Database (Skrzypek *et al.* 2017), MRS or centromere regions. Chr3 does not possess a complete MRS thus gray box marked with an X represents an absence of a mean LOH frequency/kb value.

By plotting the frequency of strains displaying LOH across the genome, in 2 kb windows, we noticed that LOH frequency tends to increase along each chromosome arm, whereas the centromere regions are often among the poorest in LOH (Figure 6B). Additionally, we observed that some MRS positively impact LOH frequency, notably *mrs-2*, *mrs-5* and *mrs-6* (Figure 6B), as the location of these repeat sequences coincide with abrupt increases in frequency along the chromosome (Figure 6B). Thus, certain MRS appear to be hotspots for recombination, often leading to LOH. Lastly, we calculated and compared the mean LOH frequency per kilobase across different regions of the genome, e.g. ORFs, intergenic regions, centromeres, repeats and MRS (+/- 5kb) (Figure 6C). With this analysis, we confirmed that all centromere display on average less LOH than other genomic regions, independently of the chromosome. As illustrated by the heatmap in Figure 6C, Chr6 and ChrR seem to display the highest and lowest LOH frequencies per kilobase, respectively, while MRS from different chromosomes demonstrate mixed frequencies of LOH (Figure 6C). Otherwise, no obvious striking differences in LOH frequencies per kilobase were observed using this approach although sub-types may enrich for LOH, something that remains to be assessed.

Discussion

Although LOH play an important role in the biology of *C. albicans*, few studies have focused on a detailed description of the genomic rearrangements within a collection of clinical isolates. We took advantage of the laboratory's large collection of sequenced clinical *C. albicans* isolates to conduct a descriptive analysis of LOH events within genetic cluster 1. Throughout this study, we established a pipeline allowing a robust identification of LOH, displaying a minimum size of 300 bp, and thoroughly characterized the latter genomic rearrangement events within cluster 1 of the *C. albicans* population. We highlighted that each strain analyzed displays numerous ubiquitous LOH although, high levels of heterozygosity are maintained as most LOH are short-tract events (≤ 10 kb). Nevertheless, long-tracts LOH were also identified and often characterized as distal LOH. Distribution of LOH in Cluster 1 revealed no obvious hotspots though, few MRS seem to positively influence LOH frequency while LOH are infrequent at centromeres. Overall, our data highlights that the population *C. albicans* predominantly evolves, in terms of LOH, using small genomic changes (similarly to laboratory evolution experiments conducted by other research teams (Ene *et al.* 2018)) and that LOH are stochastic even though certain genomic features and parameters may influence their appearance.

The genome of *C. albicans* is characterized by relatively elevated levels of heterozygosity and a high tolerance to genomic rearrangements (Jones *et al.* 2004; Fischer

et al. 2006; Hirakawa *et al.* 2015; Ropars *et al.* 2018; Wang *et al.* 2018), in particular LOH events. In a diploid organism, LOH involves the loss of one of two genomic copies that differ at the nucleotide level. LOH in *C. albicans* are considered to be pervasive and of high importance, as the totality of strains possess such genomic rearrangements and LOH have at time been associated with the acquisition of adaptive phenotypes (reviewed in (Beekman and Ene 2020)). Despite the rich accessibility to whole-genome sequencing data, few studies have focused on the characterization of LOH events. Only in recent years, accurate LOH characterization has emerged in micro-evolution studies conducted *in vivo* and *in vitro* (Ene *et al.* 2018; Sitterle *et al.* 2019), aiming to identify the contribution of these genomic rearrangements to the evolution of *C. albicans*. These authors highlighted the presence and the potential importance of short LOH events which were not specifically identified nor analyzed in prior genome comparative studies. Indeed, diversity of natural *C. albicans* isolates has been thoroughly described using both MLST and SNP/CNV calling data of whole genome sequencing, demonstrating that *C. albicans* strains regroup into distinct genomic clusters independently of strain origin (Odds *et al.* 2007; Ropars *et al.* 2018; Wang *et al.* 2018). Differentiation and diversification of genetic clusters is associated with the accumulation of SNPs and the acquisition of LOH within this predominantly clonal population (Ropars *et al.* 2018; Wang *et al.* 2018). However, these authors mostly focused on SNP data and did not specifically aim to identify and characterize all LOH within their strain collections thus, general traits of LOH within natural *C. albicans* strains remained elusive.

Consequently, we took on the project of exploring the characterization of LOH within the *C. albicans* population by focusing on a single genetic cluster and using already available sequencing data in the laboratory. We selected one of the major genetic cluster inside the population, Cluster 1, as many genomes were available and because the nucleotide diversity (Π) within this cluster was estimated to be the highest (Π = 0.478) (Ropars *et al.* 2018). A total of 54 *C. albicans* strains from Cluster 1, from different origins (Table S1), were chosen for later analysis. Using a distance phylogeny, we illustrated that strains do not cluster together by origin within this genetic cluster (Figure 1), similarly to observations conducted at the population level (Ropars *et al.* 2018). The heterozygosity represented in these divergent strains was then used to construct a highly heterozygous ancestral genome (Root), theoretically deprived of LOH which emerged within this Cluster 1 (Figure 2). We subsequently used this genome as a reference in order to identify LOH events inside each of the 54 strains. We have inventoried 7,025 LOH, where each strain displays numerous LOH impacting all chromosomes with a median of 128 LOH/strain (Figure 4), confirming that LOH are frequent within the *C. albicans* population. The vast majority of these LOH events (95.6%) were characterized as short-tract LOH spanning a maximum length of 10 kb (Figure 4F).

Additionally, numerous distal LOH events were uncovered during this analysis, with a median length 235x longer than the overall median LOH size (median LOH size: All = 1,155 bp, Distal = 271,816 bp). Distal LOH events often result from break-induced replication or mitotic crossover events leading to long LOH spanning from a given locus of the chromosome until the telomere. LOH facilitate the passage of recessive beneficial alleles through Halden's sieve (Gerstein *et al.* 2014) and can purge recessive lethal alleles from the population (Feri *et al.* 2016; Marton *et al.* 2019), both of high importance within a clonal population. Nevertheless, strains of *C. albicans* displaying greater levels of heterozygosity are considered to possess a higher fitness thus, maintenance of heterozygosity is also crucial for this opportunistic pathogen (Hickman *et al.* 2013; Hirakawa *et al.* 2015). Indeed, short-tract events of LOH would allow *C. albicans* to sustain an overall relatively high heterozygosity level, while still permitting to potentially uncover/purge recessive alleles. Appearance of short LOH was also described as a safe strategy to rapidly evolve upon small-scale evolution experiments in *C. albicans* (Ene *et al.* 2018; Beekman and Ene 2020). Together with our results, appearance of short-tract LOH appears to be a powerful evolution strategy, participating to both short- and long-term evolution of *C. albicans*.

As previously mentioned, certain LOH have been associated with the acquisition of adaptive phenotypes such as gain of antifungal resistance, increased virulence or adaptation towards commensalism (Coste *et al.* 2007; Tso *et al.* 2018; Liang *et al.* 2019). Hence, we explored the distribution of LOH events within our collection to identify hotspots and cold regions of LOH. Our distribution study of LOH events in Cluster 1, revealed that LOH seem to be quite evenly distributed across the 8 chromosomes although certain regions displayed more than 25 LOH events within 10 kb windows (Figure 6A). However, these regions did not display any striking feature type or biological process thus, are either random loci enriched for LOH or potentially associated to other parameters promoting LOH, e.g. DNA fragile sites or origins of replication. Similarly to previous observations, centromere are poor in LOH (Figure 6B-C) (Ropars *et al.* 2018; Ene *et al.* 2018; Wang *et al.* 2018) and LOH frequency increases with distancing from centromeres (Figure 6B) (Forche *et al.* 2011; Peter *et al.* 2018; Ene *et al.* 2018). Additionally, certain MRS may act as hotpots for recombination (Figure 6B), although our data shows different MRS than those described by Ene *et al.* (Ene *et al.* 2018), which may be potentially explained by differences in environmental exposures of strains between the two studies as LOH are influenced by different stresses (Forche *et al.* 2011). Overall, the general absence of striking hotspots reinforces the idea that LOH are generally stochastic genomic rearrangement events although, certain parameters can potentially mildly impact LOH frequency. Such parameters could potentially be observed upon a more refined analysis in the near future where feature types are broken-down into sub-types or other genomic

features are investigated, e.g. certain types of repeats, ORF families, DNA fragile sites, replication origins. Further analysis should also aime at identifying precise LOH cold regions which could eventually lead to the uncovering of overdominant loci, potentially participating to the maintenance of heterozygosity in *C. albicans,* comparably to observations made in clonal bee population (Goudie *et al.* 2014).

To conclude, this short descriptive study allowed to describe LOH patterns within the Cluster 1 of the *C. albicans* population, emphasizing the presence and potential importance of short-tract LOH as a long-term evolution strategy within a collection of natural isolates. Ideally, this study is to be expanded to the other major genetic clusters of the *C. albicans* population in order to have a broader and more complete view and, to potentially uncover further patterns of LOH notably, between genetic clusters or sampling origins.

Material and methods

Strains

Summary of the information regarding the 54 *C. albicans* clinical isolates from the genetic cluster 1 can be found in Table S1.

Phylogeny

A distance phylogenetic tree was constructed (BioNJ (neighbor-joining), JC (Jukes-Cantor, all mutations equal likelihood), no gaps) using Seaview to infer divergence at the nucleotide level between the 54 *C. albicans* isolates using the dataset of 54,613 confident SNPs (obtained using the same method as (Ropars *et al.* 2018)).

Root genome construction

The 54,613 confident SNPs dataset, mentioned above, was used to infer an ancestral Root genome theoretically deprived of LOH events. The comparative analysis of the SNPs allowed to define homozygous positions, if more than 90% of the strains are homozygous at these positions, or heterozygous positions if more than 30% of the strains are heterozygous. Before selection 30% as out final heterozygous criteria, we built multiple Root genome by changing this latter criterion of the heterozygous selection from 10.1% to 90%, by 5% increments. Density heatmap heterozygous SNPs within selected Root (30%) genome was built with 1kb windows with a custom script in Python and with Matplotlib library (Hunter 2007).

Identification of LOH

The identification of LOH events across the collections of *C. albicans* genomes was conducted using custom scripts (available upon demand). Initially, we extracted the heterozygous positions within the selected Root genome. Each one of these positions were compared between the Root genome and each strains and defined as h/h, if both the Root and the given strain are heterozygous at this position, or h/o, if the Root is heterozygous but the given strain is homozygous for this positions. Homozygous and heterozygous SNPs were defined using same criteria as Ropars *et al* (Ropars *et al.* 2018). All genomic regions annotated as repeats (Skrzypek *et al.* 2017) and MRS were masked before identifying LOH boundaries. LOH events are defined by LOH boundary(ies), region(s) of transition between heterozygous to homozygous and/or homozygous to heterozygous stretches. These boundaries are identified for each of the 54 isolates using comparison data. A boundary is composed of three consecutive h/h positions (≥ 300 bp) followed by three consecutive h/o positions (≥300bp) (Figure 3). Similarly to the analysis conducted by Todd *et al*. (Todd *et al.* 2019) we reinforced confidence of LOH boundaries by requiring the three consecutive SNPs to be located in a

genomic distance greater than 300 bp in order to ensure that non-annotated repeat sequences do not impact our LOH analysis. We relied on the genomic coordinates of each most junction proximal SNP to (i) locate LOH event and (ii) assess its size. The maximum LOH size (LOH_{max}) of an event is equal to the difference between the junction proximal h/h SNPs and the minimum LOH size (LOH_{min}) is equal to the difference between the junction proximal h/o SNPs (Figure 3).

Analysis of LOH

Analysis of LOH events was conducted using excel, Prism7 and Python custom scripting with Matplotlib library (Hunter 2007).

Data access

All data is available upon demand.

Acknowledgments

T.M., C.M., M.L., and C.D. designed experiments. T.M. and C.M. performed experiments. T.M., C.M., M-E.B., M.L. and C.D. analyzed data. T.M., M.L., and C.D. wrote the manuscript.

T.M. is the recipient of a PhD. fellowship from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID) and Indo-French Centre For The Promotion Of Advanced Research CEFIPRA. We acknowledge support from the French Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases [ANR10-LABX-62-IBEID]) and of the Indo-French Centre For The Promotion Of Advanced Research CEFIPRA.

Supplementary material

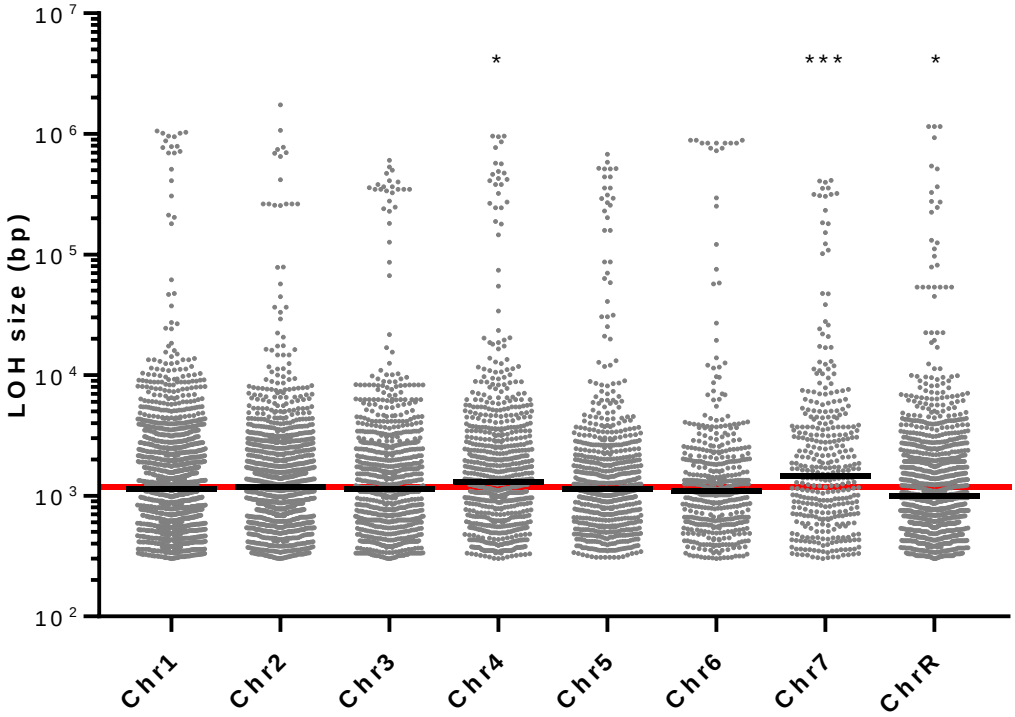

Figure S1: Distribution of LOH size by chromosomal location identified within the 54 C. albicans isolates of Cluster 1.
Each circle represents the size (bp) of an LOH event identified within the collection of genetic cluster 1 *C. albicans* isolates. Median LOH size per chromosome is identified by black horizontal lines while the overall median LOH size between chromosomes is illustrated by the red horizontal line. Statistical analysis using Krustkal-Wallis test ($p<0.0001$) determines that differences in median LOH size (bp) are significantly different between chromosome while, Mann-Whitney test between median LOH size of a chromosome and the overall median LOH size (red line) identifies Chr4, Chr7 and ChrR as possessing significantly different median LOH sizes, *: $p<0.05$, ***: $p<0.001$.

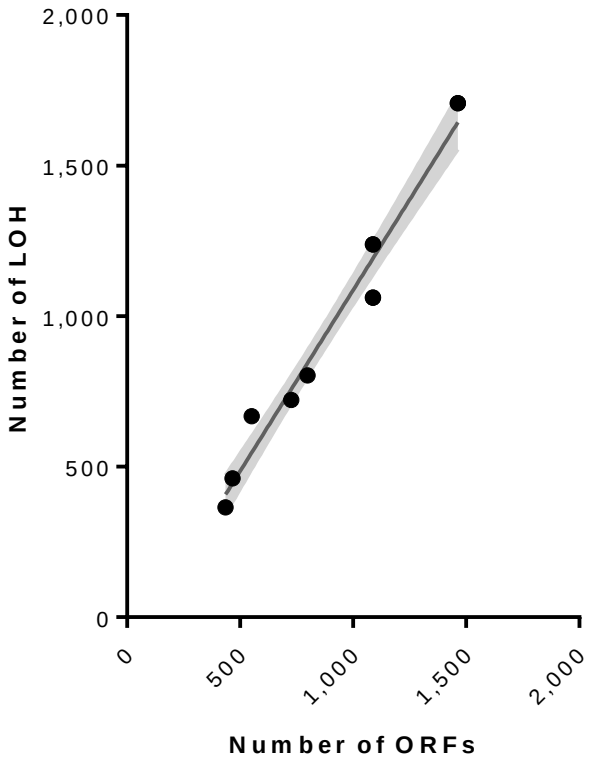

Figure S2: Correlation between the number of open reading frames (ORFs) per chromosome and the median number of LOH events per chromosome.
Plotting of the number of open reading frames (ORFs) (X axis) versus the median number of LOH events (Y axis) calculated per chromosomes. The number of ORFs per chromosome were recovered from CGD (Skrzypek *et al.* 2017). Gray line represents linear regression (+/- 95% CI, gray shade), R^2= 0.9685, Pearson's coefficient p<0.0001.

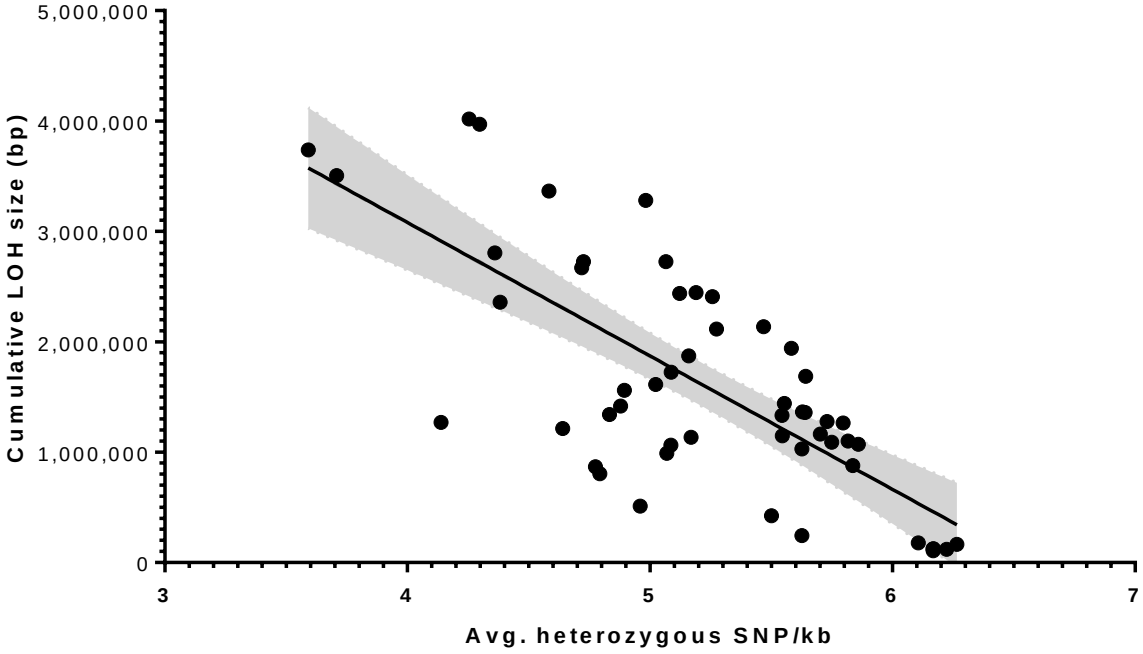

Figure S3: Correlation between the average number of heterozygous SNPs per kilobase and the cumulative LOH size across the 54 genome of C. albicans isolates.
Each circle represents a strain *C. albicans* strain (n=54), plotting of the number of average heterozygous SNP positions per kilobase (Avg. heterozygous SNP/kb) (X axis) versus the cumulative LOH size in base pairs (sum of all LOH for a given strain) (Y axis). The gray line represents linear regression (+/- 95% CI, gray shade), R^2= 0.5414, Pearson's coefficient p<0.0001.

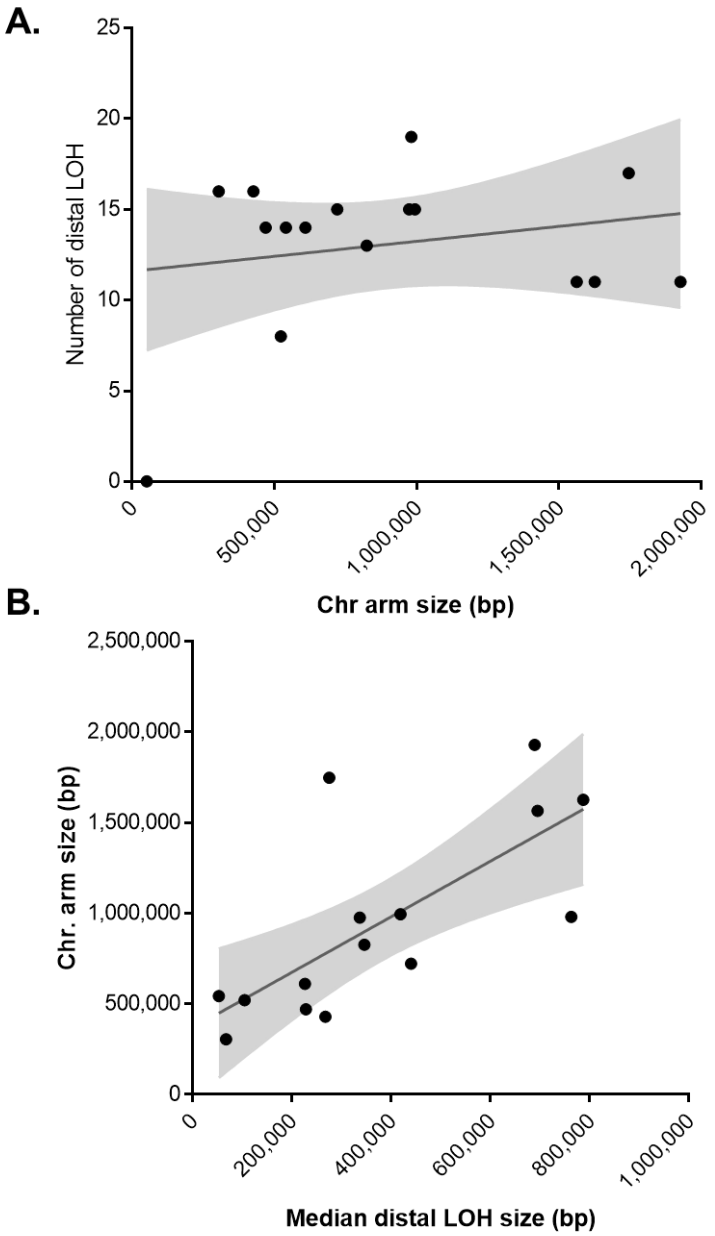

Figure S4: Correlations of distal LOH and chromosome arm parameters.
A. Plotting of the chromosome arm size (bp) versus the median number of distal LOH. The gray line represents linear regression (+/- 95% CI, gray shade), R^2= 0.04344, Pearson's coefficient p>0.05. **B.** Plotting of the chromosome arm size (bp) versus or the median distal LOH size (bp) per chromosome arm. The gray line represents linear regression (+/- 95% CI, gray shade), R^2= 0.5235, Pearson's coefficient p=0.0023.

Table S1: Strains analyzed from genetic cluster 1 of the *C. albicans* population.

Strain	Origin		Strain	Origin	
	Geographical	**Sampling**		**Geographical**	**Sampling**
CEC1424	USA	invasive	CEC4035	UK	superficial
CEC2020	Niger	superficial	CEC4106	Senegal	superficial
CEC3529	Morocco	superficial	CEC4107	Senegal	superficial
CEC3534	France	commensal	CEC4108	Senegal	superficial
CEC3544	Morocco	commensal	CEC4256	France	superficial
CEC3545	Belgium	commensal	CEC4257	France	superficial
CEC3560	Niger	superficial	CEC4259	France	invasive
CEC3599	France	superficial	CEC4479	unknown	food spoilage
CEC3603	Morocco	commensal	CEC4480	unknown	food spoilage
CEC3605	Morocco	commensal	CEC4488	unknown	food spoilage
CEC3609	Morocco	commensal	CEC4489	unknown	food spoilage
CEC3617	Morocco	commensal	CEC4496	unknown	food spoilage
CEC3621	Brazil	commensal	CEC4499	unknown	food spoilage
CEC3623	Morocco	commensal	CEC4500	unknown	food spoilage
CEC3625	Brazil	commensal	CEC4512	France	superficial
CEC3627	Brazil	commensal	CEC4523	France	superficial
CEC3636	Morocco	commensal	CEC4581	France	superficial
CEC3660	France	superficial	CEC4883	China	superficial
CEC3662	France	invasive	CEC4976	France	commensal
CEC3666	france	superficial	CEC5057	France	invasive
CEC3672	France	invasive	CEC5125	France	unknown
CEC3673	France	invasive	CEC5128	France	unknown
CEC3678	France	superficial	CEC5132	France	unknown
CEC3679	France	superficial	CEC5135	France	unknown
CEC3682	France	invasive	CEC5139	France	unknown
CEC4029	UK	superficial	CEC5145	France	unknown
CEC4032	UK	superficial	SC5314	USA	invasive

Table S2: Candidate hotspot genomic regions of LOH in cluster 1 of *C. albicans*.

	10 kb window	
Ca22chr1A_C_albicans_SC5314	1010001	1020001
Ca22chr1A_C_albicans_SC5314	2050001	2060001
Ca22chr2A_C_albicans_SC5314	1630001	1640001
Ca22chr3A_C_albicans_SC5314	100001	110001
Ca22chr3A_C_albicans_SC5314	320001	330001
Ca22chr3A_C_albicans_SC5314	490001	500001
Ca22chr3A_C_albicans_SC5314	1250001	1260001
Ca22chr3A_C_albicans_SC5314	1350001	1360001
Ca22chr4A_C_albicans_SC5314	570001	580001
Ca22chr4A_C_albicans_SC5314	1300001	1310001
Ca22chr5A_C_albicans_SC5314	1070001	1080001
Ca22chr6A_C_albicans_SC5314	100001	110001
Ca22chr6A_C_albicans_SC5314	220001	230001
Ca22chr6A_C_albicans_SC5314	510001	520001
Ca22chr6A_C_albicans_SC5314	620001	630001
Ca22chrRA_C_albicans_SC5314	50001	60001
Ca22chrRA_C_albicans_SC5314	120001	130001
Ca22chrRA_C_albicans_SC5314	610001	620001
Ca22chrRA_C_albicans_SC5314	1530001	1540001
Ca22chrRA_C_albicans_SC5314	1710001	1720001

Chapter 2: Study of the impact of recessive alleles and buffering role of repeated sequences in *Candida albicans*

Contains: Article published in mSphere (13/02/2019)

Identification of recessive lethal alleles in the diploid genome of a *C. albicans* laboratory strain unveils a potential role of repetitive sequences in buffering their deleterious impact

Timea MARTON[1,2], Adeline FERI[1,2,$], Pierre-Henri COMMERE[3], Corinne MAUFRAIS[1,4], Christophe D'ENFERT[1], Melanie LEGRAND[1,*]

[1] Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, 75015, France
[2] Univ. Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, 75015, France
[3] Institut Pasteur, Unité de Technologie et de Service Cytométrie et Biomarqueurs, Plate-forme Cytométrie, Paris, 75015, France
[4] Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), USR 3756 IP CNRS, Institut Pasteur, Paris, France
[$]Present address: Pathoquest, BioPark, 11 rue Watt, 75013 Paris, France

Abstract

The heterozygous diploid genome of *Candida albicans* is highly plastic, with frequent loss-of-heterozygosity (LOH) events. In the SC5314 laboratory strain, while LOH events are ubiquitous, a chromosome homozygosis bias is observed for certain chromosomes, whereby only one of the two homologs can occur in the homozygous state. This suggests the occurrence of recessive lethal allele(s) (RLA) preventing large-scale LOH events on these chromosomes from being stably maintained. To verify the presence of a RLA on Chr7, we utilized a system that allows (i) DNA double-strand break (DSB) induction on Chr7 by the I-*Sce*I endonuclease and (ii) detection of the resulting long-range homozygosis. I-*Sce*I successfully induced a DNA-DSB on both Chr7 homologs, generally repaired by gene conversion. Notably, cells homozygous for the right arm of Chr7B were unrecoverable confirming the presence of RLA(s) in this region. Genome data mining for RLA candidates identified a pre-mature nonsense-generating SNP, within the HapB allele of C7_03400c whose *S. cerevisiae* ortholog encodes for the essential Mtr4 RNA helicase. Complementation with a wild-type copy of *MTR4* rescued cells homozygous for the right arm of Chr7B, demonstrating that the $mtr4^{K880*}$ RLA is responsible for the Chr7 homozygosis bias in strain SC5314. Furthermore, we observed that the major repeat sequences (MRS) on Chr7 acted as hotspots for inter-homologue recombination. Such recombination events provide *C. albicans* with increased opportunities to survive DNA-DSBs whose repair can lead to homozygosis of recessive lethal or deleterious alleles. This might explain the maintenance of MRSs in this species.

Importance

Candida albicans is a major fungal pathogen, whose mode of reproduction is mainly clonal. Its genome is highly tolerant to rearrangements, in particular loss of heterozygosity events, known to unmask recessive lethal and deleterious alleles in heterozygous diploid organisms such as *C. albicans*. By co mbining a site-specific DSB inducing system and the mining of genome sequencing data of 182 *C. albicans* isolates, we were able to ascribe the chromosome 7 homozygosis bias of the *C. albicans* laboratory strain SC5314 to an heterozygous SNP introducing a premature STOP codon in the *MTR4* gene. We have also proposed genome-wide candidates for new recessive lethal alleles. We additionally observed that the major repeat sequences (MRS) on chromosome 7 acted as hotspots for inter-homologue recombination. Maintaining MRSs in *C. albicans* could favour haplotype exchange, of vital importance upon LOH events leading to homozygosis of recessive lethal or deleterious alleles which ineluctably accumulate upon clonality.

Introduction

In diploid genomes, new mutations are heterozygous and their effect is generally masked by the presence of the ancestral allele. As claimed in the Haldane's sieve only mutations that confer a fitness advantage as heterozygotes can invade the population. Although true, it does not make specific prediction about the fitness of the mutant homozygotes. Recent studies in *Saccharomyces cerevisiae* have observed maintenance of genetic variation due to heterozygote advantage, as a result of over-dominance of mutated alleles (Gerstein *et al.* 2014). In addition, Gerstein and colleagues used the model organism *S. cerevisiae* to show that recessive beneficial mutations can avoid Haldane's sieve in clonal organisms, through rapid loss-of-heterozygosity (LOH), and thus contribute to rapid evolutionary adaptation (Gerstein *et al.* 2014). Similarly, in *Candida albicans*, mutations followed by genomic rearrangements such as LOH events and isochromosome formation have been associated to the acquisition of antifungal resistance (Selmecki *et al.* 2006a; Coste *et al.* 2007), bringing forth the idea that mechanisms favoring genome plasticity could contribute to *C. albicans* fitness within the host and upon exposure to antifungals. *C. albicans* is a frequent human commensal yeast responsible for both mucosal and the majority of life-threatening nosocomial fungal infections (Prieto *et al.* 2016). Its diploid genome displays a high degree of plasticity that includes, in particular, LOH events. Despite frequent LOH, overall heterozygosity is maintained in the *C. albicans* population, as illustrated by various studies that highlighted the elevated levels of natural heterozygosity, with an heterozygous position every ~200-250 bp on average (Jones *et al.* 2004; Hirakawa *et al.* 2015; Ropars *et al.* 2018). Furthermore, several studies revealed that genome heterozygosity showed a significant correlation with faster growth rates (Hickman *et al.* 2013; Abbey *et al.* 2014; Hirakawa *et al.* 2015).

Essentially, mutations can be categorized as beneficial, harmful or neutral and can be differently assigned depending on the organism's environment. Because the mode of reproduction of *C. albicans* is mainly clonal, and therefore mimics inbreeding in higher eukaryotes, an increased number of recessive lethal alleles (RLA) in the *C. albicans* genome is expected as compared to other eukaryotes that undergo true sexual reproduction. Various types of mutations can impact the functionality of alleles and render them inactive; however mutations introducing premature STOP codons would convey the most obvious effect. Within the *C. albicans* laboratory strain SC5314, Muzzey *et al.* (Muzzey *et al.* 2013) reported almost 200 genes for which one of the alleles contains a SNP that introduces a premature STOP codon. Functional differences have already been reported between the two alleles of a heterozygous gene and, in all instances, the effect of the recessive mutation was visible only

upon homozygosis towards the mutated allele (*HIS4* (Gómez-Raja *et al.* 2008); *MBP1* (Ciudad *et al.* 2016); *GPI16/MRF2* (Feri *et al.* 2016)). Moreover, SNPs in promoter regions have been shown to alter expression regulation of two alleles (Muzzey *et al.* 2014). Of interest, LOH is pervasive in *C. albicans* isolates as homozygous regions can be found in all sequenced isolates and affect all of the chromosomes. These LOH events vary in size: they can be limited to a single chromosomal region, affect an entire chromosomal arm or even cover the entirety of a chromosome (Abbey *et al.* 2014; Hirakawa *et al.* 2015; Ropars *et al.* 2018).

Recently, a combination of molecular tools has been developed to study genome dynamics in *C. albicans*. First, an LOH reporter system takes advantage of fluorescent markers at an artificial heterozygous locus containing the BFP and GFP genes (Loll-Krippleber *et al.* 2015). Consequently, the appearance of spontaneous LOH events, for the given locus, can be monitored by the fluorescent status of cells using flow cytometry (Loll-Krippleber *et al.* 2015). Second, LOH events are often a result of DNA double strand breaks (DSB) (Malkova *et al.* 2000) resolved by means of various DNA repair mechanisms which can either be independent or dependent of homologous recombination. Feri *et al*. developed an inducible, locus-specific DNA-DSB system that uses the I-*Sce*I meganuclease from *S. cerevisiae.* When coupled to the BFP/GFP LOH reporter system, this system can be used to study the consequence of a targeted DNA-DSB on the appearance of LOH events. Indeed, Feri *et al*. have shown that I-*Sce*I-induced DNA breaks are predominantly repaired by gene conversion resulting in limited LOH. Nevertheless, various patterns of long-range LOH can also be obtained. Of note, the engineered system, alongside sequence resources, helped identify a RLA on Chr4B of *C. albicans* strain SC5314 (Feri *et al.* 2016). This RLA is the consequence of a nonsense mutation in the *GPI16* gene that encodes an essential component of the glycosylphosphatidylinositol (GPI) anchor biosynthetic machinery and explains why Chr4B is never observed in the homozygous state in *C. albicans* strains SC5314. Notably, this RLA appeared unique to strain SC5314 (Feri *et al.* 2016).

Although LOH can be observed on all 8 chromosome pairs, prior studies conducting haplotype characterization of (i) progenies from the parasexual life-cycle (Forche *et al.* 2008), (ii) homozygous diploid isolates derived from *RAD*52 double knock-out mutants (Andaluz *et al.* 2011), and (iii) haploid strains of *C. albicans* (Hickman *et al.* 2013) showed a chromosome homozygosis bias in the *C. albicans* laboratory strain SC5314. This suggests that mutations, potentially RLAs, could apply constraints on the directionality of LOH events. Indeed, the homozygosis state of some chromosomes was only observed for a given homolog, while recurrently absent for the other homolog. This is the case for chromosomes 1, 4, 6, as well as

chromosome 7 (Chr7) for which homozygosis of haplotype B (HapB) is never observed, while haplotype A homozygosis is, suggesting the presence of RLAs on Chr7 HapB (Chr7B).

In this study, we aimed to identify the RLA(s) on Chr7B using an approach similar to that developed by Feri *et al.* when searching for RLAs on Chr4B. This approach also allowed addressing the role that repetitive sequences, such as the Major Repeat Sequences (MRS), might play on the overall genome dynamics of *C. albicans*. MRS are unique to *C. albicans* and *C. dubliniensis,* and are found throughout their genomes. MRS are composed of 3 sub-regions: RB2 which contains the *FRG8* gene, the RPS region which varies in repeat numbers (thus in size) and the HOK region. *C. albicans* possess 8 MRS, one on each chromosome with the exceptions of Chr7 where the presence of one MRS on each arm is observed, and of Chr3 where an incomplete MRS is located (Chibana and Magee 2009). MRS expansion and contraction have previously been shown to be involved in chromosome loss where the chromosome copy containing the shorter MRS region is spontaneously lost (Lephart *et al.* 2005). Furthermore, MRS have also been shown to be involved in chromosome translocation (Chibana and Magee 2009), when two different chromosomes exchange large regions of an arm. Results presented below confirm that homozygosis of Chr7B is unrecoverable, and that this is the consequence of a premature STOP codon in the Chr7B-borne allele of the RNA helicase encoding gene *MTR4*. Furthermore, we highlight that repeat regions such as MRS are hotspots for inter-homolog recombination upon DNA repair and play a role in LOH dynamics in *C. albicans*.

Results

Strain engineering to promote and detect long-range LOH on Chr7.

Genome analysis revealed that the left arm of Chr7 carries only 9 heterozygous SNPs in 3 ORFs, while the right arm of Chr7 carries 784 heterozygous SNPs in 105 ORFs. Because RLAs are more likely to be found in heterozygous regions, we focused on the right arm of Chr7 to understand Chr7 homozygosis bias. To efficiently screen for the presence of RLAs on Chr7 right arm, we engineered strains carrying an artificial heterozygous BFP/GFP LOH reporter system (Loll-Krippleber *et al.* 2015) associated with an I-*Sce*I DNA-DSB-inducing system (Feri *et al.* 2016) (Figure 1A). Because we chose to insert the I-*Sce*I target site (TS) associated with the auxotophic marker *URA3* conferring uridine prototrophy in the most *mrs*-7b proximal, gene-free region found on the right arm of Chr7, this setup allows rendering a maximum number of alleles homozygous on this arm while avoiding the *mrs*-7b. The integration of the *URA3*-I-*Sce*I TS cassette can occur on either Chr7 homologs (Chr7A or Chr7B) thus transformants were screened by SNP-RFLP to identify which Chr7 haplotype was

targeted (Figure S1). Using the heterozygous SNP at position 727,328, we showed that 28/51 *C. albicans* transformants had integrated the I-*Sce*I TS on Chr7A (55%) and 23/51 on Chr7B (45%), demonstrating the absence of integration bias for this locus. Two independent transformants that had integrated the I-*Sce*I TS on Chr7 HapA (CEC5061) or HapB (CEC5062), were selected and used in subsequent analysis (Figure 1A).

Strains CEC5061 and CEC5062 underwent preliminary characterization regarding their fluorescence status, as well as their growth rate to ensure that the successive transformation steps did not significantly alter their fitness. The fluorescence status of the intermediate and final strains was verified by flow cytometry. The flow cytometry outputs clearly displayed (i) the absence of fluorescence signals in the parental strain CEC4591, (ii) a shift towards the mono-GFP gate upon integration of the P_{TDH3}-*GFP-ARG4* cassette in CEC4679, (iii) a shift towards the double-fluorescent BFP/GFP gate upon integration of the P_{TDH3}-*BFP-HIS1* cassette in CEC4685 and (iv) the double-fluorescence status of the population upon integration of the *URA3*-I-*Sce*I-TS cassette in CEC5061 and CEC5062 (Figure S2A).

Growth curves performed demonstrated that only the insertion of the *URA3*-I-*Sce*I TS cassette in CEC5061 and CEC5062 resulted in a faster growth rate, almost certainly due to the uridine prototrophy in these strains as *URA3* deletion has been shown to result in significant decreases in *C. albicans* growth rate even when *ura3*Δ strains are grown in rich medium (Kirsch and Whitney 1991)(Figure S2B).

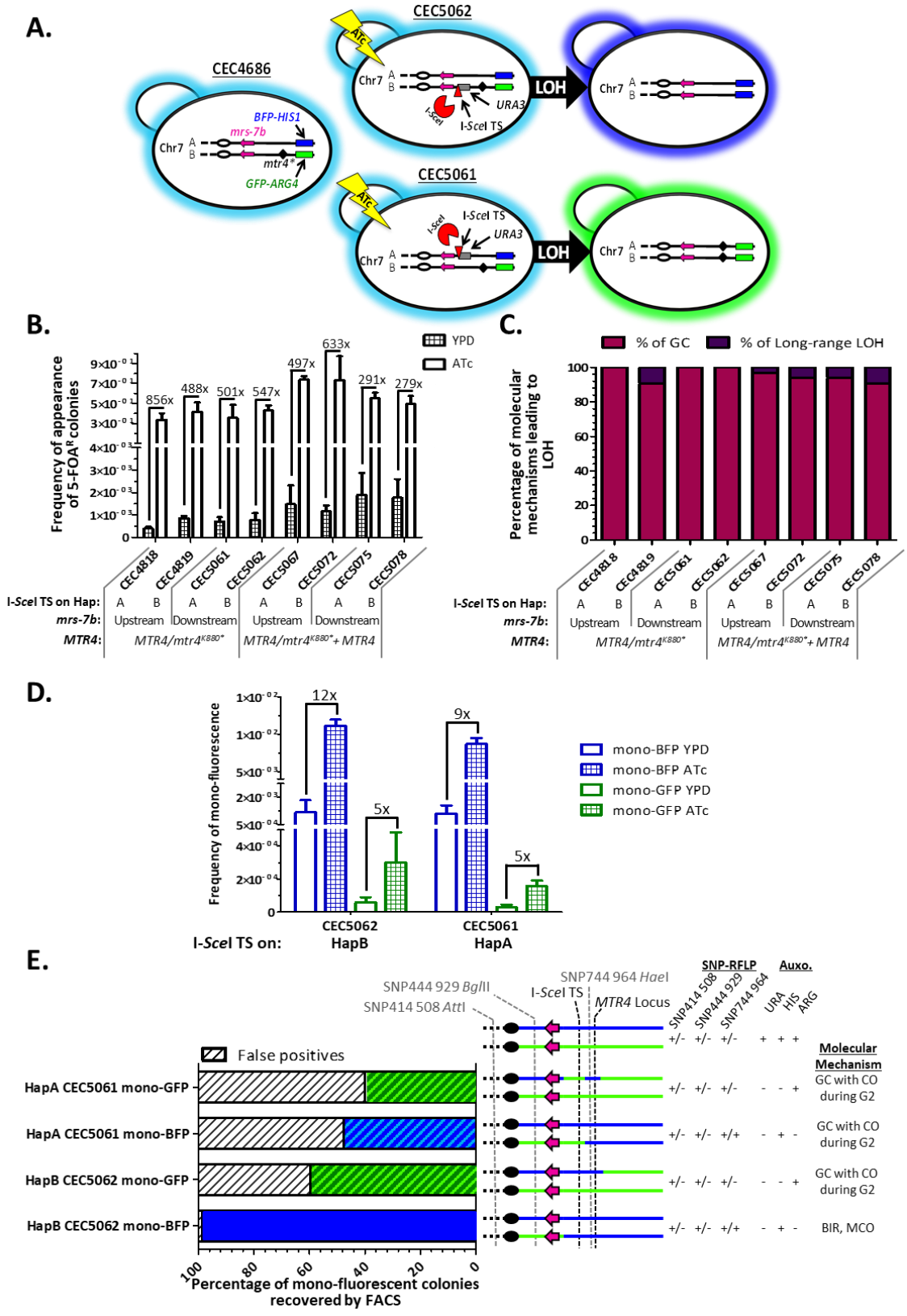

Figure 1: Coupling of a DNA double-strand-break-inducing system and a FACS-optimized LOH reporter system on Chr7.

A. Illustration of strains bearing the BFP/GFP LOH reporter system on the right arm of Chr7 associated to an I-*Sce*I-inducible DNA-DSB system on Chr7B (CEC5062) or Chr7A (CEC5061). Upon addition of ATc, the mega-endonuclease I-*Sce*I is expressed and will generate a DNA-DSB at its target sequence (I-*Sce*I TS linked to the *URA3* marker) on Chr7. Theoretically, when the DNA-DSB is repaired by BIR

or MCO, doubly-fluorescent cells harboring the I-*Sce*I-TS on BFP-bearing Chr7A will become mono-GFP while doubly-fluorescent cells harboring the I-*Sce*I-TS on GFP-bearing Chr7B will become mono-BFP. The centromere, MRS and candidate RLA on Chr7 are shown with a circle, a pink arrow and a diamond, respectively. **B.** Average frequency of 5-FOA^R colonies obtained after 8h of I-*Sce*I overexpression and recovery from 3 independent experiments (n=3, ± S.D.) alongside fold changes observed between YPD and ATc conditions. **C.** Percentage of molecular mechanisms leading to LOH amongst 5-FOA^R colonies in induced condition (n=32) **D.** Histogram representing average frequency (n=6, ± S.D.) of appearance of mono-fluorescent cells in the presence (ATc) and absence (YPD) of expression of I-*Sce*I and hence induced DNA-DSBs on Chr7B (CEC5062) or Chr7A (CEC5061). Fold changes between uninduced and induced conditions for each mono-fluorescent populations are indicated. **E.** Characterization of mono-fluorescent sorted individuals (n=32) from induced CEC5062 (I-*Sce*I TS on HapB) and CEC5061 (I-*Sce*I TS on HapA) strains. A subset of sorted individuals was characterized for their fluorescence and auxotrophy statuses in addition to SNP-RFLP for haplotype appointment, in order to profile homozygosis of the right arm of Chr7 and determine the molecular mechanism used for I-*Sce*I-induced DNA-DSB repair. In the stacked histogram, white and black striped portions represent miss-sorted cells (false positives), while green or blue portions represent the proportion of properly FACS sorted mono-fluorescent individuals. The presence of stripes on the green or blue bars indicate that those individuals are not fully homozygous for one given haplotype from the I-*Sce*I TS to the BFP/GFP LOH reporter system on the right arm of Chr7. Abbreviations of molecular mechanisms are as follows; gene conversion (GC), break-induced replication (BIR), mitotic crossover (MCO), gene conversion with crossover (GC with CO).

Validation of the I-*Sce*I DNA-DSB induction system on Chr7 by 5-FOA counter-selection.

As the I-*Sce*I TS is associated with the genetic auxotrophic marker *URA3*, we could assess the frequency of cells that have survived an I-*Sce*I-induced DNA-DSB at the TS by monitoring the frequency of appearance of 5-FOA resistant (5-FOA^R) colonies upon I-*Sce*I induction. Indeed, 5-FOA^R colonies should have lost the *URA3* genetic marker (uridine auxotrophy) and are likely to represent cells that have sustained an I-*Sce*I-dependent DNA-DSB through a LOH event, even though point mutations in the *URA3* gene cannot be excluded. Upon I-*Sce*I induction, we obtained 501 times more 5-FOA^R colonies when the I-*Sce*I TS was located on Chr7A and 547 times more 5-FOA^R colonies when the I-*Sce*I TS was located on Chr7B (Figure 1B). These data confirmed the efficiency of the I-*Sce*I-dependent DNA-DSB induction system on Chr7. The majority of I-*Sce*I DNA-DSB induced 5-FOA^R colonies (90-100%) resulted from DNA-DSB repair by gene conversion, as suggested by fluorescence and auxotrophy profiles of 32 ura$^-$ colonies (Figure 1C). Similar to what has been observed for Chr4 (Feri *et al.* 2016), DNA-DSBs by I-*Sce*I on Chr7 are predominantly repaired by gene conversion repair mechanisms, resulting in short-range LOH events.

An I-SceI-induced DNA-DSB on Chr7B leads to viable cells homozygous for the right arm of Chr7A.

Although the 5-FOA assays inform on the overall occurrence of LOH events encompassing the *URA3* gene, it does not allow to efficiently study the under-represented long-range LOH events. Thus, we also investigated LOH frequency upon I-*Sce*I expression

using flow cytometry, an assay that specifically detects long-range LOH events. As expected, upon induction of I-*Sce*I in CEC5062, possessing the I-*Sce*I TS on the GFP-bearing Chr7B, we observed a 12-fold increase in the appearance of mono-BFP cells (Figure 1D). We also observed a 5-fold increase in the appearance mono-GFP cells (Figure 1D). A subset of each population was recovered by fluorescence activated cell sorting (FACS) and further characterized. We observed that, while the majority of the mono-BFP population included true mono-BFP cells displaying complete Chr7A homozygosis distal to the I-*Sce*I TS, 100% of the rare true mono-GFP cell displayed only partial homozygosis of Chr7B. From a mechanistic point of view, the mono-BFP cells resulted most likely from the repair of the DNA-DSBs by mechanisms of break-induced replication or mitotic crossover while the mono-GFP cells could be one of the possible outcomes of DNA-DSB repair by gene conversion with crossover during the G2 phase of the cell cycle (Figure 1E).

Absence of recovery of cells being fully homozygous for the right arm of Chr7B.

Unlike targeting the right arm of Chr7B, a DNA-DSB on the right arm of Chr7A in CEC5061 should lead to a higher increase in frequency of the mono-GFP cells in comparison to the mono-BFP cells (Figure 1D). Although an augmentation in frequency of both mono-BFP and mono-GFP cells was obtained, the mono-BFP cells still appeared at a higher frequency, 8×10^{-3} ($\pm$ 8×10^{-4}), as compared to 1×10^{-4} ($\pm$ 3×10^{-5}) for the mono-GFP cells in the induced condition (Figure 1D). Characterization of a subset of FACS-sorted mono-BFP cells confirmed that the true mono-BFP cells had arisen from a DNA-DSB repaired by gene conversion with crossover during G2. Further characterization of FACS-sorted mono-GFP cells (corresponding to the expected fluorescence) revealed that the I-*Sce*I-induced homozygosis of Chr7B was only partial in the targeted region. Thus, rather than having experienced break-induced replication or mitotic crossover events extending from the I-*Sce*I site to the BFP/GFP locus, the rare mono-GFP individuals were likely to have resulted from DNA-DSB repair by gene conversion with crossover during G2 (Figure 1E). In conclusion, we were unable to recover a single individual that had undergone complete homozygosis of the right arm of Chr7B distal to the I-*Sce*I TS. This suggests that complete homozygosis of the right arm of Chr7B is associated with lethality in the SC5314 genetic background; thus confirming the chromosome homozygosis bias previously observed and validating the hypothesis that the presence of RLA(s) in this region could be at cause.

A data mining strategy identifies a heterozygous mutation in the *MTR*4 gene as a possible cause of the homozygosis bias of Chr7.

Genome sequence data obtained from a collection of 182 *C. albicans* isolates (Ropars *et al.* 2018) including the reference strain SC5314, was used to compile all heterozygous SNPs within ORFs of Chr7 and search for SNPs (i) generating a premature STOP codon, (ii) showing a heterozygous genotype in SC5314, (iii) never observed in the homozygous state in the collection of 182 genomes and (iv) located in a coding region never found to be dispensable in *C. albicans*. Only one such SNP was identified, located at position 746,359 on Chr7B. In *C. albicans* strain SC5314, this SNP causes a change from AAA (lysine) on Chr7A to TAA (STOP) on Chr7B in the C7_03400C gene. This gene is the orthologue of *S. cerevisiae MTR4* that encodes an essential ATP-dependant RNA helicase involved in RNA processing in *S. cerevisiae* (Giaever *et al.* 2002; Bernstein *et al.* 2006; Skrzypek *et al.* 2017). In *C. albicans* SC5314, the HapA allele of *MTR4* encodes a full length Mtr4 protein while the HapB allele carrying the STOP-introducing SNP encodes a truncated Mtr4^{K880*} protein that misses a C-terminal DSHCT domain, common to DEAD-box helicases (Figure 2A).

The *mtr4*^{K880*} allele is responsible for the Chr7 homozygosis bias.

To validate that the identified RLA candidate *mtr4*^{K880*} is truly responsible for the Chr7 homozygosis bias, the full-length *MTR4* ORF was placed ectopically under the control of the P_{TDH3} constitutive promoter, in strain CEC5061, giving rise to strain CEC5075 (Figure 2B). Results shown in Figure 1B-C confirmed that the DNA-DSB inducing system was functional in this strain and that the I-*Sce*I-induced DNA-DSB was predominantly repaired by gene conversion as observed in other instances. Strikingly, flow cytometry analysis revealed a significant elevation of the fold-increase of mono-GFP cells obtained upon induction of I-*Sce*I in CEC5075 by comparison to CEC5061 (Figure 2C). Upon cell sorting and characterization of these mono-GFP cells, the majority (93%) appeared as true mono-GFP individuals that had become fully homozygous for Chr7B from the I-*Sce*I TS to the BFP/GFP LOH reporter system. These cells were likely to have resulted from the repair of the induced DNA-DSB by break-induced replication or mitotic crossover, indicating that complete homozygosis of the right arm of Chr7B is compatible with viability upon addition of a functional *MTR4* allele (Figure 2D). Similar to what was observed with the CEC5061 parental strain, I-*Sce*I induction in the CEC5075 *MTR4*-complemented strain resulted in an augmentation of mono-BFP cells (Figure 2C). Eighty-eight percent of these mono-BFP cells corresponded to cells where the DNA-DSB had been repaired by gene conversion with crossover during G2 (Figure 2D).

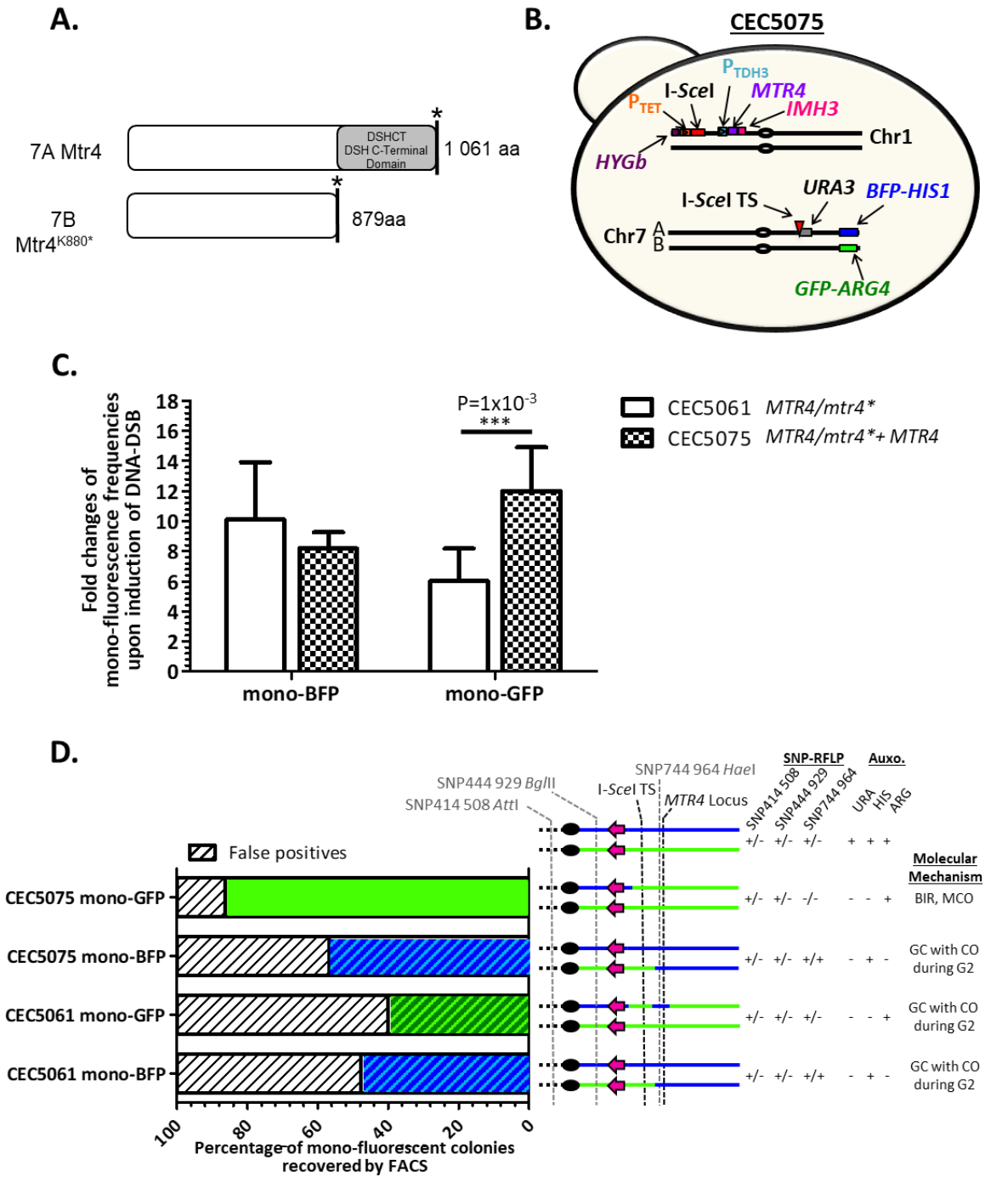

Figure 2: Complementation with a functional MTR4 allele restores viability of Chr7B homozygous cells.

A. Schematic alignment of the wild-type *MTR4* and *mtr4*^{K880*}-encoded proteins. *MTR4* encodes an ATP-dependent RNA-helicase and the nonsense mutation in the *mtr4*^{K880*} allele leads to loss of the DSH C-terminal domain shared by DEAD box helicases. **B.** Illustration of the CEC5075 complemented strain possessing the functional *MTR4* allele at the *RPS1* locus on Chr1. Centromeres are illustrated with a circle. **C.** Comparison of the average fold-increase of the mono-fluorescent populations upon DNA-DSB induction (ATc/YPD) in both parental (CEC5061) and complemented (CEC5075) strains (n=6, ±S.D.), significance was determined using a bilateral T-test (P-value). **D.** Characterization of mono-fluorescent sorted individuals (n=32) from induced CEC5061 (I-*Sce*I TS on HapA) and *MTR4*-complemented CEC5075 strains. See legend of Figure1 for explanation.

Upon DNA-DSB, Major Repeat Sequences are a source of inter-homolog recombination.

Our initial strategy for unveiling RLAs was to induce a DNA-DSB downstream of *mrs-7b* (the MRS located on the right arm of Chr7 – Figure 3A) in order to ensure that the repeat sequences would not interfere with the DNA repair mechanisms. To seek validation of this initial assumption, the I-*Sce*I TS was moved upstream of *mrs-7b*, between the centromere and *mrs-7b*, on either HapA or HapB and in presence of an ectopic copy of *MTR4* (Figure 3B). By plating on 5-FOA medium, we showed that the inducible I-*Sce*I DNA-DSB system was functional in the new location and that gene conversion was the preferred mechanism of repair, independently of the targeted haplotype (Figure 1B-C). Notably, when the I-*Sce*I TS was localized on HapB upstream of *mrs-7b* in a *MTR4*-complemented strain (CEC5072), induction of I-*Sce*I expression resulted in an almost equal increase in the mono-GFP and mono-BFP cell populations relative to non-induced conditions (Figure 3C). This contrasted to what was observed when the I-*Sce*I TS was localized on HapB downstream of *mrs-7b* in a *MTR4*-complemented strain (CEC5078), whereby a greater increase in the mono-BFP cell population than mono-GFP cell population was observed upon I-*Sce*I induction (Figure 3C). This observation could be explained by inter-homolog recombination at *mrs-7b*, which would link the I-*Sce*I TS to either the BFP or GFP markers when the TS is located upstream of *mrs-7b* (Figure 3B). This would result in an equal proportion of mono-BFP and mono-GFP cells upon repair of the I-*Sce*I-induced DNA-DSB, keeping in mind that break-induced replication or mitotic crossover are the predominant molecular mechanisms leading to long range LOH (Figure 3C). In contrast, when the I-*Sce*I TS is located downstream *mrs-7b*, it remains linked to the GFP marker, regardless of inter-homolog recombination at *mrs-7b*, thus predominantly yielding mono-BFP cells upon I-*Sce*I-induced DNA-DSB repair (Figure 3B-C). Overall, our data suggest that the MRS is a hotspot for inter-homolog recombination upon DNA repair on the right arm of Chr7.

Of interest, *C. albicans* Chr7 is characterized by the occurrence of a second MRS on the left arm, namely *mrs-7a* (Figure 3A). Strains possessing the I-*Sce*I TS between the centromere and *mrs-7a* and the BFP/GFP LOH reporter system on the left arm of Chr7 also exhibited an equal rate in mono-BFP and mono-GFP cells upon I-*Sce*I induction, whatever the location of the I-*Sce*I TS on HapA or HapB (data not shown), suggesting an increase in the number of cells that have undergone crossover at *mrs-7a*, linking the I-*Sce*I TS to either the BFP or GFP markers. Thus, *mrs-7a* also appears as a hotspot for inter-homolog recombination on Chr7.

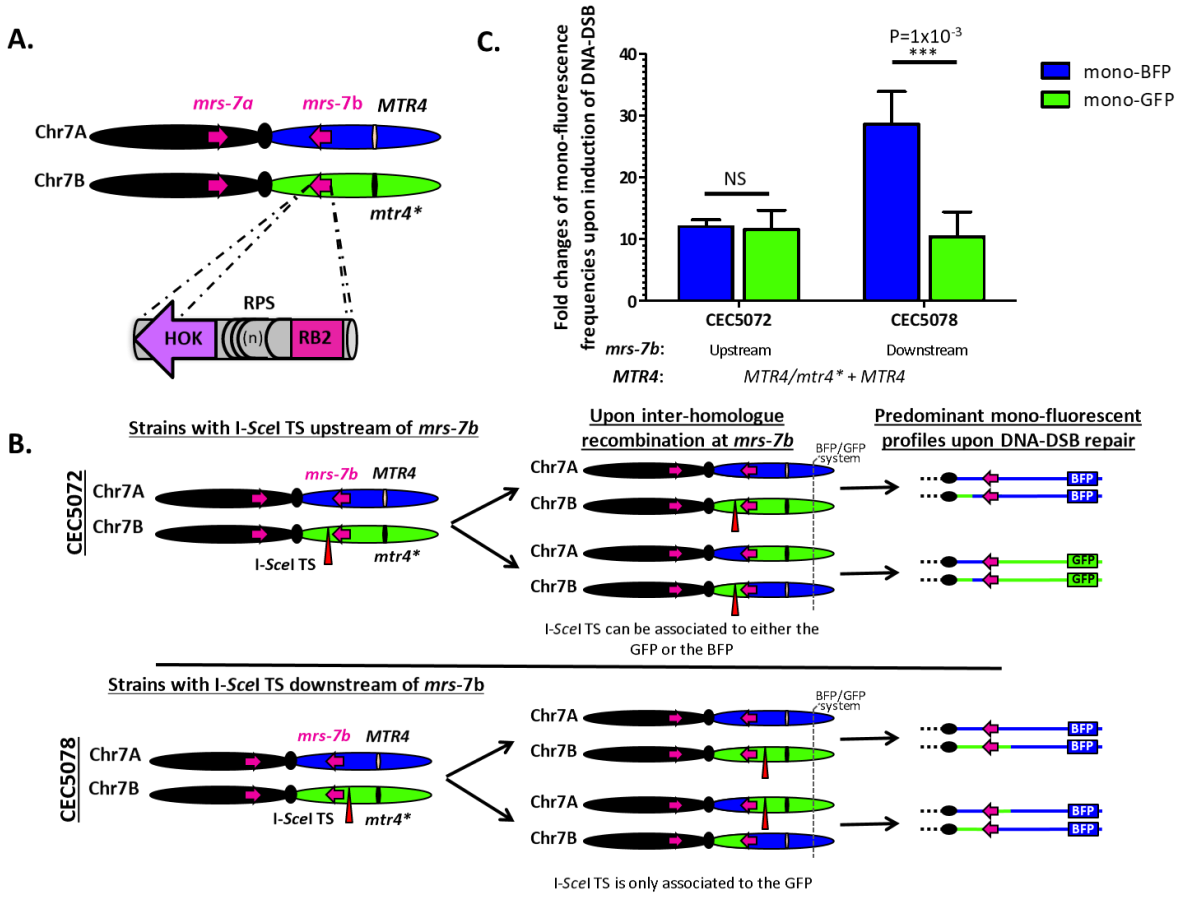

Figure 3: Upon DNA-DSB, Major Repeat Sequences are a source of inter-homolog recombination.
A. Three sub-units compose the MRS, which are hotspots for mitotic crossover. **B.** Engineering of strains where the I-*Sce*I TS is located on HapB upstream (CEC5072) or downstream (CEC5078) of *mrs*-7b. Illustration of the right arm of Chr7 upon inter-homologue recombination at *mrs-7b* followed by the predominant mono-fluorescence profiles upon DNA-DSB repair leading to long-range LOH events. **C.** Histogram showing average fold changes (n=6, ± S.D.) of mono-fluorescent populations upon I-*Sce*I induction (ATc vs YPD). Whereas no significant difference in augmentation of both populations when the I-*Sce*I TS is placed upstream of *mrs*-7b (CEC5072), a significant difference is observed when the I-*Sce*I TS is placed downstream of *mrs*-7b (CEC5078). Significance was determined using a bilateral T-test (P-value).

Discussion

Previous studies have shown that *C. albicans* strains harbor recessive lethal alleles (RLAs) that are responsible for a bias upon homozygosis of certain chromosomes, whereby only one of the two homologs can be retained in the homozygous state (Table 1). Yet, the nature of these RLAs is generally unknown. In this report, we have ascribed the Chr7 homozygosis bias of the *C. albicans* laboratory strain SC5314 to a heterozygous SNP introducing a premature STOP codon in the *MTR4* gene. Furthermore, we have unveiled the contribution of the Major Repeat Sequence (MRS) to inter-homologue recombination and hence, chromosome dynamics in *C. albicans.*

Our work focused on the homozygous bias observed for Chr7, suggesting the presence of at least one RLA on Chr7B (Table 1). Using a fluorescence-based LOH reporter system and an I-*Sce*I-dependent DNA-DSB inducing system, we have shown that while long-range homozygosis of Chr7A does not affect cell viability, long-range homozygosis of Chr7B is nonviable in *C. albicans* strain SC5314. A library of SNPs compiled from 182 clinical *C. albicans* genomes (Ropars *et al.* 2018) was searched for SNPs that would result in the most drastic outcome, i.e. a premature STOP codon. In this respect, we surely underestimated the presence of recessive lethal alleles as mutations other than premature STOP-introducing SNPs are likely to result in lethal phenotypes. Indeed, non-synonymous SNPs in coding regions have been shown to negatively impact protein function or regulation (De Gobbi *et al.* 2006; Nackley *et al.* 2006). Our approach pinpointed a SNP in the *C. albicans MTR4* gene that complementation experiments confirmed to be the RLA responsible for the lethality of individuals homozygous for Chr7B in strain SC5314. The identified SNP results in a truncated form of Mtr4 that lacks the last 182 C-terminal amino acids, encompassing a DEAD-box family DSHCT domain (Mtr4^{K880*}, Figure 2A). In *S. cerevisiae MTR4* encodes an ATP-dependent RNA-helicase whose deletion results in nuclear accumulation of unprocessed RNAs (Liang *et al.* 1996) and reduction of function results in increased sensitivity to both benomyl and nocodazole (Smith *et al.* 2011). Importantly, it has been shown that the C-terminal domain of other fungal RNA helicases is critical for their proper RNA-unwinding function (Mohr *et al.* 2008), consistent with the *mtr4^{K880*}* being a loss-of-function allele and homozygosity of this allele being lethal.

Table 1: Summary of chromosome homozygosis bias observed in the literature.

Chr	**Parasexuality** (Forche *et al.* 2008)	***RAD*52 mutants** (Ciudad *et al.* 2004)	**Obligate diploid** (Hickman *et al.* 2013)	**Bias summary**
R	no BB	no BB		None
1	no AA no BB	no BB	no BB	BB bias
2	no BB			None
3	no AA	no AA	no BB	None
4	no AA no BB	no BB	no BB	BB bias
5		no AA		None
6	no BB	no BB	no BB	BB bias
7	no BB	no BB	no BB	BB bias

Our study and that of Feri *et al.* indicate that the mining of the genomes of a large panel of *C. albicans* isolates for premature STOP-introducing SNPs is a suitable approach to identify RLAs responsible for chromosome homozygosis bias in *C. albicans* strains. As a reference, we provide in Table S4 a list of 70 genes that harbor a STOP-introducing SNP in one of the 2 alleles in the laboratory strain SC5314, including 13 alleles (Table 2) that were never found in the homozygous state in a collection of 182 genome-sequenced isolates,

representative of the *C. albicans* population (Ropars *et al.* 2018). The location of these 13 candidate RLAs across the *C. albicans* genome is shown in Figure 4. Subsequent Sanger sequencing will be necessary to confirm these alleles. Our reduced number of genes with SNPs introducing a premature STOP codon in SC5314, as compared to Muzzey *et al.*, could be explained by the stringency of our analysis. Indeed, only positions with high SNP quality and coverage depth > 20 for all 182 strains of our collection were considered for further analysis.

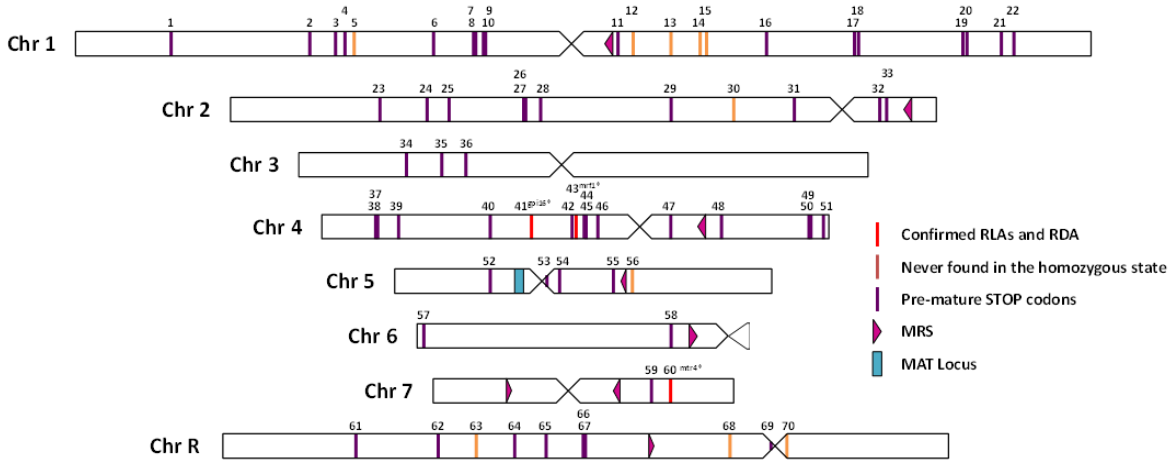

Figure 4: Schematic representation of the localization of premature STOP-inducing SNPs in the C. albicans SC5314 genome.
Representation of 70 heterozygous SNPs (vertical bars) inducing a premature STOP codon distributed on all 8 chromosomes of the reference strain SC5314. SNPs inducing premature STOPs that are never found in the homozygous state in the library of 182 genomes of clinical *C. albicans* isolates are represented in orange while confirmed recessive lethal and deleterious alleles are identified in red. Details regarding each SNP can be found in Table S4 using the assigned numbering in this figure.

A promising candidate to explain the homozygosis bias observed for Chr1 is the *gcd6** allele (C1_08600C_B) that carries a SNP introducing a premature STOP codon, truncating the protein from the 17 C-terminal amino-acids. In *S. cerevisiae*, *GCD6* encodes the catalytic epsilon subunit of the translation initiation factor eIF2B. The truncated domain is found at the C-terminus of several translation Initiation Factors and is important for mediating protein-protein interactions. *GCD6* is essential in *S. cerevisiae,* as null mutants are inviable. The same is likely to be true in *C. albicans* since heterozygous mutants are viable but null mutants have never been reported in the literature. Therefore, it can be hypothesized that homozygosis of *gcd6**, located on Chr1B, could result in cell death and be responsible for the homozygosis bias observed on Chr1 (Table 1).

Interestingly, a recent transposon mutagenesis screen in haploid *C. albicans* also argues that Ca*GCD6* is essential (Segal *et al.* 2018). Among the 13 candidate RLAs presented in Table 2, the same screen argues for essentiality of Ca*GPI16* (previously identified (Feri *et*

al. 2016)) and Ca*MTR4* (identified in this study). Although the 10 remaining genes in Table 2 were defined as nonessential genes by Segal *et al*., it is puzzling to see that the mutated alleles with a premature STOP codon were never observed in the homozygous state in the natural population of 182 *C. albicans* isolates. This observation highlights the complementarity of the study by Segal *et al.* and our work, given that gene essentiality can vary depending on growth conditions and *in vitro* assessment of gene essentiality does not necessarily correlate with *in vivo* essentiality (Segal *et al.* 2018).

Table 2: Genome-wide RLA candidates in the C. albicans strain SC5314. *

Chr	Pos.	ORF name	Gene name	Description	Hap. mut. allele	Essentiality in	
						C. albicans	*S. cerevisiae*
Ca22chr1A	1131425	C1_05380C_A		Ortholog(s) have telomeric DNA binding activity, role in DNA double-strand break processing, DNA replication initiation, chromatin silencing at silent mating-type cassette, telomere capping and shelterin complex localization	B	no info	no
Ca22chr1A	1760072	C1_08070W_A	CDR4	Putative ABC transporter superfamily; fluconazole, Sfu1, Hog1, core stress response induced; caspofungin repressed; fluconazole resistance not affected by mutation or correlated with expression; rat catheter and flow model biofilm induced	B	no	no
Ca22chr1A	1882812	C1_08600C_A	GCD6	Ortholog of S. cerevisiae Gcd6; catalytic epsilon subunit of the translation initiation factor eIF2B; genes encoding translation factors are repressed by phagocytosis by murine macrophages	B	het ok	yes
Ca22chr1A	1986384	C1_09100W_A	ZCF28	Zn(II)2Cys6 transcription factor; required for yeast cell adherence to silicone substrate; Spider biofilm induced	A	no	no
Ca22chr1B	634346	C1_03040W_B		Protein of unknown function; induced by nitric oxide; rat catheter and Spider biofilm induced	A	no info	no
Ca22chr1B	834720	C1_03960C_B		Ortholog(s) have role in TOR signaling, re-entry into mitotic cell cycle after pheromone arrest and endoplasmic reticulum membrane, endoplasmic reticulum-Golgi intermediate compartment localization	A	no	no
Ca22chr4A	659155	C4_03130W_A		Putative GPI transamidase component; possibly an essential gene, disruptants not obtained by UAU1 method	B	het ok (no UAU mutant)	yes
Ca22chr4A	796679	C4_03750C_A		Ortholog(s) have translation release factor activity, role in mitochondrial translational termination and mitochondrial inner membrane localization	B	no info	no
Ca22chr4B	531082	C4_02560C_B	UME7	Putative transcription factor with zinc cluster DNA-binding motif; similar to S. cerevisiae Ume6p, which is a transcription factor involved in the regulation of meiotic genes	No allelic variation in feature	no	no
Ca22chr7A	746359	C7_03400C_A		Ortholog of S. cerevisiae Mtr4, an ATP-dependent 3'-5' RNA helicase of the DEAD-box family; Hap43-induced gene; Spider biofilm induced	B	het ok	yes
Ca22chrRA	1603863	CR_07380C_A	KSR1	3-ketosphinganine reductase, catalyzes the second step in phytosphingosine synthesis	B	no	yes
Ca22chrRA	1782353	CR_08200C_A		Putative MRP/CFTR-subfamily ABC transporter; member of multidrug resistance-associated protein (MRP) subfamily of ABC family; similar to S. cerevisiae Bpt1p	B	het ok	no

*Confirmed RLAs in red and recessive deleterious allele in yellow

From a mechanistic point of view, our work revealed that most of the cells that have undergone a I-*Sce*I-mediated DNA-DSB on Chr7 use gene conversion as a repair mechanism, thus limiting LOH extent and the loss of genetic information. Break-induced replication, mitotic crossover, gene conversion with crossover and chromosome truncation, that lead to long-

range LOH events, in decreasing order are less frequently utilized. Our results are consistent with those of Feri *et al* on Chr4, arguing that our observation is not locus specific but can be applied genome-wide in *C. albicans*. However, we cannot exclude the possibility that the relative usage of these repair mechanisms is specific to I-*Sce*I-induced DNA-DSB, *i.e.* DNA-DSBs induced by the CRISPR-*Cas9* RNA-guided endonuclease could be preferentially handled by another mechanism of DNA repair.

In the course of this study, we also addressed the role that repetitive sequences, such as the MRS, might play on the overall genome dynamics of *C. albicans*. This was achieved by studying how MRS position affects the outcome of I-*Sce*I-induced LOH. Even though most *C. albicans* chromosomes possess a unique MRS region, the presence of two MRS regions, one on each arm, is unique to Chr7 (Chibana and Magee 2009). We observed that the presence of *mrs*-7b or *mrs*-7a between the I-*Sce*I TS and the telomere-proximal LOH reporter system on Chr7 results in equal augmentation of the mono-BFP and mono-GFP populations upon I-*Sce*I-dependent DNA-DSB (Figure 3). This contrasts to what is observed when the I-*Sce*I TS is located downstream the MRS, whereby the mono-fluorescent population arising by break-induced replication or mitotic crossover is increased relative to the mono-fluorescent population arising by gene conversion with crossover in the G2 phase of the cell cycle (a scarce molecular mechanism), unless a RLA is present. This suggests that the MRS could be a hotspot for inter-homolog mitotic crossover on Chr7. Indeed, upon recombination events at the MRS, DSBs repaired by break-induced replication or mitotic crossover would result in a relatively equal appearance of both mono-fluorescent populations when I-*Sce*I-TS is located between the centromere and the MRS (Figure 3B-C).

Mitotic crossovers at MRSs could be an intrinsic feature of these repeat regions or be triggered by either (i) stress resulting from I-*Sce*I overexpression or (ii) the physical I-*Sce*I-induced DNA-DSB. The former would imply that increased mitotic crossovers at MRS should be expected on all chromosomes, remaining to be tested. The latter would imply that increased mitotic crossovers at Chr7 MRSs are only observed upon repair of a DNA-DSB upstream of these repeated regions. These latter hypotheses are consistent with a role for stress in the enhancement of the recombination frequency near the MRS or in general, as already suggested by Lephart *et al*. (Lephart *et al.* 2005). Importantly, our results confirm the original proposal of Pujol *et al.* (Pujol *et al.* 1999) that MRSs on Chr7 are hotspots for recombination. Concretely, MRS would allow switching *C. albicans* haplotypes by generating a new combination of alleles. Information regarding the biological importance of MRS remains scarce despite the positive selection on MRS leading to the maintenance of these large and unique repeats in the *C. albicans* genome. Such recombination events provide *C. albicans* with

increased opportunities to survive DNA-DSBs whose repair can lead to homozygosis of recessive lethal or deleterious alleles. This might explain the maintenance of MRSs in this species.

Material and methods

Strains and media

C. albicans strains used in this work are derived from the reference strain SC5314. The cloning experiments were carried out using One Shot® TOP10 chemically competent *Escherichia coli* cells (ThermoFisher Scientific). All *C. albicans* strains and *E. coli* plasmids generated and used throughout this investigation are listed in Table S1 and Table S2, respectively.

C. albicans yeasts were maintained on rich yeast extract-peptone-dextrose (YPD) medium (1% yeast extract, 2% peptone, 2% dextrose). Synthetic complete (SC; 0.67% Yeast Nitrogen Base without amino acids, 2% dextrose, 0.08% Drop Out mix) and synthetic defined (SD; 0.67% Yeast Nitrogen Base without amino acids, 2% dextrose) media, both with drop out amino acids, depending on auxotrophy of the strains, were used as selective media. Solid media were obtained by adding 2% agar. 5-FOA containing agar medium (0.67% Yeast Nitrogen Base without amino acid, 0.0625% 5-Fluoroorotic Acid (Toronto Research Chemicals), 0.01% uridine, 2% glucose, 2% agar supplemented with arginine and histidine) was used to detect cells that have an unfunctional *URA3* gene. Mycophenolic acid (MPA) medium is composed of SD medium with 5µg/mL MPA (Sigma-Aldrich®). Meanwhile, *E. coli* strains were cultured on/in LB or 2YT media with appropriate antibiotics for selection purposes (Kanamycin 50µg/mL, Ticarcillin 50µg/mL and Gentamicin 10µg/mL).

In silico identification of potential recessive lethal alleles

A collection of 182 *C. albicans* genomes (Ropars *et al.* 2018) was utilized with the intention to search for potential RLAs. SNPs were compiled and filtered with the following criteria: (i) a SNP that introduces a premature STOP codon in a coding region (ORF, open reading frame), (ii) this SNP was never observed in the homozygous state, (iii) the SNP lies in a coding region never found to be dispensable in *C. albicans* and (iv) the SNP was found in the heterozygous state in the reference strain SC5314.

Plasmid constructions

Methods for the construction of plasmids used for I-*Sce*I target sequence (TS) insertion (pFA-*URA3*-I*Sce*I-TS-*HSP90*/*CUP9*) and for complementation with the wild type *MTR4* allele (CIp10-P$_{TDH3}$-*MTR4*-*IMH3*) are described in Supplementary methods.

***C. albicans* strain constructions**

Construction of strain CEC4685 is described in Supplementary methods. Briefly, this strain is derived from the common laboratory strain SN76 and carries the I-*Sce*I-encoding gene under the control of the inducible P_{TET} promoter (Park and Morschhäuser 2005); pNIMX coding for the transactivator necessary for activation of the P_{TET} promoter in presence of tetracycline derivatives (Chauvel *et al.* 2012); and the BFP/GFP LOH reporter system (Loll-Krippleber *et al.* 2015) on the right arm of Chr7. The integration of the I-*Sce*I TS was achieved by transformation of strain CEC4685 with a PCR-amplified cassette bearing the I-*Sce*I TS and Chr7 homology regions, the latter borne from primers 9 and 10 (Table S3) used in amplification of pFA-*URA3*-I*Sce*I-TS-*CDR3*/TG(GCC)2 (Feri *et al.* 2016). This led to prototroph strains named CEC5061 (I-*Sce*I TS on Hap A) and CEC5062 (I-*Sce*I TS on Hap B) (Table S1).

Re-localization of I-*Sce*I TS upstream of the MRS was conducted by transforming CEC4685 with *Hpa*I+*Sca*II-cut plasmid of pFA-*URA3*-I*Sce*I-TS-*HSP90/CUP9*, which carries the *URA3*-I-*Sce*I TS construction for integration on Chr7. The now prototroph transformants were selected on SD media. Final strains were named CEC4818 (I-*Sce*I TS on Hap A) and CEC4819 (I-*Sce*I TS on HapB) (Table S1). Strains possessing the BFP/GFP LOH reporter system on the left arm of Chr7 were generated by positioning the heterozygous BFP/GFP locus close to the left telomere (position 108,458 – 108,838) and the I-*Sce*I TS at position 367,060 – 367,145 between *mrs-7a* (position 228,342 – 242,083) and the centromere (position 425,808 – 428,708).

Complementation of CEC5061, CEC5062, CEC4818 and CEC4819 with the wild-type *MTR4* allele at the *RPS1* locus was achieved using the *Stu*I linearized CIp10-P_{TDH3}-*MTR4-IMH3* plasmid described in Supplementary methods. Transformants were selected on SD + MPA 5µg/mL (Sigma-Aldrich) media, final strains were named CEC5067, CEC5072, CEC5075 and CEC5078, respectively (Table S1).

Following each transformation step, junction PCRs were conducted to ensure proper integrations. Additional controls were conducted such as fluorescence check (flow cytometry of 20,000 cells) and auxotrophy testing of strains (by spotting on SD supplemented with either Arg, His, Ura or MPA). To ensure that the strains generated in the course of this study did not display growth defects resulting from transformation events, growth curves were generated using a TECAN Sunrise™ with OD readings (620 nm) every 10 min for a period of 48h. Doubling times were outputted from the generated OD values, and analyzed using GraphPad Prism 5.

Induction of the *Tet*-On system

In order to activate the *Tet*-On promoter and achieve I-*Sce*I protein production, single colonies were pre-cultured in liquid SC-His-Arg medium at 30°C. After overnight growth, induction was conducted in YPD + anhydrotetracycline (ATc) (3 µg/mL) (Thermofisher ACROS Organics™) for 8 h at 30°C followed by an overnight recovery in YPD.

5-Fluoroorotic acid (5-FOA) selection

Following the I-*Sce*I induction protocol, as seen above, three different cell dilutions of cultures (20,000 cells, 2,000 cells and 200 cells) grown in presence (induced) or absence (non-induced, control) of ATc were plated on 5-FOA-containing plates in triplicates. Dilutions were verified by plating a volume corresponding to 100 cells on YPD plates. Plates were incubated at 30°C for 3 days before analysis.

Cell preparation for flow cytometry and analysis

All flow cytometry analyses were conducted on the MACSQuant® Analyzer (Miltenyi Biotec) where the BFP is detected with a 405 nm laser and 425-475 nm filters and the GFP is detected with a 488 nm laser and 500-550 nm filters. Data for a maximum of 10^6 cells were analyzed using the FlowJo V10.1 software. The gates to determine the LOH frequencies were arbitrarily selected but conserved throughout sample analysis.

Cell sorting

Induced and non-induced cultures were filtered using BD Falcon™ Cell strainers in order to remove large debris and filamentous cells that could obstruct the tubing system of the cytometer. The MoFlo® Astrios™ flow cytometer was used to analyze and sort the cells of interest. For each sorted gate, 1,000 cells were recovered in 400 µL of liquid YPD medium, plated immediately after cell sorting on four YPD Petri plates and incubated at 30°C for 48h before collection of results.

SNP-RFLP

In silico identification of heterozygous SNPs affecting a restriction site on only one haplotype of Chr7 was used for haplotype characterization of strains. First, a nucleotide multiple sequence alignment by MUSCLE was conducted using the Chr7A and Chr7B sequences from the reference strain SC5314 (Skrzypek *et al.* 2017). Heterozygous SNPs were selected using the following criteria: (i) interrupts a commonly known restriction site on one haplotype, (ii) selected enzyme does not cut again within a range of 1 kb and (iii) the heterozygous SNP is present in most strains of the 182 *C. albicans* clinical strains collection (Ropars *et al.* 2018).

Secondly, primer pairs were designed to result in PCR products with different digestion profiles and used to verify the presence of the heterozygous SNP in our strain of interest.
These SNPs were used for two distinct purposes: (i) to assign the Chr7 homolog targeted by I-*Sce*I TS integration and (ii) to monitor the heterozygous status of the left and right arms of Chr7 upon DNA-DSB repair. Regions of roughly 2.5 kb surrounding the heterozygous SNPs were amplified by PCR and digested with the appropriate restriction enzyme overnight. *C. albicans* gDNA extractions, PCRs and amplicon verifications were conducted following Feri *et al.* while a list of primers used can be found in Table S3. The SNPs at positions 444,929 (*Bgl*II cutting HapA) and 727,328 (*Hpa*I cutting HapA) were used to identify the I-*Sce*I targeted haplotype of the right arm of Chr7, when the I-*Sce*I TS is located upstream and downstream of the *mrs*-7b, respectively. 2.2 kb and 2.16 kb regions around the heterozygous SNPs located at positions 414,508 (left arm) and 744,964 (right arm) utilizing *Att*I (cutting HapB) and *Hae*II (cutting HapB) enzymes, respectively, were used to assess the heterozygous status of both Chr7 arms. All SNP-RFLP sites and the enzyme sensitive haplotypes are summarized in Figure S1.

Acknowledgements

TM is the recipient of a PhD fellowship from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID). AF was the recipient of a PhD fellowship from INRA and Institut Pasteur. We acknowledge support from the French Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases, ANR-10-LABX-62-IBEID).

Supplementary material

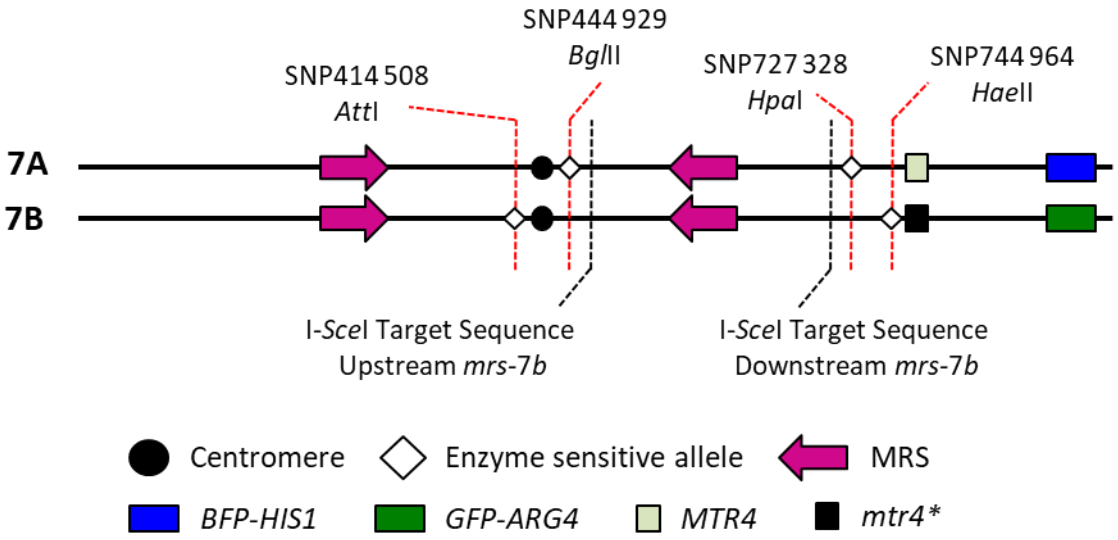

Figure S1: Overview of Chr7 in engineered strains throughout this study.
Schematic localization of the I-*Sce*I target sequences relative to the MRSs, the BFP/GFP LOH reporter system and the *MTR4*/*mtr4*^{K880*} locus. Sites used for SNP-RFLP and associated restriction enzymes are shown, where a rhombus identifies the enzyme sensitive allele.

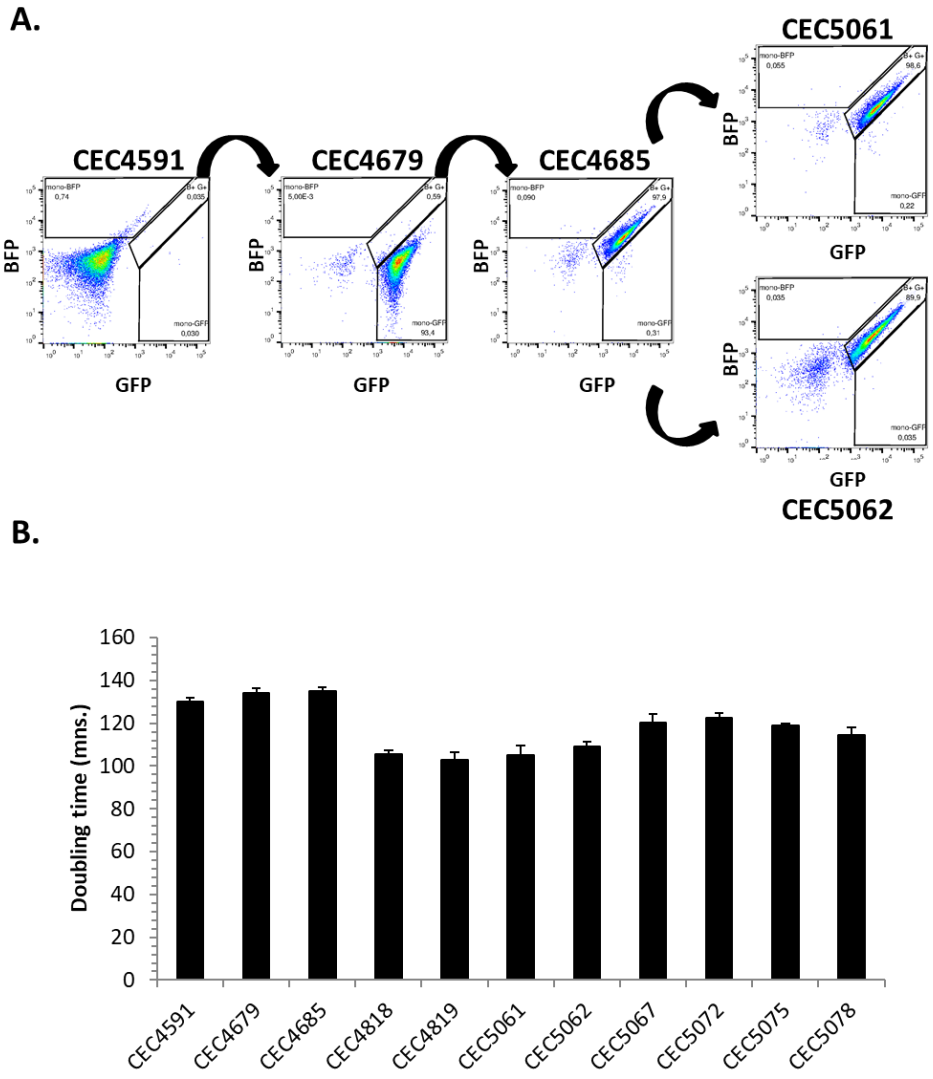

Figure S2: General characterization of strains.
A. Validation of the fluorescence of strains by flow cytometry. 20,000 events per sample are analysed.
B. Doubling times of *C. albicans* strains used throughout this study in YPD at 30°C.

Table S1: C. albicans strains used in this study.

Strain	Genotype	Characteristic	Auxotrophies	Parental Strain	Reference
CEC4591	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436	pNIMx-P_{TET}-NLS-I-*Sce*I	arg- his- uri- Hyg^R Nsn^R	CEC1448	(1)*
CEC4679	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*	CEC4591 + Ch7Right_P_{TDH3}-*ARG4-GFP*	ARG+ his- uri- Nsn^R Hyg^R	CEC4591	This study
CEC4685	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1*	CEC4679 + Chr7Right_P_{TDH3}-*HIS1-BFP*	ARG+ HIS+ uri- Nsn^R Hyg^R	CEC4679	This study
CEC4818	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:446552 to 446704::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS	CEC4685 + re-located I-*Sce*I TS on Hap A	ARG+ HIS+ URI+ Nsn^R Hyg^R	CEC4685	This study
CEC4819	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:446552 to 446704::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS	CEC4685 + re-located I-*Sce*I TS on Hap B	ARG+ HIS+ URI+ Nsn^R Hyg^R	CEC4685	This study
CEC5061	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:725867 to 726122::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS	CEC4685 + I-*Sce*I TS on Hap A	ARG+ HIS+ URI+ Nsn^R Hyg^R	CEC4685	This study
CEC5062	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:725867 to 726122::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS	CEC4685 + I-*Sce*I TS on Hap B	ARG+ HIS+ URI+ Nsn^R Hyg^R	CEC4685	This study

CEC5067	*arg4Δ/arg4Δ his1Δ/his1Δ* *ura3Δ::λimm434/ura3Δ::λimm434* *iro1Δ::λimm434/iro1Δ::λimm434* *ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:446552 to 446704::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS *RPS1*::Clp10/*RPS1*::P_{TDH3}-*MTR4-IMH3r*	CEC4818 + P_{TDH3}-*IMH3-MTR4*	ARG+ HIS+ URI+ Nsn^R Hyg^R Mpa^R	CEC4818	This study
CEC5072	*arg4Δ/arg4Δ his1Δ/his1Δ* *ura3Δ::λimm434/ura3Δ::λimm434* *iro1Δ::λimm434/iro1Δ::λimm434* *ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:446552 to 446704::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS *RPS1*::Clp10/*RPS1*::P_{TDH3}-*MTR4-IMH3r*	CEC4819 + P_{TDH3}-*IMH3-MTR4*	ARG+ HIS+ URI+ Nsn^R Hyg^R Mpa^R	CEC4819	This study
CEC5075	*arg4Δ/arg4Δ his1Δ/his1Δ* *ura3Δ::λimm434/ura3Δ::λimm434* *iro1Δ::λimm434/iro1Δ::λimm434* *ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:725867 to 726122::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS *RPS1*::Clp10/*RPS1*::P_{TDH3}-*MTR4-IMH3r*	CEC5061 + P_{TDH3}-*IMH3-MTR4*	ARG+ HIS+ URI+ Nsn^R Hyg^R Mpa^R	CEC5061	This study
CEC5078	*arg4Δ/arg4Δ his1Δ/his1Δ* *ura3Δ::λimm434/ura3Δ::λimm434* *iro1Δ::λimm434/iro1Δ::λimm434* *ADH1/adh1*::P_{TDH3}-carTA::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-*HYGb*-P_{TET}-ATG-NLS-I-*Sce*I/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr7_C_albicans_SC5314:725867 to 726122::P_{TDH3}-Cd*URA3*-I-*Sce*I-TS *RPS1*::Clp10/*RPS1*::P_{TDH3}-*MTR4-IMH3r*	CEC5062 + P_{TDH3}-*IMH3-MTR4*	ARG+ HIS+ URI+ Nsn^R Hyg^R Mpa^R	CEC5062	This study

*****(1) Feri *et al.* 2016

Table S2: Plasmids used in this study

Plasmid	Usage	Reference
pCRBluntII-P_{TDH3}-*GFP-ARG4*	Yeast Fluorescent Cassette	(2)*
pCRBluntII-P_{TDH3}-*BFP*-Cd*HIS1*	Yeast Fluorescent Cassette	(2)*
pFA-*URA3*-I*Sce*I_*CDR3*/TG(GCC)2	I-*Sce*I TS downstream *mrs*-7b	(1)*
pCRII::*HSP90*	I-*SceI* TS upstream *mrs*-7b	This study
pCRII::*CUP9*	I-*SceI* TS upstream *mrs*-7b	This study
pFA-*URA*3-I*Sce*I_*CDR3*/*CUP9*	I-*SceI* TS upstream *mrs*-7b	This study
pFA-URA3-I*Sce*I_*HSP90*/*CUP9*	I-*SceI* TS upstream *mrs*-7b	This study
pDONR207	*MTR4* recomplementation	(3)*
pDONR207-*MTR4*	*MTR4* recomplementation	This study
CIp-P_{TDH3}-GTW-*LEU2*	*MTR4* recomplementation	(1)
pCRII-*IMH3r*	*MTR4* recomplementation	This study
CIp-P_{TDH3}-GTW-*IMH3r*	*MTR4* recomplementation	This study
CIp-P_{TDH3}-*MTR4-IMH3r*	*MTR4* recomplementation	This study

***** (1) Feri *et al.* 2016, (2) Loll-Krippleber et al. 2015, (3) Chauvel *et al.* 2012

Table S3: Primers used in this study. *

Primer	Name	Sequence	Usage
1	K7_yFP_Ch7Right_Fwd	CACCGATAGGAATCTGGTCCAGAAGATTGG TTGTCATGTATTTTGGTCTATAAGAATATGA GGTGACAACTTGCATTGGTGTGTAGTAGAT TTCTGGTTGCAGGAAACAGCTATGACC	Integration of the BFP/GFP LOH reporter system
2	K7_yFP_Ch7Right_Rev	TTTTCCTTTAGGCAAGTTCGCCTTTAGAGAA GATTCCATTTATCAGAGGTTGAAACTAAAAG AGGCGAAGAGCTACTTTATTGTCACCGAGC ATGCTTGGGTTTTCCCAGTCACGACG	Integration of the BFP/GFP LOH reporter system
3	Ch7_Right_YFP_Integration_F	CCATTTGACAAGTTTAGAGGAGAC	Integration of the BFP/GFP LOH reporter system
4	Ch7_Right_YFP_Integration_R	GTTTACAAACCCTGGGTCTC	Integration of the BFP/GFP LOH reporter system
5	CaTDH3-R-BsiWI	AGAGAGCGTACGTGTAAAGTTTGTTGATGT TAA	Integration of the BFP/GFP LOH reporter system
6	S Arg1	GCTACCGATATGAGAATTTTCGTTCG	Integration of the BFP/GFP LOH reporter system
7	CdHIS1_detectF	AATGCTGCAGCTTATTGAGCGGTG	Integration of the BFP/GFP LOH reporter system
8	CdHIS1_detectR	TGCCCTTCTACCTGGAGTAATGGT	Integration of the BFP/GFP LOH reporter system
9	K7_URA3_ISceI_F	GAAATTGGTTTGTTCATTTAAAAAATGGGTT ATGCAACACTGCCCAGTTCTAGCTTGTTGTA ATTATTGTAGTGGTCC**ATTACCCTGTTATCC CTA**GCTCGTTTAAACTAGAAGGACCAC	Integration of I-SceI TS downstream of the mrs-7b
10	K7_URA3_ISceI_R	TTCCTCTCGACACTGCCCAATAAAATTGTTC TGCCTCTGTTGATAAGGCGGTACAAATTTA GTTGAGCCCATTACTTGGTTGTTTTGATCCT TTACGCATCGACGGATCCGGATGGTATA	Integration of I-SceI TS downstream of the mrs-7b
11	K7_URA3_ISceI_verif_F	ACCACCGGTAGCAACATTCT	Integration of I-SceI TS downstream of the mrs-7b
12	K7_URA3_ISceI_verif_R	ATGGTTGTGCACGTGATCAA	Integration of I-SceI TS downstream of the mrs-7b
13	URA3verfor	GAGGATTGTTTGGTAAAGGAAGAG	Integration of I-SceI TS downstream of the mrs-7b
14	C7_02030w_DWN_F	GGGAAAGTTAACCCTTGGCTGTAATGAGCC TTG	Integration of I-SceI TS upstream of the mrs-7b
15	C7_02030w_DWN_R	AAAGGGCCGCGGGTGGATGGGTAATACTT GCCTG	Integration of I-SceI TS upstream of the mrs-7b
16	C7_02040c_UP_F	GGGAAACTGCAG**ATTACCCTGTTATCCCTA** CAGTGACTAACGTTTCCAATGG	Integration of I-SceI TS upstream of the mrs-7b
17	C7_02040c_UP_R	AAAGGGAAGCTTGCTATTGCATACGGAGAT ATCTC	Integration of I-SceI TS upstream of the mrs-7b
18	Verif_C7_02030c_F	CCTGGTAGAACATGAGCTCT	Integration of I-SceI TS upstream of the mrs-7b
19	Verif_C7_02040w_R	GGAGTGTCAATTACGGATTG	Integration of I-SceI TS upstream of the mrs-7b
20	SNPs_414508_414582_F	AGCACTCGCCAATCGTGATA	SNP-RFLP
21	SNPs_414508_414582_R	TCATCATCGTCGACGTCATC	SNP-RFLP
22	SNP444929_F	AGCAGTATCAACGCCAGCAT	SNP-RFLP
23	SNP744964_F	ACATCGAACCTCTAGGCGTAG	SNP-RFLP
24	SNP273287_R	TGCTTTGGACACCATGATGC	SNP-RFLP
25	61_E06_F	GGGGACAAGTTTGTACAAAAAAGCAGGCTT GATGGATGACTTGTTTGATGTCTTTGATGAA	*MTR4* recomplementation
26	61_E06_R	GGGGACCACTTTGTACAAGAAAGCTGGGTC CATATACAATGAAGATACTTGTACAAAATC	*MTR4* recomplementation
27	MTR4_SNP_TM_F	AGTTCGACTGCCAAGAGAACA	*MTR4* recomplementation
28	MTR4_SNP_TM_R	CTGGTTCAGCAAGTTCTGGT	*MTR4* recomplementation
29	MTR4_ORF_Verif1_F	CCTCAAGCAGCTCCACCTAA	*MTR4* recomplementation
30	MTR4_ORF_Verif2_F	TCTGCTGGTAAAACCGTGGT	*MTR4* recomplementation
31	MTR4_ORF_Verif3_F	GACGACCCTGCTTCAAGTGA	*MTR4* recomplementation
32	MTR4_ORF_Verif4_F	AGACGTGGGTTGGATGATCG	*MTR4* recomplementation
33	MTR4_ORF_Verif5_F	GTACGTAAAGTTATTACTCACCCAGG	*MTR4* recomplementation
34	MTR4_ORF_Verif6_F	TGAAAGGTAGAGTAGCTGCTGA	*MTR4* recomplementation
35	CaIMH3r_SacI_F	CCCAAAGAGCTCGTCGACGGTATCGATAAG	*MTR4* recomplementation
36	CaIMH3r_SacI_R	CCCAAAGAGCTCCTGCAGGAATTCGATGTA	*MTR4* recomplementation
37	IMH3r_verif_F	AATTCGAAGGGGGCGTTCAT	*MTR4* recomplementation
38	IMH3r_verif_R	ACCGATAAACCATCTTTCTTAGGGT	*MTR4* recomplementation

*Restriction sites are underlined while the I-*Sce*I target sequence is identified in bold blue font

Table S4: List of heterozygous SNPs inducing premature STOP codons in the C. albicans SC5314 genome.

# in Figure 4	Chr-Position	Position	ORF Name		Gene Name	Present in Muzzey's table (Muzzey *et al.* 2013)	Amino Acid Impacted
1	Ca22chr1A	296097	C1_01500W	orf19.3336		yes	R/*
2	Ca22chr1A	744053	C1_03540C	orf19.3057		yes	Q/*
3	Ca22chr1A	821187	C1_03880C	orf19.4461		yes	C/*
4	Ca22chr1A	854313	C1_04060W	orf19.4482	*IFI3*	yes	W/*
5	Ca22chr1A	881789	C1_04250C	orf19.1060		yes	L/*
6	Ca22chr1A	1131425	C1_05380C	orf19.427		yes	Q/*
7	Ca22chr1A	1255338	C1_05970W	orf19.2449		yes	E/*
8	Ca22chr1A	1255367	C1_05970W	orf19.2449		yes	Y/*
9	Ca22chr1A	1286182	C1_06140C	orf19.2431		yes	Y/*
10	Ca22chr1A	1286232	C1_06140C	orf19.2431		yes	R/*
11	Ca22chr1A	1711007	C1_07860W	orf19.5057		yes	S/*
12	Ca22chr1A	1760072	C1_08070W	orf19.5079	*CDR4*	yes	Y/*
13	Ca22chr1A	1882812	C1_08600C	orf19.407	*GCD6*	yes	W/*
14	Ca22chr1A	1972370	C1_09060C	orf19.4763		yes	L/*
15	Ca22chr1A	1986384	C1_09100W	orf19.4767	*ZCF28*	yes	S/*
16	Ca22chr1A	2175131	C1_09960W	orf19.4862.2	*PET100*	yes	L/*
17	Ca22chr1A	2461704	C1_11200W	orf19.2296		yes	Q/*
18	Ca22chr1A	2466747	C1_11210C	orf19.681	*HAP43*	yes	Y/*
19	Ca22chr1A	2801839	C1_12860C	orf19.4916		yes	Q/*
20	Ca22chr1A	2808532	C1_12900W	orf19.4921		yes	Q/*
21	Ca22chr1A	2915254	C1_13370W	orf19.4963		yes	L/*
22	Ca22chr1A	2959172	C1_13490C	orf19.4981		yes	Q/*
23	Ca22chr2A	466036	C2_02350C	orf19.1555	*SAC3*	yes	S/*
24	Ca22chr2A	615045	C2_03050W	orf19.5796	*SHE9*	yes	L/*
25	Ca22chr2A	693284	C2_03340W	orf19.894		yes	Y/*
26	Ca22chr2A	928706	C2_04430W	orf19.4509		yes	Y/*
27	Ca22chr2A	929603	C2_04440W	orf19.4508		yes	L/*
28	Ca22chr2A	980303	C2_04670W	orf19.149		yes	Y/*
29	Ca22chr2A	1388153	C2_06760C	orf19.2227		yes	L/*
30	Ca22chr2A	1585971	C2_07750W	orf19.2204.2	*RHO2*	yes	W/*
31	Ca22chr2A	1779322	C2_08750W	orf19.3600		yes	E/*
32	Ca22chr2A	2054077	C2_10030C	orf19.1779	*MP65*	yes	E/*
33	Ca22chr2A	2065445	C2_10070W	orf19.1774		yes	Q/*
34	Ca22chr3A	336960	C3_01550C	orf19.1690	*TOS1*	yes	Y/*
35	Ca22chr3A	445420	C3_02020W	orf19.1646		yes	R/*
36	Ca22chr3A	530036	C3_02420C	orf19.230		yes	L/*
37	Ca22chr4A	166194	C4_00910C	orf19.4703		yes	R/*
38	Ca22chr4A	166523	C4_00910C	orf19.4703		yes	W/*
39	Ca22chr4A	235553	C4_01170C	orf19.4673	*BMT9*	yes	Q/*
40	Ca22chr4A	531045	C4_02560C	orf19.2745	*UME7*	yes	L/*
41	Ca22chr4A	659155	C4_03130W	orf19.2677	*GPI16*	yes	R/*
42	Ca22chr4A	792650	C4_03720C	orf19.1306		yes	S/*
43	Ca22chr4A	796679	C4_03750C	orf19.1303	*MRF2*	yes	R/*

44	Ca22chr4A	833072	C4_03910W	orf19.1596	*FGR28*	yes	L/*
45	Ca22chr4A	833643	C4_03910W	orf19.1596	*FGR28*	yes	Y/*
46	Ca22chr4A	874950	C4_04080C	orf19.5302	*PGA31*	yes	Y/*
47	Ca22chr4A	1104475	C4_05120C	orf19.741		yes	W/*
48	Ca22chr4A	1262912	C4_05720W	orf19.1257		yes	W/*
49	Ca22chr4A	1536852	C4_06900W	orf19.3121	*GST1*	yes	K/*
50	Ca22chr4A	1544443	C4_06950W	orf19.3114		yes	W/*
51	Ca22chr4A	1580763	C4_07130W	orf19.3091		yes	E/*
52	Ca22chr5A	296955	C5_01330W	orf19.1935		yes	W/*
53	Ca22chr5A	484392	C5_02180C	orf19.4225	*LEU3*	yes	Q/*
54	Ca22chr5A	524833	C5_02360C	orf19.4244		yes	W/*
55	Ca22chr5A	691310	C5_03120W	orf19.4341		yes	Q/*
56	Ca22chr5A	747018	C5_03300C	orf19.2650		yes	S/*
57	Ca22chr6A	24866	C6_00230W	orf19.1181		yes	W/*
58	Ca22chr6A	801191	C6_03710W	orf19.5742	*ALS9*	yes	S/*
59	Ca22chr7A	692267	C7_03190C	orf19.5137		yes	Y/*
60	Ca22chr7A	746359	C7_03400C	orf19.1335	*MTR4*	yes	K/*
61	Ca22chrRA	419533	CR_01870C	orf19.2585		yes	Y/*
62	Ca22chrRA	677885	CR_02930W	orf19.2850		yes	L/*
63	Ca22chrRA	803707	CR_03620C	orf19.4380.1		yes	K/*
64	Ca22chrRA	922066	CR_04110W	orf19.494		yes	C/*
65	Ca22chrRA	1018552	CR_04670C	orf19.1737		yes	Y/*
66	Ca22chrRA	1136971	CR_05270W	orf19.1008		yes	W/*
67	Ca22chrRA	1137061	CR_05270W	orf19.1008		yes	W/*
68	Ca22chrRA	1603863	CR_07380C	orf19.6131	*KSR1*	yes	R/*
69	Ca22chrRA	1734063	CR_07930C	orf19.605		yes	E/*
70	Ca22chrRA	1782353	CR_08200C	orf19.6382		yes	Q/*

*Grey represent multiple pre-mature STOP codons within the same ORF, Confirmed RLAs in red and recessive deleterious allele in yellow

Supplementary methods

Plasmid constructions

A plasmid carrying a replacement cassette to introduce the I-*Sce*I target sequence (TS) on the right arm of Chr7, near the centromere, was constructed to be integrated in the 8 kb gene-free region between *HSP90* and *CUP9***.** This plasmid carries the *URA3* auxotrophic marker and the I-*Sce*I TS (5'-ATTACCCTGTTATCCCTA-3') flanked by two 1 kb genomic regions homologous to the Chr7 integration locus in proximity of the centromere (Figure S1). Both Chr4 homology regions, from the pFA-*URA3*-I*Sce*I-TS-*CDR3*/TG(GCC)2 plasmid (Feri *et al.* 2016), were replaced by Chr7 homology regions. These two homology portions are composed of sequences from the Chr7 intergenic region located between *HSP90* and *CUP9* from SC5314 gDNA. The first region at position 445,752-446,549, closer to *HSP90*, was amplified with primers 18 and 19 (Table S3), which possess *Sac*II and *Hpa*I restriction site tails, respectively. The second amplified region (primers 16 and 17) at position 446,703-447,696, closer to *CUP9*, carries tails that contain the I-*Sce*I TS-*Pst*I and *Hind*III restriction sites, respectively. These PCR fragments were cloned into the TOPO®-TA pCRII plasmid yielding plasmids pCRII::*HSP90* and pCRII::*CUP9*. pFA-*URA3*-I*Sce*I-TS-*CDR3*/TG(GCC)2 was digested with *Pst*I and *Hind*III to replace one of the Chr4 regions with the *Pst*I,*Hind*III-digested fragment from pCRII::*CUP9*. The resulting plasmid pFA-*URA3*-I*Sce*I-TS-*CDR3*/*CUP9* and pCRII::*HSP90* were both digested with *Hpa*I and *Sca*II to replace the second Chr4 region by the second Chr7 region, resulting in pFA-*URA3*-I*Sce*I-TS-*HSP90/CUP9*.

For *MTR4* complementation at the *RPS1* locus, the CIp10-P_{TDH3}-*MTR4-IMH3* plasmid containing the wildtype allele of *MTR4* expressed under the control of the constitutive promoter P_{TDH3} and associated to the *IMH3* gene conferring resistance to MPA was constructed. To this aim, the *IMH3* gene was amplified by PCR (Expand™ Long range dNTPack, Roche) from *pSF*II (Morschhäuser *et al.* 2005) and cloned in the TOPO®-TA pCRII vector (ThermoFisher Scientific). The *IMH3* fragment was then recovered from the resulting plasmid by *Sac*II digestion and cloned in the *Sac*II-linearized CIp10-P_{TDH3}-GTW-*LEU2* (Feri *et al.* 2016) giving rise to CIp10-P_{TDH3}-GTW-*IMH3*. The ORF coding for *MTR4* was amplified by PCR using primers 25 and 26 (Table S3) containing aatP1 and attP2 sites for Gateway™-recombinational cloning and transferred in pDONR207 using the BP clonase™ as previously described (Legrand *et al.* 2018). Several recombined plasmids were sequenced to confirm the cloned haplotype. pDONR207::*MTR4A* containing *MTR4* haplotype A was then used in combination with CIp10-P_{TDH3}-GTW-*IMH3* and LR clonase™ (Invitrogen) to yield CIp10-P_{TDH3}-*MTR4-IMH3.*

C. albicans strain constructions

The parental strain CEC4591 (Table S1) was sequentially transformed by heat shock according to the Lithium Acetate/PEG protocol (Gola *et al.* 2003), in order to obtain the final constructions represented in Figure 1A and Supplemental Figure S1. CEC4591 is derived from the common laboratory strain SN76 and carries the I-*Sce*I gene under the control of the inducible P_{TET} promoter (Park and Morschhäuser 2005), as well as pNIMX (Chauvel *et al.* 2012) coding for the transactivator necessary for the proper activation of the P_{TET} promoter in presence of tetracycline derivatives. The strategy was to introduce (i) the BFP/GFP LOH reporter system on the right arm of Chr7 near the telomeres and (ii) the I-*Sce*I TS on the right arm of Chr7 near the centromeres of each Chr7 homolog. CEC4591 was first transformed with the P_{TDH3}-*GFP-ARG4* (Loll-Krippleber *et al.* 2015) cassette that was designed to integrate by homologous recombination on the right arm of Chr7 at position 911,043-911,337, in a roughly 5 kb intergenic region (Figure1A and Figure S1). To do so, 120 bp primers were used, which are composed of 100 bp tails possessing the complementary sequence of the Chr7 integration locus and 20 bp complementary to the P_{TDH3}-*GFP-ARG4* cassette. Thus, primers 1 and 2 (Table S3) were used to amplify by PCR the P_{TDH3}-*GFP-ARG4* integration cassette from plasmid pCRBluntII-P_{TDH3}-*GFP-ARG4*. The cassette was amplified in a total PCR reaction volume of 500 µL, precipitated overnight in 50 µL of sodium acetate and 1 mL of 100% ethanol, pelleted by centrifugation at maximum speed for 30 min at 4°C and suspended in 25 µL of 1X TE Buffer for transformation in *C. albicans*. Transformants were selected on SC-Arg agar medium and validated by PCR to verify both boundaries of the integration, giving rise to CEC4679. Similarly, CEC4679 was subsequently transformed with the P_{TDH3}-*BFP-HIS1* cassette (Loll-Krippleber *et al.* 2015) that was designed to integrate at the same allelic locus as the P_{TDH3}-*GFP-ARG4* cassette hence resulting in the complete artificial heterozygous BFP/GFP locus. The P_{TDH3}-*BFP-HIS1* integration cassette was amplified from plasmid pCRBluntII-P_{TDH3}-*BFP*-Cd*HIS*1. Transformants were selected on SD+Uri plates and junctions were validated by PCR (Table S3). This insured that transformants possess simultaneously the P_{TDH3}-*GFP-ARG4* and the P_{TDH3}-*BFP-HIS1* cassettes, each one localized on a different chromosomal homologue. The resulting strain was CEC4685.

Chapter 3: Extension of homozygosis towards centromere upon DNA-DSB repair by break-induced replication in *Candida albicans.*

Break-induced replication often leads to extended homozygosis tract ascends towards centromere in *Candida albicans*.

Contains: Article in preparation, to be submitted to G3-Genetics

Timea MARTON[1,2*], Murielle CHAUVEL[1*], Adeline FERI[1,2*], Corinne MAUFRAIS[1,3], Christophe D'ENFERT[1], Mélanie LEGRAND[1]

[1] Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, France
[2] Université Paris Diderot, Sorbonne Paris Cité, Paris, France
[3] Hub de Bioinformatique et Biostatistique, Département de Biologie Computationnelle, USR 3756 IP CNRS, Institut Pasteur, Paris, France

* Co-first authors

Abstract

Genomic rearrangement events have been associated with the acquisition of adaptive phenotypes, allowing organisms to generate new favorable genetic variation combinations. Thus, understanding the molecular mechanisms leading to genomic rearrangements is key to better understand *Candida albicans* biology and pathogenesis . The genome of *C. albicans* is highly plastic displaying numerous genomic rearrangement events, which are often the by-product of DNA break repairs. For example, DNA double-strand breaks (DSB) repair using various homologous-recombination pathways are a major source of loss-of-heterozygosity (LOH), observed ubiquitously in both clinical and laboratory strains of *C. albicans.* Mechanisms such as break-induced replication (BIR) or mitotic crossover (MCO) result in long-tracts of LOH, spanning hundreds of kilobases until the telomere. Analysis of I-*Sce*I-induced BIR/MCO tracts in *C. albicans* revealed unexpectedly that the homozygosis tracts can ascend several kilobases towards the centromere, displaying homozygosis between the break site and the centromere. We sought to explain which molecular mechanism(s) was(were) responsible for this phenotype by characterizing a series *C. albicans* DNA repair mutants, including *pol32*$^{-/-}$, *msh2*$^{-/-}$, *spo11*$^{-/-}$, *mph1*$^{-/-}$ and *mus81*$^{-/-}$. This study permitted us to describe the deletion effect of these genes on genome stability and highlight functional differences between *Saccharomyces cerevisiae* (model DNA repair organism) and *C. albicans.* Additionally, we demonstrated that ascending LOH tracts towards the centromere are associated with intrinsic features of BIR and potentially involve the mismatch repair pathway which acts upon natural heterozygous positions. Overall, this mechanistic approach to study LOH deepens our limited characterization of DNA repair pathways in *C. albicans* and brings forth the notion that centromere proximal alleles from DNA break sites are not guarded from undergoing LOH.

Introduction

Genomic instability is defined by the increased frequency of mutations or genetic rearrangements. It facilitates rapid adaptation to changing environments and is often considered a hallmark of pathogenic species and cancer (Maslov and Vijg 2009; Darmon and Leach 2014; Covo 2020). Numerous studies, which aim to better understand the biological implications of genomic instability, rely on fundamental mechanistic studies conducted in the baker's yeast, *Saccharomyces cerevisiae,* to investigate the effect of perturbations in molecular mechanisms involved in DNA maintenance with regards to genomic rearrangements. The opportunistic fungal pathogen *Candida albicans* is an emerging powerful model organism to study the effect of genome dynamics on fungal pathogenesis, notably due to its diploid highly heterozygous and plastic genome (Hirakawa *et al.* 2015; Ropars *et al.* 2018; Wang *et al.* 2018). The genome of *C. albicans* is highly tolerant to genomic rearrangements (Fischer *et al.* 2006), as copy number variations and loss-of-heterozygosity (LOH) are frequently acquired upon exposure to stress conditions (Selmecki *et al.* 2005; Forche *et al.* 2011; Harrison *et al.* 2014; Ene *et al.* 2018). Indeed, LOH of various sizes are ubiquitous to both laboratory and clinical isolates, and can span across several nucleotides or even affect entire chromosomes (Ropars *et al.* 2018; Ene *et al.* 2018; Wang *et al.* 2018). In recent years, several gross chromosomal rearrangements have been associated with adaptive phenotypes, including antifungal resistance (Selmecki *et al.* 2008; Todd *et al.* 2019), or niche specialization by augmentation or diminution of virulence *in vivo* (Tso *et al.* 2018; Liang *et al.* 2019; Forche *et al.* 2019). Thus, understanding the molecular mechanisms leading to genomic rearrangements is key to better understand how *C. albicans* can quickly and efficiently switch from a gut commensal to a potentially life-threatening pathogen.

Despite the beneficial outcomes of genomic instability, genomic maintenance is essential to ensure the transmission of a complete and faithful DNA sequence to the subsequent generations. All cells recurrently suffer DNA damages caused by intrinsic and extrinsic factors leading to single-strand or double-strand breaks (DSB), threatening genomic integrity. DNA-DSB repair can be conducted through numerous molecular mechanisms, which can be classified into two major categories: non-homologous end joining (NHEJ) and homologous-recombination (HR)-mediated repair. NHEJ is described as a "cut and paste" DNA repair mechanism because it involves little (if any) modifications of the broken DNA ends, with the DNA break being identified and ligated together. In contrast, several different pathways of HR-mediated repair exist, all requiring extensive modifications of the broken DNA ends and an intact repair template, such as the homologous chromosome or the sister chromatid, in order to complete the repair. Indeed, as described in *S. cerevisiae,* HR-mediated

repair is a lengthier process requiring DNA damage detection, 5'-end resection leading to single-strand DNA (ssDNA), homology search, ssDNA invasion and DNA synthesis, and therefore necessitates the coordination of several cellular processes. Additionally, DNA-DSB repair can be associated with the generation of genomic rearrangements, i.e. LOH, translocations or copy number variations (Malkova *et al.* 2000; Ramakrishnan *et al.* 2018; Piazza *et al.* 2019). For instance, different repair pathways will result in different LOH sizes, ranging from a single locus to an entire chromosome. When only one end of the DSB shares homology with a template, an efficient HR pathway known as break-induced replication (BIR) may come into play to repair the break, leading to long-tract LOH which can span over hundreds of thousands of bases, stretching from the break point to the telomere. Furthermore, BIR is characterized as highly inaccurate, due to frequent template switching and slippage of polymerase δ resulting in 1,000-fold increase in indels (Smith *et al.* 2007; Deem *et al.* 2011). Additionally, BIR in *S. cerevisiae* requires extensive 5'-end resection leading to the exposure of long ssDNA which is more sensitive to DNA damaging agents and may lead to the formation of point mutation clusters (Sakofsky *et al.* 2014, 2019; Elango *et al.* 2019).

In *C. albicans*, only a limited number of studies focused on the effects of DNA repair pathway disruption on overall genomic stability, mainly through the characterization of deletion mutants of genes mostly involved in excision repair (BER, NER, MMR) or DSB repair. More recently, two studies investigated the outcomes of I-*Sce*I- and Cas9 nuclease-mediated DNA-DSB in *C. albicans,* by characterizing the repair mechanisms at stake (Feri *et al.* 2016; Vyas *et al.* 2018). These studies demonstrated that DNA-DSB is predominantly associated to HR repair in *C. albicans*, as illustrated by the frequent occurrence of LOH upon DNA-DSB repair (Feri *et al.* 2016; Vyas *et al.* 2018). These LOH, that are frequently observed in the *C. albicans* population, are counter-intuitive in regards to the several studies that have highlighted the importance of maintaining heterozygosity as it has been shown to correlate with strain fitness (Hickman *et al.* 2013; Hirakawa *et al.* 2015). There appears to be an equilibrium between genomic stability, which ensure longevity of the lineage and maintenance of genomic integrity, and genomic instability which promotes the acquisition of new allele combinations potentially favoring adaptation. This equilibrium may be influenced by extrinsic factors because the nature of molecular mechanisms leading to LOH in *C. albicans* varies based on growth conditions, e.g. exposure to H_2O_2 predominantly results in short-tract LOH while fluconazole and growth at 39°C favor the appearance of long-tract LOH (Forche *et al.* 2011). In the context of a human pathogen, environmental stimuli that would trigger genome instability could activate specific DNA repair pathways, consequently dictating the LOH size and therefore the translation of these homozygosis tracts into phenotypic diversity. Fidelity of repair mechanisms can also play an important role in the latter described equilibrium though, this

has not yet been investigated in *C. albicans*. The reliability of DNA damage repair remains to be thoroughly investigated as it plays an important role in diploid heterozygous strains where maximum heterozygosity should ideally be maintained as it has been shown to be beneficial in the long-term.

The previously developed I-*Sce*I inducible DNA-DSB system coupled to a fluorescence-based LOH reporter system, permits the detection, quantification, and sorting of *C. albicans* cells that had undergone a long-tract LOH, as a result of homologous recombination (Feri *et al.* 2016). By using this set up, we sought out to investigate the fidelity of BIR in *C. albicans,* as it was proposed to be the main DNA repair mechanism at the origin of the long-tract LOH often observed in laboratory and clinical isolates (Hirakawa *et al.* 2015; Ropars *et al.* 2018; Wang *et al.* 2018). Analysis of whole-genome sequencing data of several isolates which have undergone I-*Sce*I-dependent BIR/MCO revealed that long-tract LOH are quite often associated with events of homozygosis between the I-*Sce*I break site and the centromere. In order to identify the potential molecular mechanisms involved in these ascending LOH tracts, we constructed deletion mutants for the following genes (i) *POL32*: a component of the Pol δ complex involved in HR-mediated repair, (ii) *MSH2*: a key component of the mismatch repair (MMR) pathway, (iii) *SPO11*: involved in the initiation of meiotic recombination events, (iv) *MPH1*: dissolvase of the Holliday junctions or (v) *MUS81*: resolvase of Holliday junctions. Thus, we evaluated the impact of these mutations on *C. albicans* biology and on the homozygosis events occurring at the I-*Sce*I break site upon BIR/MCO. We show that *pol32*$^{-/-}$*, msh2*$^{-/-}$*, mph1*$^{-/-}$ or *mus81*$^{-/-}$ deletions lead to increased genomic instability and/or impact the molecular mechanisms used upon I-*Sce*I DNA-DSB induction in *C. albicans.* Furthermore, deletion of either *POL32*, *MPH1* or *MSH2* gene, increased the overall precision of BIR-repair as homozygosis tracts start much closer to the I-*Sce*I break site. Therefore, the size of homozygosis regions previously observed between the I-*Sce*I break site and the centromere is reduced. Similarly to *S. cerevisiae,* extensive 5' resection occurs during BIR in *C. albicans*, which promotes overflow of homozygosis towards the centromere. Additionally, this extensive 5' resection leads to long 3' ssDNA which upon strand invasion, heteroduplex DNA formation and D-loop extension promotes the activity of MMR machinery which "corrects" natural heterozygous positions found in *C. albicans,* extending the region impacted by LOH by thousands of bases. Overall, this study permitted to characterize the role of three new genes (*pol32*$^{-/-}$*, mph1*$^{-/-}$ and *mus81*$^{-/-}$) on genome stability in *C. albicans* and introduced the notion that repair through BIR is imprecise often associated with ascending LOH towards the centromere.

Results

Unexpected long-tract homozygosis centromere-proximal to the I-*Sce*I-induced double-strand break

We previously described the use of a LOH reporter system coupled with a DNA-DSB inducing system taking advantage of the *S. cerevisiae* I-*Sce*I meganuclease to investigate the mechanisms involved in DNA-DSB repair in *C. albicans* (Loll-Krippleber *et al.* 2015; Feri *et al.* 2016). We distinguished different types of molecular events leading to long-tract homozygosis upon DNA-DSB induction: (i) I-*Sce*I-independent spontaneous events, (ii) I-*Sce*I-induced BIR/MCO, and (iii) I-*Sce*I-induced gene conversion (GC) with crossover (CO) (Feri *et al.* 2016). Here, we aimed to assess the fidelity of homology-directed repair mechanisms resulting in long-tract LOH events in *C. albicans*. The strains generated by Feri *et al*. possess (i) the BFP/GFP LOH reporter system (position 471,021-481,176) and (ii) the I-*Sce*I target sequence (position 776,698-778,104) ("I-*Sce*I TargetA" on Hap A or "I-*Sce*I TargetB" on Hap B) on the left arm of Chr4 (Chr4L) and (iii) a full-length copy of *GPI16* at the *RPS1* locus in order to compensate for the recessive lethal allele on Chr4B in the SC5314 background, as well as (iv) the I-*Sce*I gene under the control of the inducible promoter P_{TET} (Feri *et al.* 2016)) (Figure S1). These strains allow the induction of a DNA-DSB on Chr4L, the recovery of cells having undergone LOH upon DNA-DSB repair and the precise study of the resulting homozygosis tracts. We selected five clones derived from both parental strains (i.e. carrying the I-*Sce*I target sequence either on Chr4A or Chr4B) and, having undergone long-tract LOH based on auxotrophy and SNP-RFLP characterization (Figure S2). These strains were whole-genome sequenced. Allele balance at heterozygous positions (ABHet) and sequencing depth confirmed that all ten sequenced isolates had undergone BIR/MCO (Figure 1, Table S1).

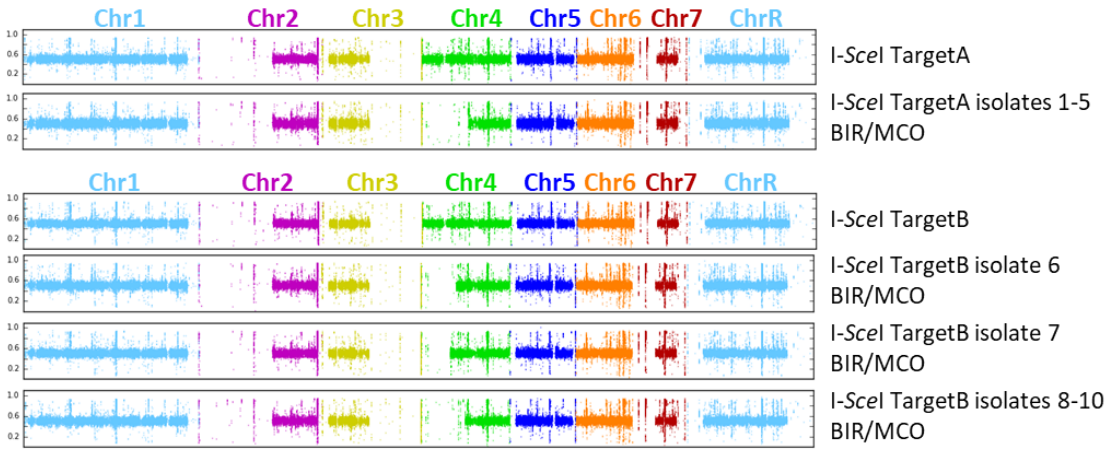

Figure 1: Allele balance at heterozygous position plots for wild type C. albicans strains of this study.
Plots showing the allele balance at heterozygous positions across the eight chromosomes. Sequencing confirms the presence of homozygosis, a long tract LOH, on the left arm of Chr4 (either haplotype A or B according to the strain) in mono-fluorescent isolates. Additional LOH events have previously been described in parental strains of "I-*Sce*I TargetA" and "I-*Sce*I TargetB" (LOH on Chr2 (Loll-Krippleber *et al.* 2015)) and in SC5314 (LOH on Chr3 and Chr7 (Abbey *et al.* 2011)).

In order to assess the fidelity of BIR/MCO in *C. albicans*, we conducted a detailed characterization of the Chr4 genomic region between positions 650,000 and 1,050,000, surrounding the I-*Sce*I break site. All heterozygous SNPs found in both parental strains were identified within this region of interest. Allele status at these positions was investigated for each of the ten *C. albicans* clones having undergone BIR/MCO upon I-*Sce*I induction, to precisely identify the start site of the homozygosis tract. As illustrated in Figure 2, our data revealed a great variability in terms of anchoring position for the selected LOH, ranging from 898 bp to 118,465 bp upstream the I-SceI target site in seven out of ten clones. Clone 5 displays homozygosis with interspersed heterozygous regions which extended almost exactly from the I-*Sce*I site towards the telomere, likely to represent template switching (Figure 2). For two out of ten clones (6 and 7), the homozygosis tract detected at the BFP/GFP locus does not go up all the way to the I-*Sce*I site, despite the actual cut by I-*Sce*I as illustrated by the loss of the I-*Sce*I target site-associated auxotrophic marker *URA3*. In these two clones, the DNA-DSB has likely been repaired by gene conversion and has been accompanied by an I-*Sce*I-independent BIR/MCO event distal to the I-*Sce*I site starting at positions 628,652 and 523,190, respectively.

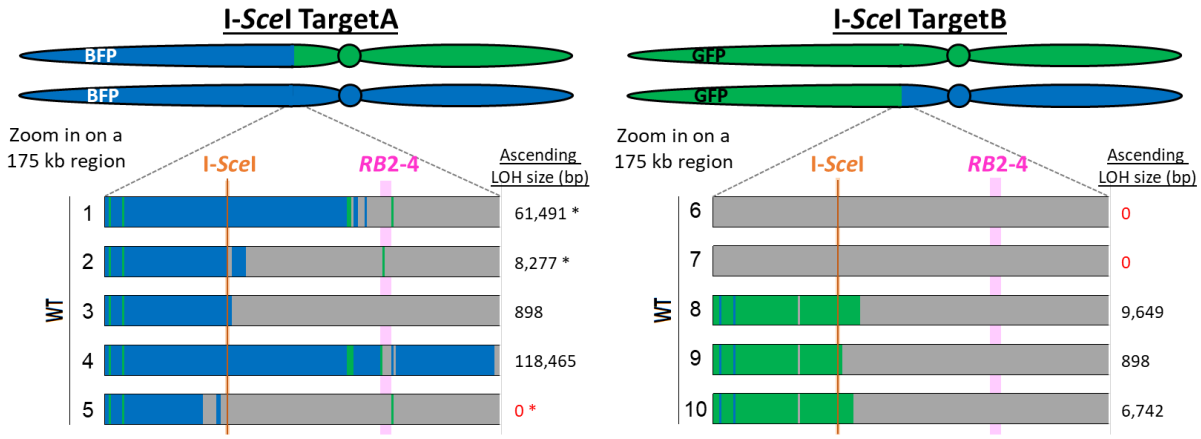

Figure 2: Results of sequencing data analysis at DNA-DSB site in cells having undergone BIR/MCO in WT genetic background.
Each diagram represents an individual sequenced isolate having undergone BIR/MCO, as determined by flow cytometry and SNP typing, following I-*Sce*I induction targeting DNA-DSB on haplotype A (I-*Sce*I TargetA) or haplotype B (I-*Sce*I TargetB). Representation of a 175 kb region surrounding the I-*Sce*I-break site (orange) was built by monitoring the homozygous/heterozygous status using 1 kb bin sizes. Several unexpected homozygosis regions located between the I-*Sce*I-break site and the centromere are identifiable within these diagrams, length (bp) indicated on the right side of each diagram. Identified in red writing are isolates having undergone spontaneous LOH on Chr4L and those with an asterisk (*) display interspersed regions of heterozygosis within long homozygosis tracts. The location of the repeat sequence *RB2-4* is identified in pink.

Phenotypic characterization of deletion mutants of genes involved in various aspects of DNA repair and recombination

Although these extended regions of homozygosis towards the centromere may result from I-*Sce*I-independent BIR, spontaneous BIR events are still rare and are unlikely to have occurred here with such a high frequency, suggesting that these extended homozygosis regions are an intrinsic feature of the molecular mechanisms involved in the repair of I-*Sce*I-induced DNA-DSB in *C. albicans*. In order to identify the molecular mechanism(s) at stake, we sought out to evaluate the impact of deleting various genes known to be involved in various aspects of DNA repair and recombination on the homozygosis events occurring at the I-*Sce*I site upon HR-mediated DNA repair through BIR/MCO. Homozygous deletion mutants were constructed in both strains carrying the I-*Sce*I target sequence on Chr4A or Chr4B for the following genes: (i) the *POL32* gene, encoding a component of the Pol δ complex involved in DNA synthesis during BIR, (ii) the *MSH2* gene, involved in identification of MMR substrates, (iii) the *SPO11* gene, whose *S. cerevisiae* ortholog is involved in meiotic recombination, (iv) the *MPH1* gene, involved in D-loop displacement and favoring non-crossover outcomes of HR-repair or (v) the *MUS81* gene encoding a resolvase involved in crossover events. Because of a heterozygous null mutation in the *GPI16* gene on Chr4L, these mutant strains were also complemented with WT *GPI16* allele to ensure viability of Chr4B homozygous cells, (Feri *et al.* 2016). Deletion

mutants were obtained for all genes using the CRISPR-Cas9 approach and further characterized.

- *Growth rate and cell morphology*

Growth rates of the deletion mutants were evaluated in standard laboratory conditions (YPD 30°C). The *mus81*$^{-/-}$ mutants displayed a drastic growth defect compared to WT strains in 24h growth kinetic experiments. The other mutant strains, *pol32*$^{-/-}$, *msh2*$^{-/-}$, *spo11*$^{-/-}$ and *mph1*$^{-/-}$ displayed doubling times similar to WT strains though *pol32*$^{-/-}$ mutant strains appeared to have a slight growth defect (Figure S3). Upon observations of colonies on agar media, we observed that *mus81*$^{-/-}$ mutants formed much smaller colonies with wrinkle edges while the other mutant strains formed smooth round colonies comparable to WT strains (Figure 3A). In liquid cultures (YPD 30°C), similarly to WT, *msh2*$^{-/-}$, *spo11*$^{-/-}$ and *mph1*$^{-/-}$ mutant strains grow in yeast form. However, pseudohyphae were observable in cultures of *pol32*$^{-/-}$ mutants and even more frequent with *mus81*$^{-/-}$ mutants, reminiscent of genotoxic stress-induced filamentation (Figure 3A). In contrast, no obvious defects in filamentation were observed in any mutant strains in liquid filamentation-inducing conditions (YPD + 10% serum or SPIDER media at 37°C for 3h), as compared to WT strains (Figure S4).

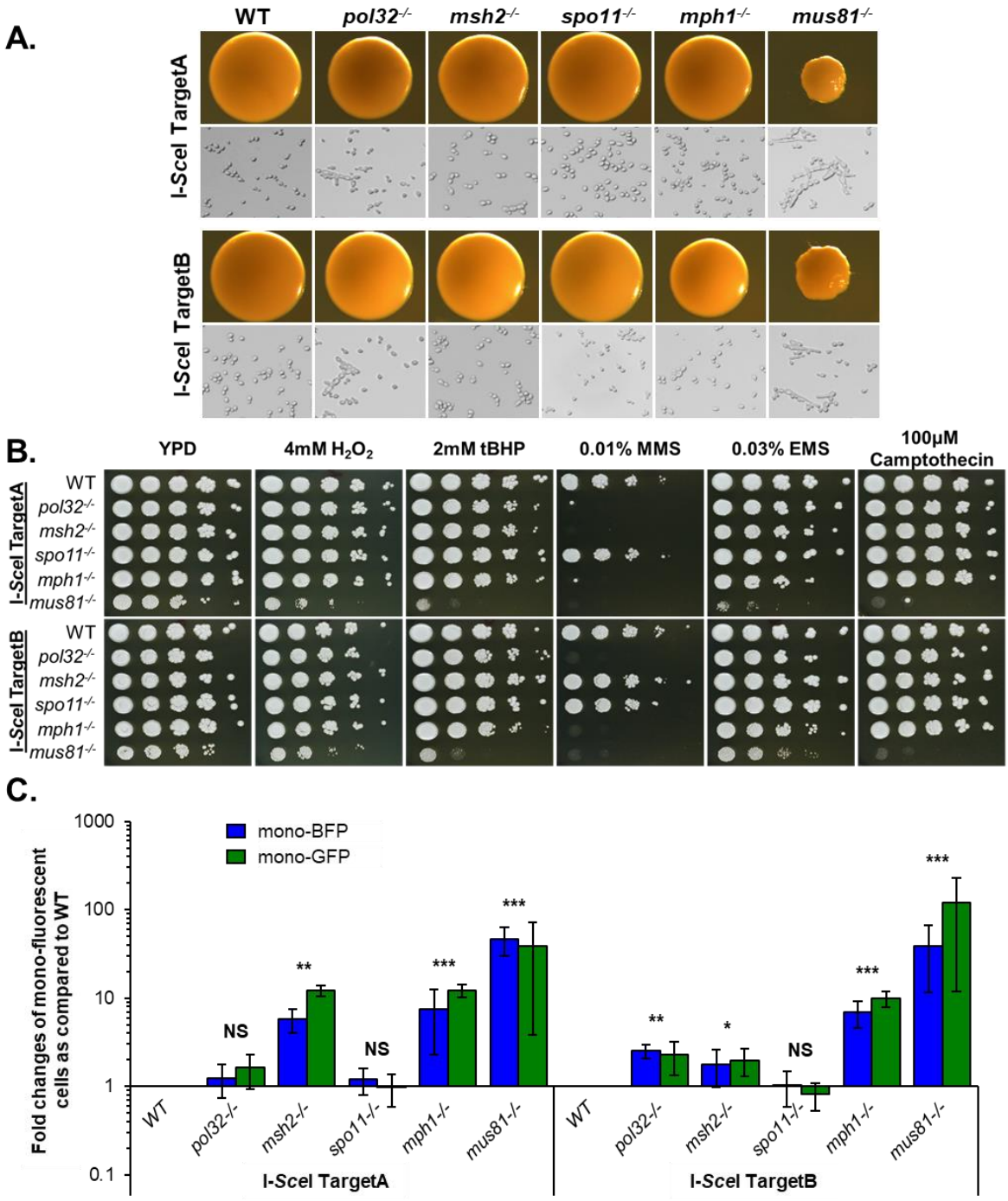

Figure 3: Characterization of DNA repair KO mutants.

A. Phenotypic characterization of cell and colony morphologies of wildtype and KO mutant strains. Photos of colonies were taken after 48h of growth on YPD agar at 30°C, whereas liquid culture were imaged after an overnight incubation. **B.** Phenotypic characterization of genotoxic sensitivity of wildtype and KO mutant strains. Spot assay conducted on YPD medium containing either; 4mM H_2O_2, 2mM tBHP, 0.01% MMS, 0.03% EMS or 100µM camptothecin. Images were taken after 2 days at 30°C. **C.** Characterization of spontaneous genomic instability of KO mutants as compared to WT strains. Histogram representing the fold changes (+/- SD) of mono-fluorescent cells (mono-BFP or mono-GFP) between WT and each KO strains. Statistical differences in overall mono-fluorescent cells (sum of mono-BFP and mono-GFP) were assessed by T-test (n=4), where NS: $p>0.05$, *: $p<0.05$, **: $p<0.01$ and ***: $p<0.001$. Example of FACS profile in Figure S6.

- *Sensitivity to DNA-damaging agents*

We also conducted drop dilution assays to evaluate the sensitivity of the deletion mutants to various DNA-damaging agents, namely oxidative agents (4 mM H_2O_2 and 2 mM tBHP), DNA alkylating agents (0.01% MMS and 0.03% EMS) and 100µM of camptothecin. Growth, as compared to WT, was observed after 48h at 30°C. While the *pol32*$^{-/-}$ and *mph1*$^{-/-}$ mutants are sensitive only to 0.01% MMS, the *mus81*$^{-/-}$ deletion mutants demonstrate sensitivity upon exposure to all five DNA-damaging agents. (Figure 3B). In contrast, the *spo11*$^{-/-}$ mutants display a phenotype identical to WT strains on all tested media (Figure 3B). While growth of the two *msh2*$^{-/-}$ strains does not seem to be affected by H_2O_2, tBHP, EMS and camptothecin, the two strains exhibit contrasting sensitivity profiles on 0.01% MMS, with the strain carrying the I-*Sce*I target sequence on HapA being more sensitive to MMS than the strain carrying the I-*Sce*I target sequence on HapB (Figure 3B). Discrepancies between *msh2*$^{-/-}$ deletion mutants may be associated with the accumulation of point mutations, as disruption of MMR pathway is associated with a mutator phenotype (Drotschmann *et al.* 1999; Healey *et al.* 2016; Boyce *et al.* 2017).

- *Chromosome 4 instability*

Because the genes targeted for deletion are key players of DNA metabolism, we also assessed genome stability in the mutant strains. We took advantage of the BFP/GFP LOH reporter system, localized on Chr4L, to evaluate genome stability in the deletion mutants under standard laboratory conditions (YPD, 30°C). While *SPO11* deletion does not appear to affect LOH frequency at the BFP/GFP locus as compared to WT, and deletion of *POL32* or *MSH2* results in only a very small augmentation in LOH frequency, the highest increases were observed for *mph1*$^{-/-}$ and *mus81*$^{-/-}$ mutants with a fold change of 8x and 58x, respectively, as compared to WT (T-test, $p<0.05$) (Figure 3C).

Additionally, we assessed the functionality of the inducible DNA-DSB system in the deletion mutant strains. Because the I-*Sce*I target sequence is associated with the *URA3* auxotrophic marker, the occurrence of a DNA-DSB at the I-*Sce*I target site, followed by efficient repair, can be monitored by counter-selection on 5-FOA-containing medium. The 5-FOA^R colonies come from cells which have likely undergone and successfully repaired an I-*Sce*I-mediated DNA-DSB. Therefore, fold changes between the frequencies of 5-FOA^R colonies in non-induced vs induced conditions can be used as indicators of cleavage efficacy. The inducible I-*Sce*I cleavage system appears to be functional in all five deletion mutants as illustrated by fold changes ranging from 11 to 1400 (Table 1). Noticeably, the *pol32*$^{-/-}$ and *mus81*$^{-/-}$ mutants display fold changes that are in the lower range, below 100, and that could

be explained by an increased mortality rate of the cells undergoing a DNA-DSB in these genetic backgrounds.

Table 1: Fold changes of frequencies of 5-FOA^R colonies obtained between I-SceI induced and non-induced conditions.

I-*Sce*I TargetA				**I-*Sce*I TargetB**			
Strain	**Condition**	**5FOA^R freq**	**Fold Changes**	**Strain**	**Condition**	**5FOA^R freq**	**Fold Changes**
WT	YPD	1.77E-05	**1108**	**WT**	YPD	4.59E-04	**318**
	YPD+ATc	1.96E-02			YPD+ATc	1.46E-01	
pol32$^{-/-}$	YPD	1.46E-04	**11**	***pol32$^{-/-}$***	YPD	3.87E-04	**38**
	YPD+ATc	1.65E-03			YPD+ATc	1.46E-02	
msh2$^{-/-}$	YPD	1.17E-04	**261**	***msh2$^{-/-}$***	YPD	2.26E-04	**315**
	YPD+ATc	3.06E-02			YPD+ATc	7.12E-02	
spo11$^{-/-}$	YPD	1.67E-05	**1468**	***spo11$^{-/-}$***	YPD	6.40E-04	**261**
	YPD+ATc	2.45E-02			YPD+ATc	1.67E-01	
mph1$^{-/-}$	YPD	9.16E-05	**425**	***mph1$^{-/-}$***	YPD	7.73E-04	**105**
	YPD+ATc	3.90E-02			YPD+ATc	8.10E-02	
mus81$^{-/-}$	YPD	2.76E-04	**46**	***mus81$^{-/-}$***	YPD	4.17E-03	**29**
	YPD+ATc	1.27E-02			YPD+ATc	1.19E-01	

Functionality of the inducible I-*Sce*I cleavage system was also assessed by flow cytometry, focusing this time on DNA-DSB repair resulting in long-tract LOH. In Figure 4, we compared the fold changes in terms of increase in mono-fluorescent cells in all strains, and observed an augmentation in mono-fluorescent cells upon I-*Sce*I induction in all strains, ranging from 1.5- to 75.2-fold. We observe that deletion of the gene of interest reduces the frequency of mono-fluorescent cells in comparison to the WT, except *spo11$^{-/-}$* mutants which display fold changes that are close to the WT. Appearance of long-tract LOH upon I-*Sce*I induction is decreased in *pol32$^{-/-}$*, *mus81$^{-/-}$*, *mph1$^{-/-}$* and *msh2$^{-/-}$* strains, in decreasing order of severity as compared to WT (Figure 4), suggesting a potential shift in the molecular mechanisms used to mend the induced DNA-DSB.

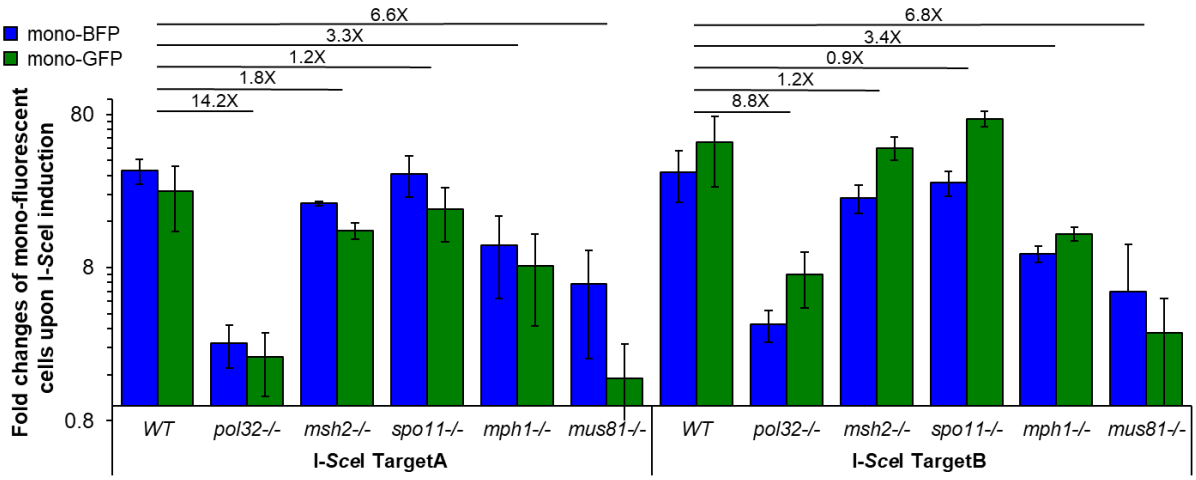

Figure 4: Quantification of I-SceI-induced long-tract LOH events in WT and KO strains.
Average fold changes of mono-fluorescent cells (+/- SD) between I-*Sce*I induced and non-induced conditions (n=4). Histogram representing the average fold changes (n=4) (+/- SD). Ratio between each mutant strain and appropriate WT strain is indicated above the histogram. Example of FACS profile in Figure S6.

- *Analysis of repair mechanisms*

Characterization of repair mechanisms was conducted in induced conditions, where DNA-DSB are predominantly I-*Sce*I-dependent. We took advantage of the auxotrophic markers associated with each component of the BFP/GFP LOH reporter system and with the I-*Sce*I target sequence, combined with the SNP-RFLP allele typing method to evaluate the size of LOH events, as summarized in Figure S2. The size of I-*Sce*I-dependent LOH events in each mutant was assessed by analyzing a total of 58 to 62 5-FOA^R colonies (2 strains x 26 to 32 colonies each), using cytometry and spotting on supplemented media to check fluorescence status and auxotrophy status, respectively. As illustrated in Figure 5A, the majority of colonies (≥80%), which underwent I-*Sce*I-dependent DNA-DSB, are characterized by short-tract LOH, likely resulting from gene conversion (GC). Thus, similarly to WT strain, I-*Sce*I-dependent DNA-DSB in the deletion mutants are mainly repaired through GC, a molecular mechanism which limits the LOH to the DNA-DSB locus. We then focused on characterizing the molecular mechanisms associated with I-*Sce*I-dependent long-tract LOH events. To do so, we defined two populations of mono-fluorescent cells for characterization after I-*Sce*I-dependent DNA cleavage: the “expected” population which includes cells that have lost the fluorescence marker carried by the Chr4 homolog targeted by I-*Sce*I and the “unexpected” population which includes cells which have become mono-fluorescent but retain the fluorescence marker carried by the Chr4 homolog targeted by I-*Sce*I; each population being representative of distinct molecular mechanisms at stake in I-SceI-dependent long tract LOH. To specifically assess the impact of each gene of interest on the molecular mechanisms leading to I-*Sce*I-dependent long-tract LOH, we performed fluorescence-activated cell sorting on both expected and unexpected mono-fluorescent populations of WT and deletion mutants

and recovered 70 to 80 colonies per population (35-40 colonies/mono-fluorescent population, for two strains), that were further characterized by flow cytometry and SNP-typing (Figure S2). We found that LOH events giving rise to the “unexpected” population are all the result of GC with CO in both the WT and mutant strains Figure 5B. For the “expected” populations, we observed three profiles: (i) the *spo11*$^{-/-}$ mutants that behave like the WT with 80% of the long-tract LOH resulting from BIR/MCO/CT and 20% from chromosome loss, (ii) the *mph1*$^{-/-}$ and *msh2*$^{-/-}$ mutants for which we observed a 2.5- to 5-fold decrease respectively in the frequency of long-tract LOH resulting from chromosome loss and (iii) the *pol32*$^{-/-}$ and *mus81*$^{-/-}$ mutants for which we reported a 2- to 3.5-fold increase in chromosome loss Figure 5B.Overall, apart from *SPO11*, deletion of the four other candidate genes altered the distribution of the investigated molecular mechanisms upon I-*Sce*I-mediated DNA-DSB.

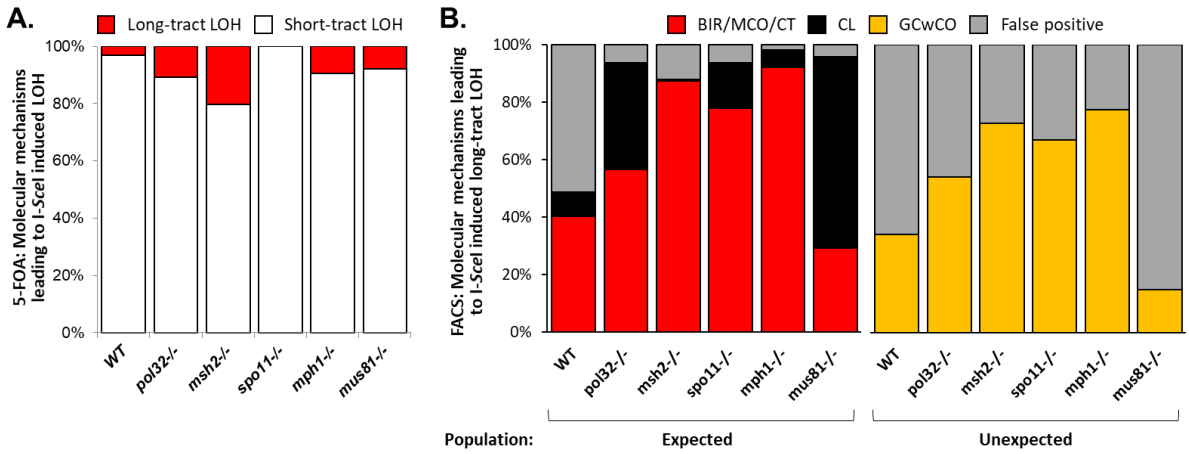

Figure 5: Analysis of molecular mechanism used in I-SceI induced conditions.
A. Percentage of molecular mechanisms leading to LOH among 5-FOA^{R} colonies in I-*Sce*I induced condition (n= 58-62). Long-tract LOH span from the I-*Sce*I target site until the BFP/GFP LOH reporter system while short-tract LOH are restricted to the I-*Sce*I break site. **B.** Characterization of mono-fluorescent-sorted colonies within expected and unexpected populations in I-*Sce*I induced condition (n=70-80). A subset of sorted cells was characterised by flow cytometry, auxotrophy spotting and SNP-typing to profile the homozygosis tract on the left arm of Chr4 and determine the molecular mechanism used to for I-*Sce*I-induced DNA-DSB. Abbreviations of molecular mechanisms are as follows; gene conversion (GC), break-induced replication (BIR), mitotic crossover (MCO), gene conversion with crossover (GC with CO).Colonies were regrouped in four categories: BIR/MCO/CT (I-*Sce*I-dependant LOH spanning from the I-*Sce*I break site until the Chr4L telomere), CL (I-*Sce*I-dependant LOH spanning across entire Chr4), GC w CO (I-*Sce*I-dependant LOH spanning from the I-*Sce*I break site until the Chr4L telomere) and False positive (I-*Sce*I-independent event or mono-fluorescence associated to a point mutation). For detailed analysis of molecular mechanisms please refer to Figure S2.

DNA repair mechanisms involved in events of homozygosis between the I-*Sce*I break site and the centromere

As mentioned above, we reported the existence of long-tract homozygosis centromere-proximal to the I-*Sce*I-induced double-strand break in the WT. We further investigated whether the absence of the candidate genes altered this phenomenon. To do so, we sequenced the genomes of nine to ten isolates having undergone I-*Sce*I-dependent BIR/MCO, for each deletion background and the WT (5 isolates of each haplotype). Whole-genome sequencing data allowed us to confirm the event of BIR/MCO on Chr4L using ABHet plots (Figure S5, Table S1). We first took advantage of the data to investigate the occurrence of gross chromosomal rearrangement (GCR) events occurring within the 59 *C. albicans* sequenced strains. Excluding the long-tract LOH event on Chr4L, the heat map in Figure 6 shows the distribution of GCRs (LOH and aneuploidy events) amongst the 8 chromosomes and the different genetic backgrounds. Overall, the *pol32*$^{-/-}$ and *mus81*$^{-/-}$ mutants possess the most genomic rearrangements with 5 and 18 GCRs within the ten *pol32*$^{-/-}$ and the nine *mus81*$^{-/-}$ sequenced strains, respectively (Figure 6). In contrast, WT and *msh2*$^{-/-}$ genetic backgrounds display no additional GCRs other than those induced by I-*Sce*I on Chr4L (Figure 6). Interestingly, from a chromosome perspective, Chr5 seems to be the most impacted by GCRs in this data set, exhibiting LOH and aneuploidy (3n or 4n) events in equal proportion (Table S2). Also, in the *mus81*$^{-/-}$ genetic background, Chr4 appears to be highly impacted by aneuploidies (3n or 4n) (Figure 6, Table S2).

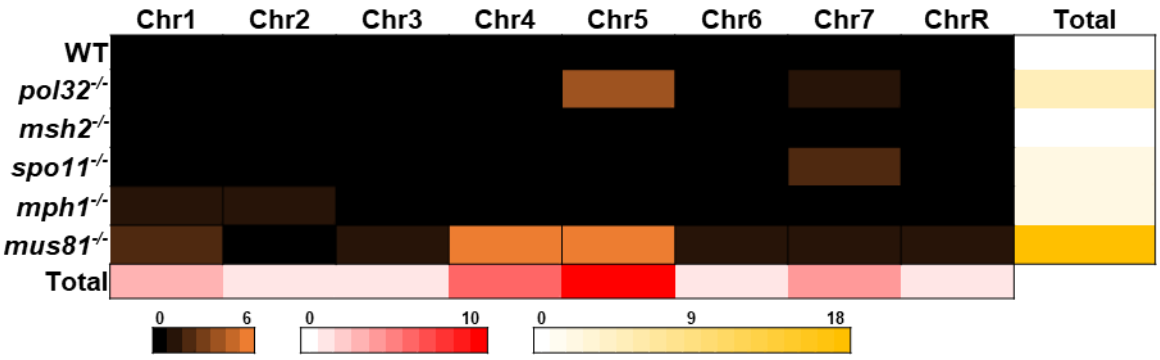

Figure 6: Distribution of genome-wide gross chromosomal rearrangements in wild-type and deletion strains.
Heatmap of gross chromosome rearrangements (GCRs) identified within the 59 *C. albicans* sequenced isolates, showing the density of GCRs per chromosome (n=59, red color bar) and per genetic background (n=9-10 strains, yellow color bar). Summary on GCRs identified within sequenced strains can be found in Table S2.

We then conducted an in-depth analysis of the 400 kb region surrounding the I-*Sce*I-break site as described above for the WT strains. By investigating the status of 2,845 naturally heterozygous positions on Chr4L between positions 650,000 and 1,050,000, we precisely identified which positions have become homozygous upon I-*Sce*I-cleavage in each sequenced isolate. As illustrated in Figure 7, 33% (16/49) of the mutant sequenced isolates having undergone an I-*Sce*I-dependent BIR/MCO upon cleavage of Chr4 (HapA or HapB) displayed homozygosis which extended almost exactly from the I-*Sce*I- break site towards the telomere (from position 778,104 to the telomere). While, the remaining 67% (33/49) of mutant isolates present an extension of the homozygous region towards the centromere (Figure 7). Of the latter isolates, two (I-*Sce*I TargetB *mus81*$^{-/-}$ isolates 3 and 4) possess homozygosis tracts which extend until the position 1,044,280, beyond *CEN*4 (located at position 992,473-996,110), and two others (in blue: I-*Sce*I TargetA *spo11*$^{-/-}$ isolate 2 and I-*Sce*I TargetA *mus81*$^{-/-}$ isolate 2) display homozygosis until the *RB*2-4 repeat sequence (pink) (Figure 7). Indeed, homozygosis tracts of these 4 isolates probably results from spontaneous LOH events. The other 29 events of homozygosis ascending towards the centromere in mutant isolates range up to 201 kb away from the I-*Sce*I-break site (Figure 7). Additionally, within the analyzed region of interest, we observed an alternating pattern between homozygosis of the HapA (green) and HapB (blue) shared by all strains (Figure 7). The latter is most likely due to haplotyping errors found within the SC5314 reference genome and not associated with intrinsic features of BIR/MCO molecular mechanisms. However, 13/59 strains display homozygosis tracts with interspersed regions on heterozygosis (identified by an asterisk in Figure 7), suggesting events of template switching during BIR, independently of genetic background. Regions of heterozygosis within homozygosis tracts are also observed at repeat regions such at *RB*2-4 locus (highlighted in pink on Figure 7), this heterozygosis is most certainly artefactual and associated to the misassembly of repeat regions in the reference genome used to align sequencing reads. Repeat regions are often difficult to accurately assemble due to their repetitive nature. Additionally, I-*Sce*I + TargetB *msh2*$^{-/-}$ isolate 1 displays initially homozygosis of HapB (blue) followed by HapA (green) homozygosis until the telomere, this can potentially be explained by two independent events or the occurrence of a spontaneous GC with CO event (Figure 7).

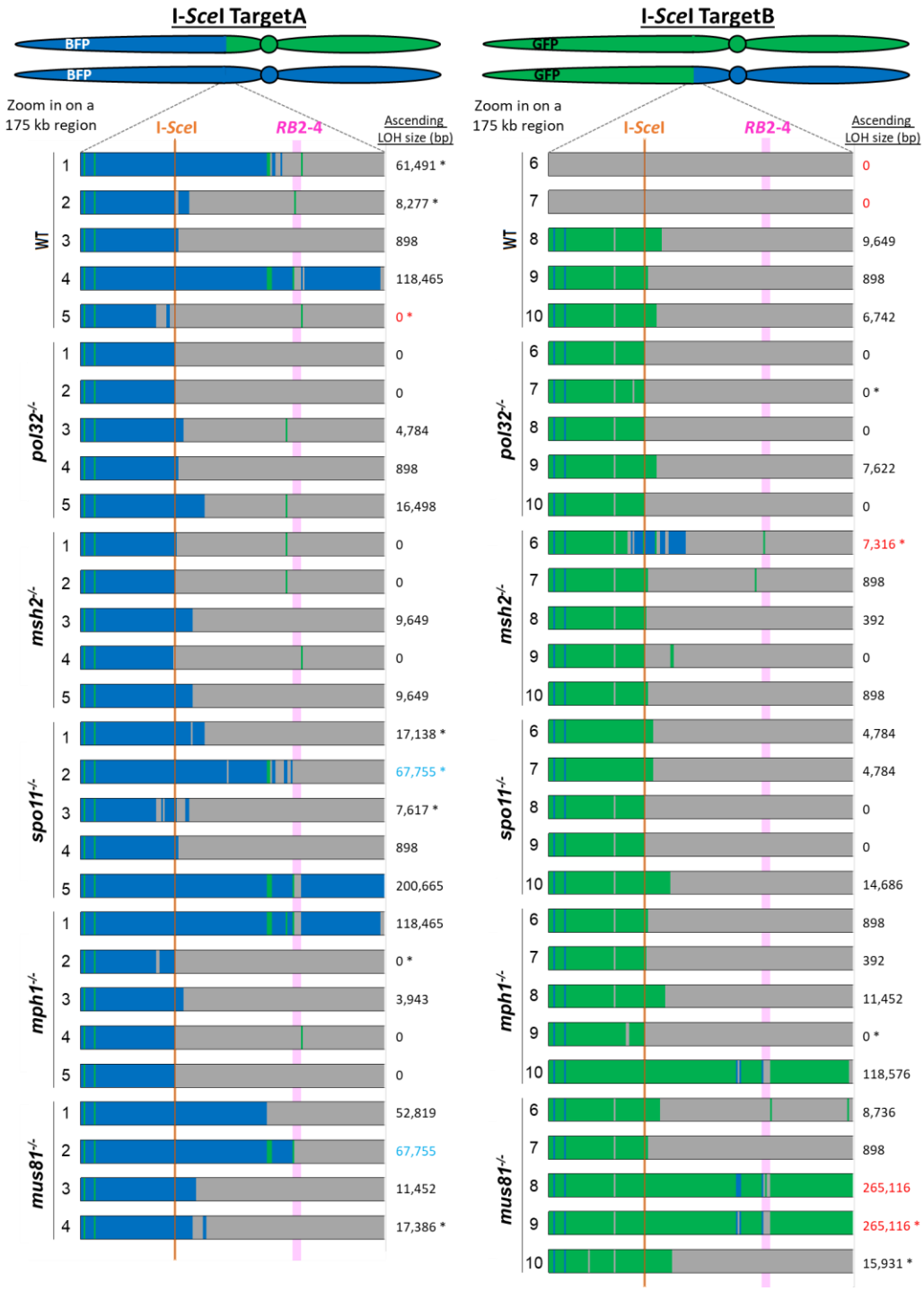

Figure 7: Results of sequencing data analysis at DNA-DSB site in cells having undergone BIR/MCO.

Each diagram represents an individual sequenced isolate (n=4-5) having undergone BIR/MCO, as determined by flow cytometry and SNP typing, following I-*Sce*I induction targeting DNA-DSB on haplotype A (I-*Sce*I TargetA) or haplotype B (I-*Sce*I TargetB) in the corresponding genetic backgrounds. Representation of a 175 kb region surrounding the I-*Sce*I-break site (orange) was built by monitoring the homozygous/heterozygous status using 1 kb bin sizes. Numerous unexpected homozygosis regions located between the I-*Sce*I-break site and the centromere are identifiable within these diagrams, with the length of the homozygosis centromere-proximal to the DNA-DSB indicated on the right side of each diagram in bp. Labelled in light blue writings are isolates where homozygosis starts in the RB2-4 repeat sequences (represented in pink), labelled in red writings are isolates having undergone non-RB2-4 associated spontaneous LOH and by an asterisk (*): isolates displaying interspersed regions of heterozygosis within long homozygosis tracts.

We then proceeded to quantify and compare the length of homozygosis tracts observed between the I-*Sce*I-break site and the centromere between the different genetic backgrounds. We precisely identified the start of the long LOH-tract on Chr4L as the position of the most centromere proximal heterozygous SNP which became homozygous, followed by two other newly homozygous positions within a >300 bp genomic region. Thus, the difference between the position of the homozygosis tract start and the position of the I-*Sce*I-break site on Chr4L gives the length of the region of homozygosis ascending towards the centromere (indicated in Figure 7), which can consequently be compared between genetic backgrounds. In order to accurately compare the effect of various deletion backgrounds on homozygosis tracts observed between the I-*Sce*I-break site and the centromere, we chose to remove all I-*Sce*I-independent LOH events on Chr4L (8 strains, indicated in blue and red on Figure 7). Among the remaining 51 isolates, the majority, 88% (45/51), possess ascending LOH tracts ranging from 0 to 17.4 kb while 12% of isolates (6/51) display > 50 kb long ascending homozygosis towards the centromere (Figure 8A). When comparing the median ascending LOH size across the six genetic backgrounds, a reduction in median length is observed within the *pol32*$^{-/-}$ (0 kb), *msh2*$^{-/-}$ (0.4 kb) and *mph1*$^{-/-}$ (0.7 kb) deletion strains as compared to WT (8.3 kb) (Figure 8A). While, deletions of *SPO*11 or *MUS*81 genes do not appear to have a negative impact on median length of homozygosis tracts between the I-*Sce*I-break site and the centromere. The same trends can be observed upon individual analysis of each targeted haplotype (I-*Sce*I TargetA or I-*Sce*I TargetB) (Figure 8B).

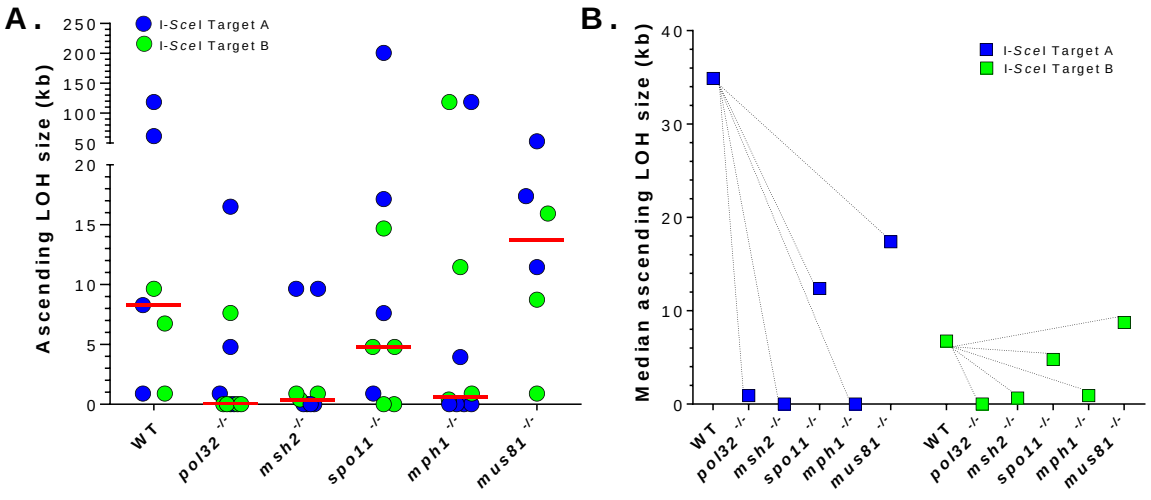

Figure 8: Comparison of ascending LOH-tracts towards CEN4 across genetic backgrounds.
A. Plotting of the length of homozygosis observed between the I-*Sce*I-break site and the centromere (*CEN*4), also referred to as ascending LOH, in the context of I-*Sce*I-dependent BIR/MCO events (blue: DNA-DSB on haplotype A, green: DNA-DSB on haplotype B) in the WT and mutant strains. The red line indicates the median ascending LOH size for each genetic background. **B.** Plot of median ascending LOH sizes for each genetic background and both I-*Sce*I-targeted haplotypes (haplotype A – TargetA or haplotype B – TargetB).

Discussion

By taking advantage of the previously developed inducible I-*Sce*I-mediated DNA-DSB system coupled to the BFP/GFP LOH reporter system (Feri *et al.* 2016), we assessed the fidelity of homology-directed repair mechanisms in *C. albicans* leading to long-tract LOH events, focusing on the BIR/MCO mechanisms. To do so, we performed whole-genome sequencing to precisely analyze the region surrounding the I-*Sce*I-break site on Chr4L in cells which displayed I-*Sce*I-dependent BIR/MCO profiles, as determined by flow cytometry and SNP typing. This analysis revealed that BIR/MCO is associated with homozygosis tracts that often extend beyond the I-*Sce*I target site towards the *CEN4* centromere and extend as expected on the other side to the left arm telomere of Chr4. We then demonstrated that deletion of *POL32*, *MPH1* or *MSH2* genes, reduced the median ascending LOH size towards the centromere, suggesting that the overflow of homozygosis towards the centromere constitutes intrinsic features of the BIR/MCO mechanisms and also involves the MMR machinery in *C. albicans* upon directed I-*Sce*I-mediated DNA-DSB.

In the human pathogenic yeast *C. albicans*, LOH are ubiquitously observed throughout the genomes of both laboratory and clinical isolates (Abbey *et al.* 2011; Ropars *et al.* 2018; Wang *et al.* 2018) and are frequently detected upon exposure to various stress conditions (Forche *et al.* 2011). A major source of LOH are DNA-DSBs repaired by HR-mediated molecular mechanisms leading to various LOH sizes (Malkova *et al.* 2000).Most of the knowledge of these mechanisms is based on fundamental studies conducted in *S. cerevisiae*. Although LOH are an important aspect of *C. albicans* biology, as illustrated by their association with the fixation of beneficial alleles leading to adaptive phenotypes (Coste *et al.* 2007; Tso *et al.* 2018; Liang and Bennett 2019), the precise characterization of genes and pathways involved in DNA repair in *C. albicans* remains superficial and “phenotype” oriented, when the deciphering of molecular mechanisms underlying such an important feature of *C. albicans* genome is crucial for the biological relevance of LOH to be fully understood. Studies have suggested that HR-mediated DNA-DSB repair is preferred to NHEJ in *C. albicans* (Feri *et al.* 2016; Vyas *et al.* 2018).

In order to investigate the fidelity of homology-directed repair, notably BIR/MCO, we characterized the effect of deletion mutation for five genes (*POL32*, *MSH2*, *SPO11*, *MPH1* and *MUS81*) on DNA-DSB repair, of which three (*POL32*, *MPH1* and *MUS81*) had not been previously investigated in *C. albicans.* Our characterization of mutant strains includes (i) a general phenotypic characterization of the mutants (in terms of growth rate, cell/colony morphology, filamentation and sensitivity to genotoxic stresses), (ii) the evaluation of

spontaneous genomic instability and (iii) the characterization of molecular mechanisms at stake upon I-*Sce*I-mediated DNA-DSB, in order to conclude on any striking genetic rewiring between *S. cerevisiae* and *C. albicans*.

The *POL32* gene is a non-essential accessory component of the polymerase δ protein complex required for BIR initiation (Huang *et al.* 2002; Makarova and Burgers 2015). Phenotypic characterization of homozygous deletion of *POL32* in *C. albicans* revealed a cell morphology phenotype, reminiscent of genotoxic-stress induced filamentation (Shi *et al.* 2007), an increased sensitivity to MMS and only a slight increase in LOH frequency (Figure 3). Similarly, the *pol32*$^{-/-}$ deletion mutants exhibited a reduced frequency of 5-FOA^{R} colonies and of cells displaying long-tract LOH (assessed by flow-cytometry) upon I-*Sce*I-dependent DNA-DSB (Table 1, Figure 5B), suggesting that the disruption of this accessory component of the Pol δ complex annihilates the efficient repair of DNA-DSB in *C. albicans.* Although *POL32* has been shown to be dispensable for GC in *S. cerevisiae* (Lydeard *et al.* 2007), it might not be the case in *C. albicans*. However, we currently cannot refute the notion that I-*Sce*I cleaving activity could also be less efficient in *POL32* deletion background as compared to WT. Viability tests are ongoing to check this hypothesis. Nevertheless, short-tract LOH are predominant upon I-*Sce*I-dependent DNA-DSB in all strains including *pol32*$^{-/-}$ strains (Figure 5A). Also, the characterization of long-tract LOH revealed an increased abundance of chromosome loss (Figure 5B), suggesting that *POL32* is more specifically implicated in BIR in *C. albicans*, similarly to observations made in *S. cerevisiae* (Lydeard *et al.* 2007). Interestingly, in the study by Lydeard *et al.*, the authors reported *POL32* as being essential for *RAD51*-dependent BIR (Lydeard *et al.* 2007). Indeed, Malkova *et al.* had previously reported *RAD51*-dependent and *RAD51*-independent BIR processes in *S. cerevisiae* (Malkova *et al.* 2001, 2005). The two processes differ in terms of site of initiation with the *RAD51*-independent BIR that initiates at a site that can be far from the DSB site (30 kb closer to the centromere), predominantly making use of a facilitator of BIR (FBI) sequence located in the vicinity of the origin site (Malkova *et al.* 2001), while the *RAD51*-dependent BIR does not depend on a FBI sequence to initiate repair and therefore begins within a few kb of the DSB (Malkova *et al.* 2005). Therefore, one would expect the *RAD51*-independent BIR process to be the only mechanism functional in *pol32*$^{-/-}$ deletion mutants, with long-tract LOH events that should rely on FBIs and initiate far from the I-*Sce*I-induced DSB. However, we observed the opposite with an enrichment for long-tract LOH that initiate within a few kb of the I-*Sce*I-induced DSB. These observations suggest differences between *S. cerevisiae* and *C. albicans* in terms of *RAD51*/*POL32* interactions or properties of the *RAD51*-independent BIR.

In *S. cerevisiae*, the *MUS81* gene is best known for its nuclease activities important for resolution of replication and recombination intermediates, notably resolution of Holliday junctions and for promoting CO events (Ho *et al.* 2010). In *C. albicans*, the *mus81*$^{-/-}$ strains displayed the most drastic phenotypes, as illustrated by their reduced growth, their shriveled colony morphology that correlates with the presence of numerous pseudohyphal cells and a high sensitivity to several genotoxic compounds (Figure S3, Figure 3A-B). As mentioned above, the formation of (pseudo)hyphal cells has been reported in *C. albicans* as a consequence of cell cycle disruption upon genotoxic stress(Shi *et al.* 2007). More recently, *S. cerevisiae MUS81* was shown to participate in cell cycle regulation (G2 to M transition) (Pfander and Matos 2017). Our observations suggest that, similarly to S*. cerevisiae*, *MUS81* could play a role in cell cycle progression in *C. albicans*. Additionally, whole-genome sequencing of *C. albicans mus81*$^{-/-}$ isolates highlighted numerous GCRs, notably aneuploidy events (Figure 6, Table S2). Increased rates of aneuploidy have also been observed in mouse and human *mus81*$^{-/-}$ cells (McPherson *et al.* 2004; Pamidi *et al.* 2007). *MUS81* is likely to be crucial to remove recombination and replication intermediates that link sister chromatids. In cells lacking *MUS81*, these intermediates would perdure, preventing proper chromosome disjunction and resulting in aneuploidy. The importance of *MUS81* in HR-mediated DNA repair is illustrated by the increase in the overall spontaneous genome instability in strains lacking *MUS81* (Figure 3C). Also, similarly to the *pol32*$^{-/-}$ mutants, we observed a reduced frequency of 5-FOA^{R} colonies and of cells displaying long-tract LOH (assessed by flow-cytometry) upon I-*Sce*I-dependent DNA-DSB (Table 1, Figure 5B) in the *mus81*$^{-/-}$ mutants, suggesting that even the GC process is impaired. In *Schizosaccharomyces pombe*, *mus81* mutants have normal frequencies of GC but reduced frequencies of CO (Smith *et al.* 2003), thus GC and CO can be genetically separated. This might not be the case in *C. albicans* as both GC and CO seem to be reduced in *mus81*$^{-/-}$ mutants as illustrated by the decrease in GC with CO events (Figure 5B).

We examined the impact of a third gene involved in HR-mediated repair, *MPH1*, which encodes a helicase which promotes D-loop displacement and NCO outcomes during HR-mediated DNA repair (Prakash *et al.* 2009). Our results suggest that a similar role is undertaken by *MPH1* in *C. albicans*, as its deletion enhances spontaneous genomic instability (Figure 3B) and resulted in an augmentation in GC with CO as shown by characterization of I-*Sce*I-dependent long-tract LOH events (Figure 5B). Additionally, similarly to observations made in *S. cerevisiae* where *mph1*$^{-/-}$ strains favor BIR (Mehta *et al.* 2017), our characterization illustrates an increase in BIR events within I-*Sce*I-dependent long-tract LOH events (Figure 5B). The anti-CO function of *MPH1* was shown to require the activity of the MMR component MutSα (Tay *et al.* 2010). Thus, unsurprisingly we observed an increase in GC with CO within

our *msh2*$^{-/-}$ genetic background (Figure 5B). The *MSH2* knock-out phenotype is the same as previously reported in *C. albicans* by Legrand *et al.* (Legrand *et al.* 2007). However, we described a slightly increased genome instability and observed a contrasting sensitivity to MMS between our two strains (Figure 3B). Defects in the MMR pathway lead to mutation accumulation in cells (Drotschmann *et al.* 1999; Shcherbakova and Kunkel 1999; Goellner 2020). Thus, we suspect that the *msh2*$^{-/-}$ I-*Sce*I Target A strain underwent a mutation rendering it sensitive to MMS as it did not demonstrate any other obvious discrepancies with the *msh2*$^{-/-}$ I-*Sce*I Target B strain, e.g. discrepancies between *msh2*$^{-/-}$ strains were not observed in latter analysis of the proportion of I-*Sce*I-dependent molecular mechanisms.

Lastly, we also investigated the effects of deletion of the *SPO11* recombinase and observed a WT-like phenotype in all aspects. This was somewhat expected, as *C. albicans SPO11* was recently shown to play an active recombinogenic role only in tetraploid cells (Anderson *et al.* 2019). Overall, our thoroughly phenotypic characterization highlighted new leads worth of further investigation to better understand the subtle differences between *S. cerevisiae* and *C. albicans* in terms of maintenance of genome integrity.

We then studied the importance of each of these five genes on ascending LOH tracts towards the centromere upon DNA-DSB repair. We initially observed that, as expected, homozygosis would span from the DNA-DSB site towards the telomere but would also ascend up to 118 kb from the DNA-DSB break site towards the centromere in isolates having undergone BIR/MCO as a consequence of I-*Sce*I-dependent DNA-DSB (Figure 2). This observation implies that the effect of LOH is not only restricted to the alleles distal to the DSB, as commonly assumed, but could also impact alleles proximal to the DSB, and this over hundreds kilobases. We sought out to understand which molecular mechanisms could be involved in these unexpected homozygous regions by investigating precisely the homozygosis tracts situated around I-*Sce*I-break sites of multiple isolates that displayed I-*Sce*I-dependent BIR/MCO profiles (Figure S2) from each previously described deletion background (Figure 7). Our data demonstrated that null mutants of either *pol32*$^{-/-}$, *msh2*$^{-/-}$ or *mph1*$^{-/-}$, displayed shorter median ascending LOH tracts towards the centromere as compared to the wild type genetic background, independently of which haplotype was targeted by I-*Sce*I nuclease (Figure 8). Deletion of *POL*32 in *S. cerevisiae* is often used to inhibit BIR as it was described to be essential for BIR and dispensable for GC (Lydeard *et al.* 2007). The fact that we observe shorter ascending LOH tracts upon *POL*32 deletion implies that these overflows of LOH towards the centromere are associated with the BIR molecular mechanism as inhibition of the BIR process reduces their appearance (Figure 8). BIR has been described in detail in *S. cerevisiae,* and shown to require extensive 5' resection as the majority of BIR events required

> 2.5 kb of 5’ resection, of which 50% of these events showed ≥ 15.5 kb resection stretches and a few events even spanned more than 27.5 kb (Chung *et al.* 2010). Our observations suggest that extensive 5’ resection is also a feature of BIR in *C. albicans* and thus could explain why homozygosis tracts often do not start exactly in the close vicinity of DNA-DSB and overflow towards the centromere. Indeed, extensive 5’ resection leading to a long 3’ single-strand DNA render more efficient the homology search in heterozygous organisms. We also described a reduction in ascending LOH size within the *MSH2* deletion background suggesting that the MMR machinery may also be implicated. The MMR pathway plays an important role by augmenting replication fidelity and genome stability through correction of base mismatch errors. The MMR machinery is spatially and temporally coupled to DNA replication and is active behind proofreading replicative DNA polymerases Polδ, resulting in a reduction in polymerase activity-associated mutations (Drotschmann *et al.* 1999; Harfe and Jinks-Robertson 2000; Healey *et al.* 2016; Boyce *et al.* 2017). The MMR machinery is also recruited along the DNA replication machinery at the site of DNA-DSB repaired by BIR. In these 5’ resection regions, MMR may “correct” natural heterozygous positions resulting in homozygosis between the DNA-DSB site and the centromere. Lastly, we observed that the median ascending LOH size was also reduced in *mph1*$^{-/-}$ genetic background (Figure 8A). This could be explained by the stabilization of the D-loop upon loss of heteroduplex DNA dissolvase Mph1. Thus, we hypothesis that either less 5’ resection could be needed to stabilize BIR, which is consistent with the finding of Mehta *et al.* showing that BIR is favored upon *MPH*1 deletion (Mehta *et al.* 2017), or that *MPH*1 deletion could favors crossovers or MCO. With our current system, we cannot differentiate BIR from MCO events although BIR has been described as a highly inaccurate process leading to > 1,000 times more indels than during normal replication in *S. cerevisiae* (Deem *et al.* 2011). However, each one of the above hypothesis regarding reduction of median ascending LOH size cannot explain the >20kb ascending events, thus we cannot refute that these latter events occurred spontaneously and are not associated with I-*Sce*I-dependent DNA-DSB repair. Of interest, upon removal of the >20kb ascending LOH events, we still observe an effect of *pol32*$^{-/-}$, *msh2*$^{-/-}$ and *mph1*$^{-/-}$ deletion on the median LOH size.

Overall, BIR in *C. albicans* is frequently associated to extended regions of homozygosis between the DNA-DSB locus and the centromere. We also propose that the MMR pathway could act upon natural heterozygous positions during DNA-DSB repair by BIR and participate in the overflow of homozygosis towards the centromere. For the first time, we describe that long-tract homozygosis start sites do not perfectly coincide directly with DNA break sites in *C. albicans.* This notion is striking in this pathogenic yeast as heterozygosity levels are positively correlated with strain fitness, while here we show that LOH through BIR

is a messy process in *C. albicans* which potentially impacts dozens of additional alleles proximal to the centromere.

Material and methods

Strains and culturing conditions

C. albicans strains described in the study are derived from the reference strains SC5314, parental strain CEC4088 and CEC4012(Leu$^-$) (Feri *et al.* 2016) possessing the BFP/GFP LOH reporter system (positions 471,021-481,176) and the I-*Sce*I target-sequence (positions 776,698-778,104) on the left arm of Chr4. Yeast cells were cultured on/in rich YPD medium (1% yeast extract, 2% peptone, 2% dextrose). Synthetic Defined (SD) (0.67% yeast nitrogen base without amino acids, 2% dextrose) and Synthetic Complete (SC) (0.67% yeast nitrogen base without amino acids, 2% dextrose, 0.08% drop-out mix with all the essential amino-acids) media were used for selection. A yeast carbon base medium with bovine serum albumin (YCB-BSA) and supplemented with amino-acids was used to recycle *LEU2* auxotrophic marker durinf strain construction. Solid media were obtained by adding 2% agar.

All *C. albicans* strains are listed in Table S3.

Induction of Tet-On system

The I-*Sce*I gene is found under the control of the Tet-On promoter. Activation of I-SceI protein production and induction of DNA DSB at targeted site, is conducted by inoculating YPD + anhydrotetracycline (ATc) (3 µg/mL) (Thermofisher ACROS Organics™) with a preculture (liquid SC-His-Arg medium overnight at 30°C). Induction is performed for 8 h at 30°C followed by an overnight recovery in YPD.

Cell sorting

Large debris and filamentous cells that could obstruct the tubing system of the cytometer were filtered using BD Falcon™ Cell strainers from induced and non-induced cultures. The MoFlo® Astrios™ flow cytometer was used to analyze and sort the cells of interest. For each sorted gate, 1,000 cells were recovered in 400 µL of liquid YPD medium, plated immediately after cell sorting on four YPD Petri plates and incubated at 30°C for 48h before collection of results.

Characterization of molecular mechanism used leading to LOH

Single colonies recovered from either cell sorting or 5-FOA plates were isolated and culture in YPD at 30°C overnight. Characterization of the molecular mechanism leading to LOH was conducted as described below for each single colony isolated. Functionality of auxotrophic

markers was evaluated by drop tests on SC medium with appropriate drop-out amino acid, depending on tested marker. Overnight saturated cultures in YPD of selected strains were spotted on YPD (control), SC-His (test presence of *BFP-HIS*1 cassette), SC-Arg (test presence of *GFP-ARG*4 cassette) and SC-Ura (test presence of *URA*3-TS cassette) and placed at 30°C for 24h to observe presence and absence of growth. The fluorescence status of everyone was also assessed using flow cytometry (MACSQuant® Analyzer (Miltenyi Biotec)), where 10 000 cells were analysed per culture and identified as mono-BFP, mono-GFP or biofluorescent (BFP and GFP). An aliquot of each culture was used to extract DNA permitting genotyping of Chr4 and the accurate identification of each LOH size. The gDNA was extracted using the EPICENTER MasterPure™ Yeast DNA Purification Kit followed by DNA elution in 100 µL of sterile water. Allele composition at two SNP positions, SNP156 (left arm of Chr4) and SNP95 (right arm of Chr4) was examined using the SNP-RFLP technique as described in Forche *et al.* (Forche *et al.* 2009b), with restriction enzyme *Taq*I and *Alu*I, respectively. Theoretically, upon I-*Sce*I mediated DNA-DSB the *URA*3 marker associated to the I-*Sce*I target sequence is lost. Thus, we screen for presence of absence of *URA*3 by conducting PCR reactions using primers from Table S4. By taking in account all results obtained throughout the phenotypic and genotypic profiling, we are were able to confirm the presence of LOH and identify the size of each LOH tract, consequently the molecular mechanism resulting in LOH.

KO strains

- *Construction*

To investigate which molecular mechanism(s) is (are) responsible for extending the homozygosis from the DNA DSB locus towards the centromere, we investigated the effects of knocking-out 5 different genes (*POL*32, *MSH*2, *SPO*11, *MPH*1 and *MUS*81), key players in different DNA repair pathways. These KO strains were constructed in SC5314 derived, CEC4088 (Leu^-) and CEC4012 (Leu^-) possessing the I-*Sce*I TS on HapA and HapB, respectively. With the transient CRISPR-Cas9 system (Min *et al.* 2016), homozygous deletion mutants were obtained in one round of transformation using a recyclable leucine flipper cassettes. We constructed a FLP-*LEU*2 plasmid by switching out the *SAT*1 gene in the pSFS1A, pFLP-*SAT*1 plasmid (Reuss *et al.* 2004), with the *LEU*2 gene. We amplified the *LEU*2 gene from the pFA-Cm*LEU*2 plasmid (Schaub *et al.* 2006) using primers containing *Pvu*II and *Nsi*I restriction sites and ligated into the backbone of pSFS1, where removal of *SAT*1 was conducted by double digestion with *Blp*1 and *Pst*I followed by Klenow treatment. By recycling of the *LEU*2 auxotrophic marker, each strain was then complemented with a full length copy of *GPI*16, as Chr4L harbors a recessive lethal allele on HapB between the I-*Sce*I

TS and the BFP/GFP LOH reporter system, rescuing individuals which become homozygous for Chr4B (Feri *et al.* 2016).

We used a transient CRISPR-Cas9 system (Min *et al.* 2016), which does not necessitate the genomic integration of either Cas9 or sgRNAs. The construction of sgRNAs and amplification of Cas9 cassettes from the pV1093 plasmid were conducted as described in Min *et al.*. Five unique 20 bp sgRNAs (Table S4) were designed using CHOPCHOP (Montague *et al.* 2014), guiding Cas9 cleavage activity to each locus of each target genes (*POL*32, *MSH*2, *SPO*11, *MPH*1 and *MUS*81). Repair templates are constructed by PCR amplification using 120 bp primers (Table S4), each composed of 20 bp complementary to the FLP-*LEU*2 cassette and 100 bp tails possessing the complementary sequences of the KO gene locus. Each primer pair was used to amplify the FLP-*LEU*2 cassette from plasmid pFLP-*LEU*2. Each cassette was amplified in a total PCR volume of 500 µL, precipitated in 100% ethanol and re-suspended in 100 µL of distilled sterile water. For each transformation, competent cells were transformed with approximately 5 µg of appropriate DNA cassette. CEC4088 and CEC4012 cells were co-transformed with 5 µg of the appropriate FLP-*LEU*2 cassette, 1 µg of the Cas9 cassette and 1 µg of sgRNA using the Lithium Acetate/PEG transformation protocol. Transformants were then selected on SC-Leu medium and, junction PCRs and internal target gene PCRs were performed in order identify homozygous KO transformants and to ensure proper integration of FLP-*LEU*2 cassette at the targeted locus (using primers from Table S4). Selected homozygous KO transformants were culture overnight at 30°C in liquid YCB-BSA medium supplemented with 100 X amino acids (Arg10g/L His10g/L Uri 2g/L Leu 6g/L) and platted on SD-agar plates containing various concentrations of leucine (0.02 g/L, 0.012 g/L, 0.006 g/L, 0.0012 g/L and 0.0006 g/L) in order to select for colonies which have flipped-out the *LEU*2 marker, resulting in *LEU*2 loss. The latter strains underwent a second transformation round to integrate *Stu*I linearized CIp-P_{TDH3}-*LEU*2-*GPI*16 plasmid (Feri *et al.* 2016) at the *RPS*1 locus on Chr1, *GPI*16 complementation. Transformants were then selected on SD medium and junction PCRs were conducted to ensure proper integration at *RPS*1 locus, using primers in Table S4. All 10 constructed strains can be found in Table S3.

- *Characterization*

The characterization of the 10 KO strains (Table S3) consisted of morphology, doubling time, filamentation and genotoxic stress sensitivity assays.

The colony morphology of all strains was also assessed on solid YPD medium, at 30°C for 3 days. Images were taken with a Leica M80 Stereomicroscope at a x25 magnification. Images were captured with a DMC 2900 camera, using the Leica Application Suite (LAS) imaging

software. Cell phenotypes were observed using overnight culture in YPD at 30°C with an Olympus IX83 microscope, 20x objective. Filamentation phenotypes were observed after 3 - 4h in YPD, YPD + 10% serum and Spider medium broths and imaged with the Olympus IX83 microscope, 20x objective.

The doubling times were evaluated in YPD media at 30°C by measuring the optical density with TECAN Infinite over 24h.

Sensitivity of strains to various genotoxic stressors was evaluated using drop test assay. Cultures grown at 30°C overnight were spotted YPD, YPD 4mM H_2O_2, YPD 0.1 mM menadione, YPD 2mM tBHP, YPD 100 µM camptothecin, YPD 0.03% EMS and YPD 0.01% MMS. Images were taken after 2-3 days of incubation at 30°C with a PhenoBooth Colony Counter (Singer Instruments).

Quantification of LOH frequency using flow cytometry

All flow cytometry analyses were conducted on the MACSQuant® Analyzer (Miltenyi Biotec). Data for a maximum of 10^6 cells per sample were analyzed using the FlowJo V10.1 software. The gates to determine the LOH frequencies were arbitrarily selected but conserved throughout sample analysis.

5-Fluoroorotic acid (5-FOA) assay

Following the I-*Sce*I induction protocol, as seen above, three different cell dilutions of cultures (20,000 cells, 2,000 cells and 200 cells) grown in presence (induced) or absence (non-induced, control) of ATc were plated on 5-FOA-containing plates in triplicates. Dilutions were verified by plating a volume corresponding to 100 cells on YPD plates. Plates were incubated at 30°C for 3 days before analysis.

DNA extraction and whole genome sequencing

Wildtype heterozygous (I-*Sce*I TargetA or TargetB) strains along with 4-5 isolates having undergone I-*Sce*I-dependent BIR/MCO, based on flow cytometry and SNP-typing, for each deletion background (WT, *pol32*$^{-/-}$, *msh2*$^{-/-}$, *spo11*$^{-/-}$, *mph1*$^{-/-}$ and *mus81*$^{-/-}$) were cultured in 5 mL of liquid SD medium overnight at 30°C and DNA was extracted following the manufacturer's protocol using the QIAGEN QIAamp DNA Mini Kit. The DNA was eluted in a total volume of 100µL. The genomes were sent for whole-genome sequencing at Novogene, Illumina sequencing technology, NextSeq500 platforms were used to generate 150 bp paired-ends reads (sequencing data available upon request).

Analysis of DNA-DSB site

Sequences and genomic variations were analyzed as previously described in (Ropars *et al.* 2018; Sitterle *et al.* 2019). Each set of paired-end reads was mapped against the *C. albicans* reference genome, SC5314 haplotype A (version A22-s07-m01-r57), using Minimap2 version (Li 2018). SAMtools version 1.9 and Picard tools version 2.8.1 (http://broadinstitute.github.io/picard) were then used to filter, sort and convert SAM files. SNPs were called using Genome Analysis Toolkit version 3.6, according to the GATK Best Practices. SNPs were filtered using the following parameters: VariantFiltration, QD<2.0, LowQD, ReadPosRankSum<-8.0, LowRankSum, FS>60.0, HightFS, MQRankSum < -12.5, MQRankSum, MQ < 40.0, LowMQ, HaplotypeScore > 13.0. Coverages were also calculated using the Genome Analysis Toolkit. The GATK variant filtration walker (VariantAnnotator) was used to add allele balance information to VCF files. The value of allele balance at heterozygous sites (ABHet) is a number that varies between 0 and 1. ABHet is calculated as the number of reference reads from individuals with heterozygous genotypes divided by the total number of reads from such individuals.

In order to princely analyze I-*Sce*I break sites, heterozygous positions between 650,000-1,050,000 for Chr4 were extracted from the GATK SNPs calling files (filtered with best practice VCF, mentioned above) from sequencing data of heterozygous WT I-*Sce*I TargetA and I-*Sce*I TargetB strains. This generated two heterozygous positions reference lists, subsequently used to for comparison with each strain. For each sequenced isolate we also extracted all information from the GATK SNPs calling files (VCF) for genomic region between positions 650,000-1,050,000 on Chr4 and using a Python custom script (available upon request) we investigated the status of each position of interest within each sequenced isolate. Because we aligned read data onto haplotype A of SC5314 our interpretation for each strain was: for a given position, if the VCF file indicated (i) a defined heterozygous SNP we assigned the position as heterozygous, (ii) a homozygous SNP we considered that positions to be homozygous for haplotype B or (iii) if no information for this position was found in the VCF file it was assigned as a homozygous haplotype A SNP. The final tables which summarize the comparison results between (appropriate) WT heterozygous strain and each sequenced isolate were then used to precisely analyze the region of interest and evaluate the ascending LOH size by identifying the start of homozygosis tracts.

Acknowledgments

We are grateful to Pierre-Henri Commere for help with the cell sorting experiments.

T.M., A.F., M.L., and C.D. designed experiments. M.C. and T.M. performed experiments. T.M., M.C., C.M., M.L. and C.D. analyzed data. T.M., M.L., and C.D. wrote the manuscript.

T.M. is the recipient of a PhD. fellowship from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID) and Indo-French Centre for the Promotion of Advanced Research CEFIPRA (CEFIPRA). We acknowledge support from the French Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases [ANR10-LABX-62-IBEID]) and of the Indo-French Centre for the Promotion of Advanced Research CEFIPRA (CEFIPRA). A.F. was the recipient of a Ph.D. grant from INRA Jouy-en-Josas and Institut Pasteur.

Supplementary material

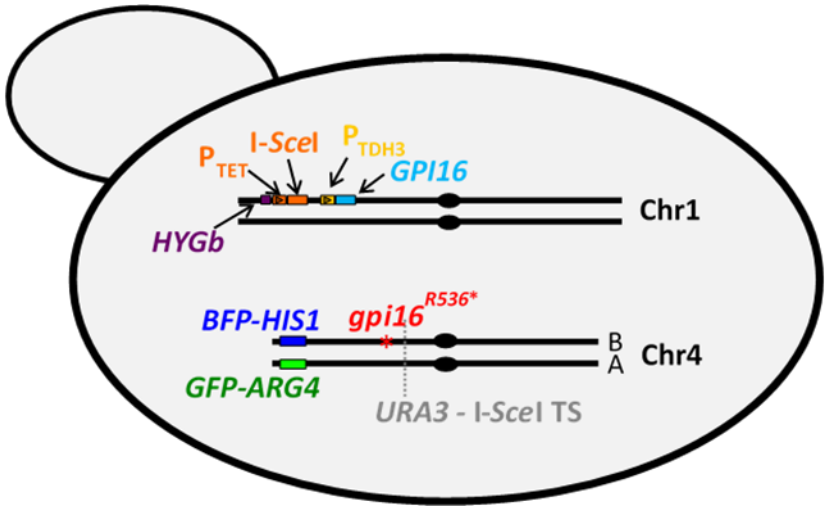

Figure S1: Coupling of a DNA double-strand-break-inducing system and a FACS-optimized LOH reporter system on Chr4.
Illustration of the strain bearing the BFP/GFP LOH reporter system (Loll-Krippleber *et al.* 2015) on the left arm of Chr4 with the inducible DNA-DSB system (Feri *et al.* 2016). DNA-DSB can be targeted on either haplotype, depending on the location of the I-*Sce*I- target sequence associated to the *URA3* auxotrophic marker. Upon addition of tetracycline (or other derivative molecules, including anhydrotetracycline) the tetracycline promoter (P_{TET}) is activated thus the coding sequence of the I-*Sce*I megaendonuclease is expressed (associated to the hygromycin resistance marker (*HYGb*), subsequently permitting to induce locus specific DNA-DSB at its target sequence. The BFP/GFP LOH reporter system is utilized to visualize long-tract LOH events which encompass this artificial heterozygous locus, composed of the BFP (associated to the histidine auxotrophy marker, *HIS1*) located on a homologues and the GFP (associated to the arginine auxotrophy marker, *ARG4*) on the second homologue. The strains constructed in this study derive from the SC5314 reference strain possessing a recessive deleterious allele, $gpi16^{R536*}$ located on Chr4B between the I-*Sce*I target sequence and the LOH reporter system (Feri *et al.* 2016). Such alleles have been previously shown to impact the directionality of LOH events thus, to overcome this limitation all strains used in this study possess an additional functional *GPI16* allele integrated at the *RPS1* locus on Chr1.

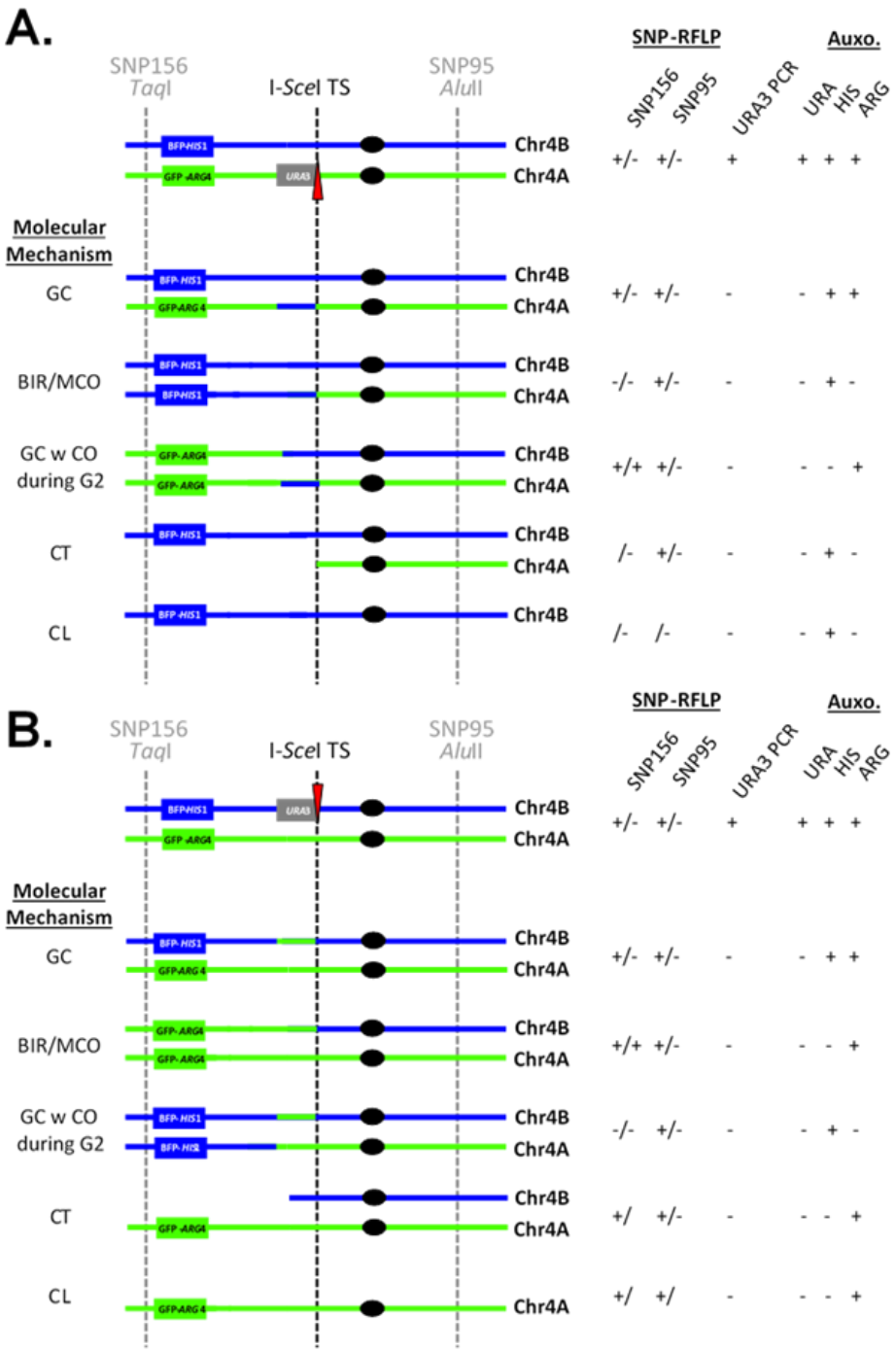

Figure S2: Characterization of monofluorescent-sorted individuals leading to the identification of a molecular mechanism.

Monofluorescence-sorted individuals underwent: (i) flow cytometry analysis to determine the fluorescence status (BFP^+/GFP^-, BFP^-/GFP^+ or BFP^+/GFP^+), (ii) SNP typing on both arms of Chr4 using the SNP-RFLP technique (SNP number from Forche *et al.* (Forche *et al.* 2009b) and restriction enzyme used), (iii) *URA3* internal PCR to determine if I-*Sce*I- target sequence was cleaved leading to loss of *URA3* auxotrophic marker and (iv) auxotrophy spot tests to evaluate the presence or absence of *URA3* (associated to the I-*Sce*I-target sequence), *HIS1* (associated to the BFP) and *ARG4* (associated to the GFP). These results were interpreted as illustrated in panels **A.** (analysis for strains possessing the I-*Sce*I- target sequence on haplotype A or Chr4L) and **B.** (analysis for strains possessing the I-*Sce*I-target sequence on haplotype B or Chr4L) in order to identify the length of the LOH tract and the appropriate molecular mechanism. Abbreviations of molecular mechanisms are as follows; gene conversion (GC), break-induced replication (BIR), mitotic crossover (MCO), gene conversion with crossover (GC with CO), chromosome truncation (CT), chromosome loss (CL).

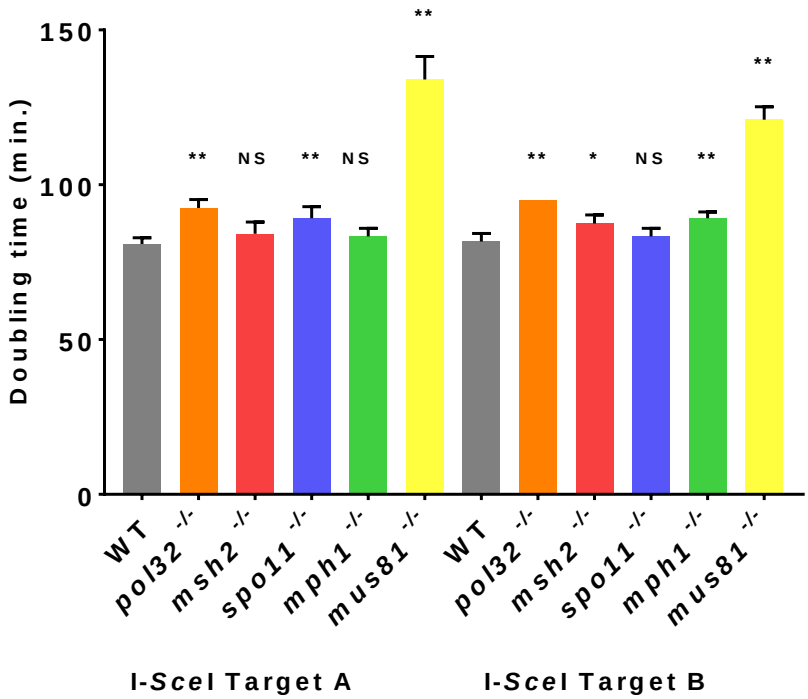

Figure S3: Doubling time of C. albicans strains.
Average doubling time (mins.) evaluated from optical density 24h kinetic at 30°C (n= 5-6), error bars represent standard deviation. The significance of doubling time differences between appropriate wild-type (WT) strain and deletion strains (*pol32*$^{-/-}$, *msh2*$^{-/-}$, *spo11*$^{-/-}$, *mph1*$^{-/-}$ and *mus81*$^{-/-}$) was evaluated using Mann-Whitney Test (NS: $p>0.05$, *: $p<0.05$ and **: $p<0.01$).

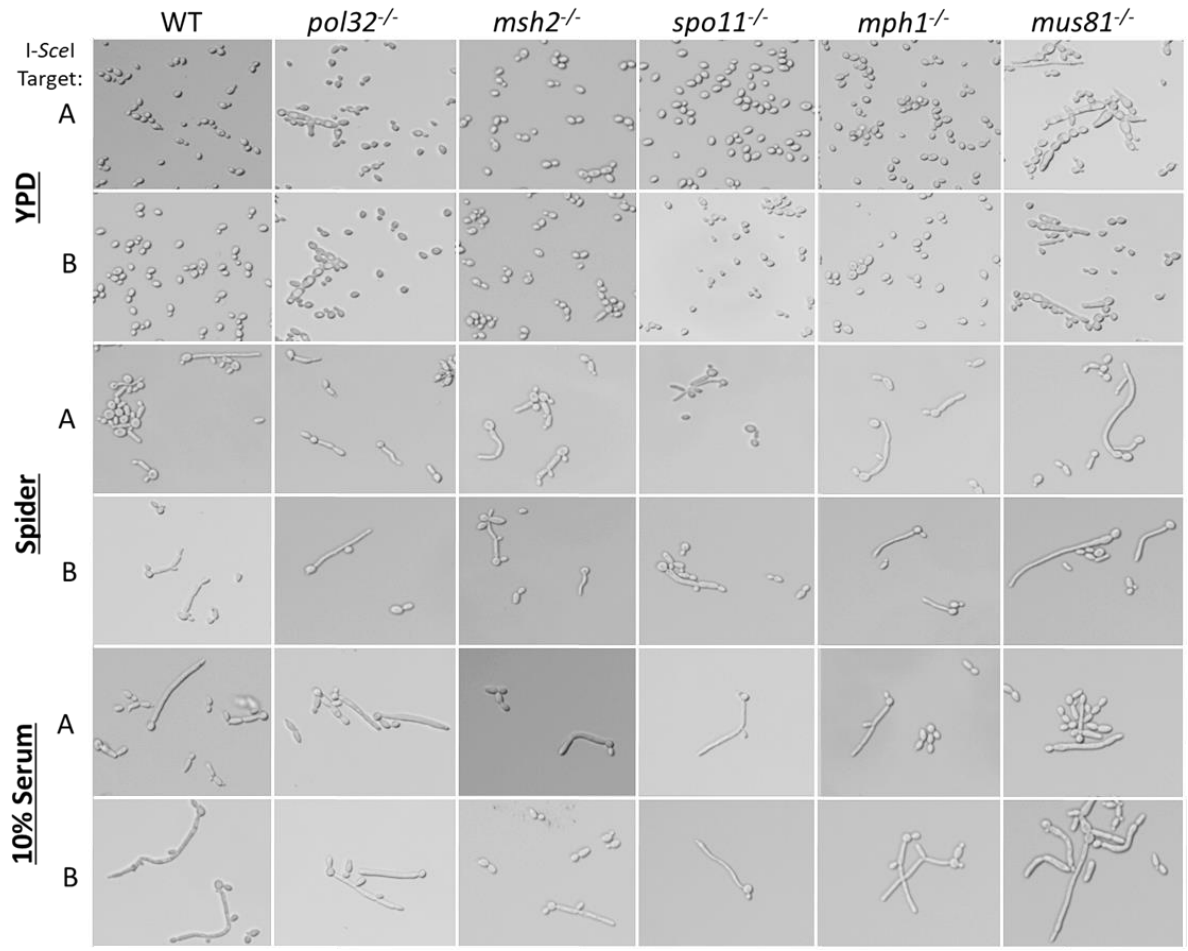

Figure S4: Filamentation assay.
Characterization of the filamentation capacity of different genetic backgrounds (*pol32$^{-/-}$, msh2$^{-/-}$, spo11$^{-/-}$, mph1$^{-/-}$ and mus81$^{-/-}$*) was observed after 3 - 4h in YPD, Spider medium and YPD + 10% serum broths and imaged with the Olympus IX83 microscope, 20x objective.

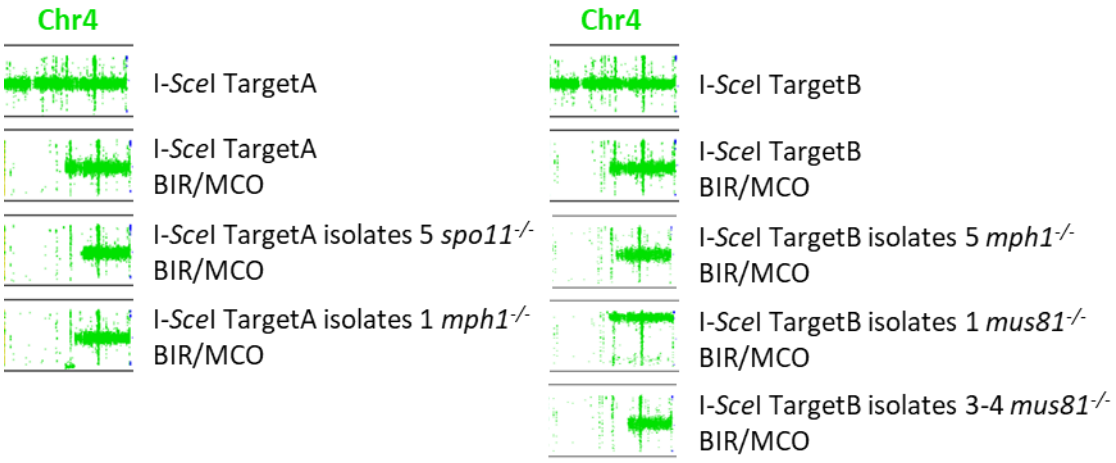

Figure S5: Allele balance at heterozygous site plots for Chr4 for all knock-out C. albicans strains sequenced in this study.
Plots showing the allele balance at heterozygous sites across chromosome 4. Sequencing confirms the presence of homozygosis, a long tract LOH, on the left arm of Chr4 (either haplotype A or B according to the strain) in mono-fluorescent isolates.

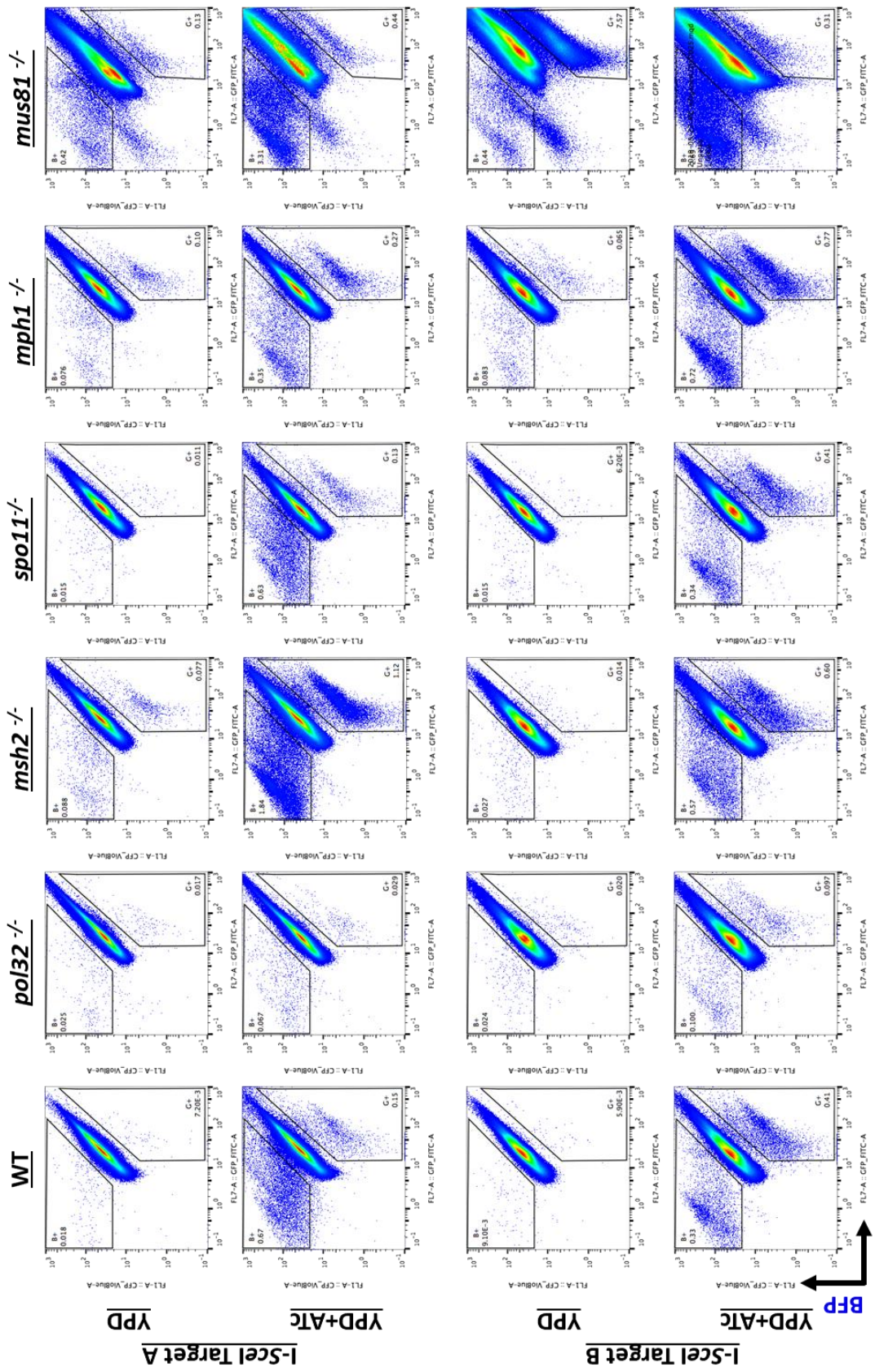

Figure S6: Flow cytometry profile illustrating genomic stability of wild type and knock-out strains.
Utilization of the BFP/GFP LOH reporter system located on Chr4L to assess genomic stability in standard growth conditions (YPD) and upon I-*Sce*I induction (YPD+ATc) leading to DNA-DSB on Chr4L. Each FACS profile represents a total of 10^6 analyzed cells.

Table S1: Sequencing depth on chromosome 4 of sequenced isolates displaying I-SceI-dependent BIR/MCO. *

Strain		Cov Chr4 by kb	Log2 Cov Chr4 by kb	SD Chr4 by kb	Log2 Cov Chr4 ROI by kb	SD Chr4 ROI by kb	Strain		Cov Chr4 by kb	Log2 Cov Chr4 by kb	SD Chr4 by kb	Log2 Cov Chr4 ROI by kb	SD Chr4 ROI by kb
Htz WT		167.96	0.88	0.20	0.88	0.17	**Htz WT**		208.58	0.90	0.21	0.90	0.15
WT CEC4430 I-*Sce*I HapB *+ GPI16*	**1**	135.01	0.82	0.39	0.83	0.19	**WT** CEC4429 I-*Sce*I HapB *+ GPI16*	**1**	214.01	0.88	0.38	0.84	0.34
	2	93.66	0.84	0.32	0.83	0.19		**2**	192.14	0.91	0.21	0.91	0.16
	3	112.71	0.84	0.41	0.85	0.17		**3**	188.11	0.89	0.24	0.91	0.20
	4	141.37	0.80	0.39	0.81	0.16		**4**	187.54	0.90	0.32	0.90	0.29
	5	105.57	0.83	0.47	0.85	0.19		**5**	181.70	0.89	0.26	0.90	0.22
***pol32*$^{-/-}$** CEC5503 I-*Sce*I HapB *+ GPI16*	**1**	151.36	0.76	0.40	0.77	0.19	***pol32*$^{-/-}$** CEC5508 I-*Sce*I HapB *+ GPI16*	**1**	174.50	0.88	0.26	0.89	0.23
	2	153.83	0.73	0.31	0.73	0.20		**2**	192.80	0.84	0.43	0.81	0.40
	3	172.30	0.84	0.41	0.84	0.17		**3**	173.48	0.86	0.41	0.83	0.37
	4	114.92	0.78	0.38	0.80	0.18		**4**	151.66	0.87	0.37	0.85	0.33
	5	156.89	0.80	0.32	0.81	0.18		**5**	163.14	0.88	0.31	0.88	0.27
***msh2*$^{-/-}$** CEC5504 I-*Sce*I HapB *+ GPI16*	**1**	154.75	0.82	0.36	0.82	0.18	***msh2*$^{-/-}$** CEC5509 I-*Sce*I HapB *+ GPI16*	**1**	205.84	0.91	0.46	0.88	0.34
	2	164.78	0.84	0.36	0.86	0.18		**2**	171.04	0.90	0.44	0.88	0.30
	3	155.97	0.83	0.29	0.84	0.17		**3**	173.38	0.93	0.25	0.93	0.17
	4	134.81	0.83	0.37	0.83	0.19		**4**	208.57	0.94	0.24	0.94	0.15
	5	135.87	0.81	0.43	0.83	0.17		**5**	167.43	0.93	0.23	0.94	0.15
***spo11*$^{-/-}$** CEC5505 I-*Sce*I HapB *+ GPI16*	**1**	177.55	0.84	0.46	0.85	0.18	***spo11*$^{-/-}$** CEC5510 I-*Sce*I HapB *+ GPI16*	**1**	168.04	0.91	0.20	0.91	0.16
	2	134.41	0.79	0.41	0.78	0.24		**2**	145.20	0.90	0.21	0.91	0.16
	3	177.26	0.80	0.34	0.79	0.22		**3**	132.49	0.92	0.21	0.93	0.16
	4	142.36	0.87	0.35	0.86	0.24		**4**	176.64	0.93	0.20	0.94	0.14
	5	140.48	0.83	0.25	0.84	0.18		**5**	179.99	0.91	0.21	0.93	0.14
***mph1*$^{-/-}$** CEC5506 I-*Sce*I HapB *+ GPI16*	**1**	170.47	0.85	0.44	0.88	0.15	***mph1*$^{-/-}$** CEC5511 I-*Sce*I HapB *+ GPI16*	**1**	130.18	0.90	0.27	0.90	0.14
	2	185.55	0.75	0.50	0.74	0.31		**2**	140.98	0.90	0.27	0.92	0.14
	3	164.13	0.86	0.34	0.88	0.17		**3**	129.43	0.90	0.27	0.92	0.14
	4	132.58	0.88	0.36	0.91	0.18		**4**	174.68	0.91	0.27	0.92	0.15
	5	138.56	0.88	0.44	0.91	0.20		**5**	126.66	0.91	0.27	0.94	0.20
***mus81*$^{-/-}$** CEC5507 I-*Sce*I HapB *+ GPI16*	**1**	197.12	0.93	0.46	0.90	0.31	***mus81*$^{-/-}$** CEC5512 I-*Sce*I HapB *+ GPI16*	**1**	152.94	1.06	0.21	1.07	0.18
	2	167.69	0.96	0.55	0.93	0.50		**2**	211.77	1.46	0.70	1.72	0.61
	3	176.91	0.82	0.46	0.83	0.23		**3**	128.41	0.85	0.19	0.86	0.18
	4	211.84	0.87	0.47	0.86	0.24		**4**	126.81	0.88	0.20	0.88	0.19
								5	134.97	1.06	0.22	1.07	0.16

* aneuploidies are written in red

Table S2: Summary of gross chromosomal rearrangements (GCR) within sequenced isolates displaying I-SceI-dependent BIR/MCO. *

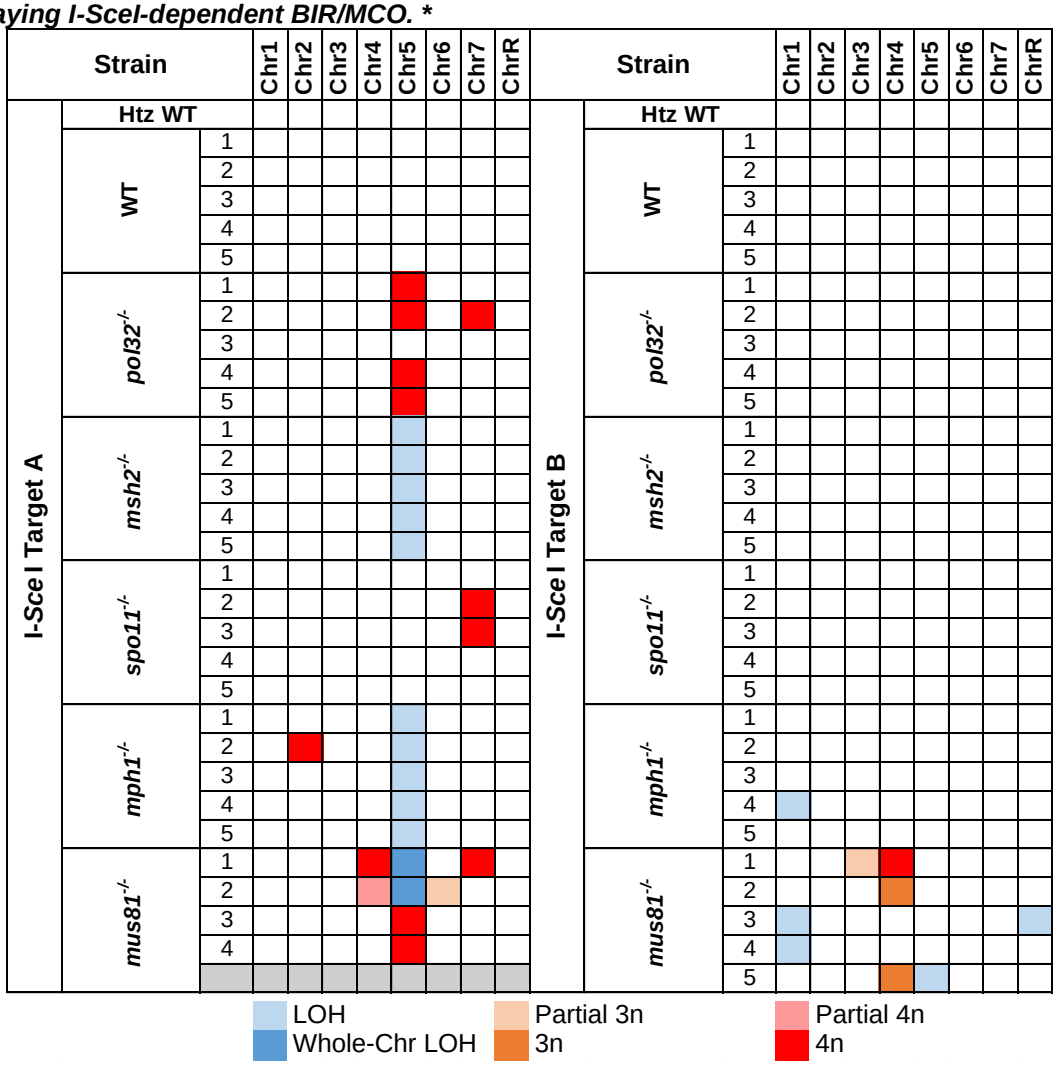

Table S3: Strains used in this study.

Strain			Genotype		Ref.
		CEC4088	SN148 Ca21chr4_C_albicans_SC5314:473390to 476401Δ::P_{TDH3}-*GFP*-*CaARG4*/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::P_{TDH3}-*BFP*-*CdHIS1 ADH1/adh1*::pNIMx Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::P_{TET}-I-*Sce*I-*HYGb*-XOG1HOL1/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::pFA-ISceI_TS-*URA3*-CDR3-tG(GCC)2/Ca21chr4_C_albicans_SC5314:775939 to 779223	ARG+ HIS+ leu-URI+ NTC^R HYG^R	(1)*
		CEC4012	SN148 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::P_{TDH3}-G*FP*-*CaARG4*/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::P_{TDH3}-*BFP*-*CdHIS1 ADH1/adh1*::pNIMxCa21chr1_C_albicans_SC5314:625304 to 626436Δ:: P_{TET}-I-*Sce*I-*HYGb*-XOG1HOL1/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::pFA-ISceI_TS-*URA3*-CDR3-tG(GCC)2/Ca21chr4_C_albicans_SC5314:775939 to 779223	ARG+ HIS+ leu-URI+ NTC^R HYG^R	(1)*
I-SceI Target A	**WT**	**CEC4429**	CEC4088 *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	(1)*
	***pol32*$^{-/-}$**	**CEC5503**	CEC4088 *pol32*Δ::FRT/*pol32*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***msh2*$^{-/-}$**	**CEC5504**	CEC4088 *msh2*Δ::FRT/*msh2*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***spo11*$^{-/-}$**	**CEC5505**	CEC4088 *spo11*Δ::FRT/*spo11*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***mph1*$^{-/-}$**	**CEC5506**	CEC4088 *mph1*Δ::FRT/*mph1*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***mus81*$^{-/-}$**	**CEC5507**	CEC4088 *mus81*Δ::FRT/*mus81*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
I-SceI Target B	**WT**	**CEC4430**	CEC4012 *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	(1)*
	***pol32*$^{-/-}$**	**CEC5508**	CEC4012 *pol32*Δ::FRT/*pol32*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***msh2*$^{-/-}$**	**CEC5509**	CEC4012 *msh2*Δ::FRT/*msh2*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***spo11*$^{-/-}$**	**CEC5510**	CEC4012 *spo11*Δ::FRT/*spo11*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***mph1*$^{-/-}$**	**CEC5511**	CEC4012 *mph1*Δ::FRT/*mph1*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
	***mus81*$^{-/-}$**	**CEC5512**	CEC4012 *mus81*Δ::FRT/*mus81*Δ::FRT *RPS1/RPS1*::P_{TDH3}-C4_03130W_A-*CdLEU2*	Prototroph NTC^R HYG^R	This study
E. coli		**ECC1449**	PV1093	Amp^R	(2)(3) *
		ECC1370	CIp-pTDH3-*LEU2*-*GPI*16	Amp^R	(1) *
		ECC337	pFA-CmLEU2	Amp^R	(5) *
		ECC98	pSFS1A	Cm^R	(4) *
		ECC1589	FLP-*LEU*2	Cm^R	This study

* (1) Feri *et al*. (2016), (2) Vyas *et al*. (2015), (3) Min *et al.* (2016), (4) Reuss *et al.* (2004), (5) Schaub *et al.* (2006)

Table S4: List of primers used throughout this study.

	Primer names	Sequence	Use
pFLP-LEU2	Leu2_F_PvuII	GGGAAACAGCTGAAGCTTCGTACGCTGC	Amplification of *LEU2* gene from pFA-CmLEU plasmid for pFLP-*LEU2* construction. In red, *Pvu*II restriction site.
	Leu2_R_NsiI	GGGAAAATGCATCATAGGCCACTAGTGG ATCTG	Amplification of *LEU2* gene from pFA-CmLEU plasmid for pFLP-*LEU2* construction. In red, *Nsi*I restriction site.
Construction of KO strains using transient CRISPR-Cas9	SNR52_F	AAGAAAGAAAGAAAACCAGGAGTGAA	Guide RNA construction (2)(3) *
	sgRNA_R	ACAAATATTTAAACTCGGGACCTGG	Guide RNA construction (2)(3) *
	SN52_N3	GCGGCCGCAAGTGATTAGACT	Guide RNA construction (2)(3) *
	sgRNA_N3	GCAGCTCAGTGATTAAGAGTAAAGATGG	Guide RNA construction (2)(3) *
	sgRNA_F_MSH2	AATAGCAGCAATAGATGAGCGTTTTAGAG CTAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
	SNR52_R_MSH2	GCTCATCTATTGCTGCTATTCAAATTAAA AATAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
	sgRNA_F_POL32	TACTTTCATCATACGTGTCCGTTTTAGAG CTAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
	SNR52_R_POL32	GGACACGTATGATGAAAGTACAAATTAAA AATAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
	sgRNA_F_MUS81	GCCACATCCATAAACCCTCGGTTTTAGAG CTAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
	SNR52_R_MUS81	CGAGGGTTTATGGATGTGGCCAAATTAA AAATAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
	sgRNA_F_SPO11	AATATTCATCCAACTCAGAAGTTTTAGAG CTAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
	SNR52_R_SPO11	TTCTGAGTTGGATGAATATTCAAATTAAA AATAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
	sgRNA_F_MPH1	GTTGCATTACCCACTGGTCTGTTTTAGAG CTAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
	SNR52_R_MPH1	AGACCAGTGGGTAATGCAACCAAATTAAA AATAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
	SPO11_KO-FLP_Fwd	AATCATATCATCAATTTTCTGAAAACGGG AAGACTAACTTTCTTTAAGAAACAAATTAC AGGTGCTCCTAAATATTTGGGAACAAGTT ACATCCTCATTATACGACTCACTATAGGG CG	Repair template for KO of *SPO*11 with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	SPO11_KO-FLP_Rev	AAACAGGGGCTTGTAAAAAAAATTTCCAC CACCAAAACATTTTGAAGAAAAAAAACTT TGTCGTGTAAGCAAATCAGAACACTCGA CTCACGTCACGTGCGCTCTAGAACTAGT GGATC	Repair template for KO of *SPO*11 with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	MUS81_KO-FLP_Fwd	ATTATACCGTTTTGTTTTGAAACACGCCT AAAAACTCGTCTGTTTACTGATTTTATTTC ACCTTCTTTATTCCATATATCATTAACTCA GTTACACACCATACGACTCACTATAGGG CGA	Repair template for KO of *MUS*81 with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	MUS81_KO-FLP_Rev	TCGTTTCACTAACTTAGAACAAACGATAC AAAATCTGTCCTCAAAAATATTTGTACATC TATACACTACCAAATAATTCTTGCAACCA ACTTTCGATTCTCGCTCTAGAACTAGTGG ATC	Repair template for KO of *MUS*81 with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	MPH1_KO-FLP_Fwd	TGATTGAGACATATTGAAGTATTGGAAGT ATCTCTGGTGTTGTATTGCCTGGTGTAAT TATTGAATAAACATACAAAACATAAGTCC GGTATTTTTTTCCTACGACTCACTATAGG GCGA	Repair template for KO of *MPH*1 with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	MPH1_KO-FLP_Rev	TATAGTCGCGAAGAAATATTTCATCACCA GTAATTTTGGTGCTCGCACTACTATTCTA TTCGTGTATTATGCATATATATAGATTTGG AATTAAATGTATCGCTCTAGAACTAGTGG ATC	Repair template for KO of *MPH*1 with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	POL32_KO-FLP_fwd	ATTGTCCACGCCAAAAATTGAAAATATAC AGAGAGAGAGAAAAAAAAGGATGTCAGA ATATCATTACTTTTTTCATTCAATTTTATA GGCGATAATGACGTACGACTCACTATAG GGCGA	Repair template for KO of *POL32* with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	POL32_KO-FLP_rev	AATTTAGTTTTGTCTGTCATTATTATGTAG TTAGGATCCAATGTAATAGTTTTTAGCCC CATCAAAAAAAGTATTTACTCGCGTAATT	Repair template for KO of *POL32* with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid

		TTTTGTTTGTTCCGCTCTAGAACTAGTGGATC	
	MSH2_KO-FLP_fwd	AGGTGCACGCAAAGTTTTTTTTTTAAACTCTCCCCCCTTATTTATTCTCCAGACGCGTCTTTTATCCAAACACCCTTCTCTTTCTCTTACCTCACCCACCTACGACTCACTATAGGGCGA	Repair template for KO of *MSH2* with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	MSH2_KO-FLP_rev	TGGTTCGTGTGTGACACAATTAATTTTATTTTGTTTATCTCGGAGATTTACCTCCGGAATGAAACCTTCTCTCTCTCTCAACGGATTACTGTCTACCATTCGCTCTAGAACTAGTGGATC	Repair template for KO of *MSH2* with *FLP-LEU2*. Underlined: Sequence complementary to *FLP-LEU2* containing plasmid
	CaCas9_F	ATCTCATTAGATTTGGAACTTGTGGGTT	Cas9 amplification (2)(3) *
	CaCas9_R	TTCGAGCGTCCCAAAACCTTCT	Cas9 amplification (2)(3) *
Transformant verification	SPO11_fwd_ver	CAGGTCATTTCTCCCAACAG	Junction PCR of *FLP-LEU2* cassettes
	SPO11_rev_ver	GGCGTATAATGGAATATCGTGG	Junction PCR of *FLP-LEU2* cassettes
	MUS81_fwd_ver	AGTTGATGGTCTGACAGTGG	Junction PCR of *FLP-LEU2* cassettes
	MUS81_rev_ver	TTCGTATTCAGTGGGTGGAC	Junction PCR of *FLP-LEU2* cassettes
	MPH1_fwd_ver	GTTGGCTATACCAAGATGCTG	Junction PCR of *FLP-LEU2* cassettes
	MPH1_rev_ver	GACGAATCCTCACTGATACAC	Junction PCR of *FLP-LEU2* cassettes
	POL32_fwd_ver	GAAGAGCTTTCGGTCAACTG	Junction PCR of *FLP-LEU2* cassettes
	POL32_rev_ver	CCGAGTCACTACAGAACAAAC	Junction PCR of *FLP-LEU2* cassettes
	MSH2_fwd_ver	AGCTCTTCTCTGGTCTTCTC	Junction PCR of *FLP-LEU2* cassettes
	MSH2_rev_ver	TCCGTTGAGAGAGAGAGAAG	Junction PCR of *FLP-LEU2* cassettes
	SAP2p_rev	TTACCCGTTTCACTTTCTGC	Junction PCR of *FLP-LEU2* cassettes
	LEU2_fwd	CGGTTCTGCTCCTGATTTAC	Junction PCR of *FLP-LEU2* cassettes
	SPO11_fwd_int	ACCGAGACTAGAAGGTTTGC	Homozygous KO vs. Heterozygous KO Verification
	SPO11_rev_int	GCTGCGATTCAAGCAAGTAG	Homozygous KO vs. Heterozygous KO Verification
	MUS81_fwd_int	GCCATTGGGTTATCATCTCC	Homozygous KO vs. Heterozygous KO Verification
	MUS81_rev_int	GTTTGGGTTAAGGAGGACTG	Homozygous KO vs. Heterozygous KO Verification
	MPH1_fwd_int	GTCGAAATCGAGCAGAGATC	Homozygous KO vs. Heterozygous KO Verification
	MPH1_rev_int	CCAACCATGCCCAATAACTG	Homozygous KO vs. Heterozygous KO Verification
	POL32_-intF	ATCTCCTGTGAAGCCAGAAC	Homozygous KO vs. Heterozygous KO Verification
	POL32_intR	CCCGTTGAAGGTTTAGTGAC	Homozygous KO vs. Heterozygous KO Verification
	MSH2_intF	GTGGTTAGTGCAGCCAAATC	Homozygous KO vs. Heterozygous KO Verification
	MSH2_rev_int	GTCAAACAGGGTATCAAGGTC	Homozygous KO vs. Heterozygous KO Verification
Colony typing	SNP95.F-New	CATGCCCGCTTGAAACTACC	Digestion par *Alu*I à 37°C
	SNP95.R-New	GTCAGTGATTCAGTTGAAGTGG	Digestion par *Alu*I à 37°C
	SNP156.F-New	ACAGAACAGTAGATTCCAAC	Digestion par *Taq*I à 65°C
	SNP156.R-New	TGGGTTTGGACATCAGGTTCAA	Digestion par *Taq*I à 65°C
	URA3_rev_AF	GTTGTCCTAATCCATCACCT	*URA3* PCR
	URA3_fwd_AF	AACTCATGCCTCACCAGTAG	*URA3* PCR

***** (2) Vyas *et al.* (2015), (3) Min *et al.* (2016)

Chapter 4: The use of CRISPR-Cas9 to target homologous recombination limits transformation-induced genomic changes in *Candida albicans*

Contains: Article published in mSphere (02/09/2020)

The use of CRISPR-Cas9 to target homologous recombination limits transformation-induced genomic changes in *Candida albicans*

Timea MARTON[1,2], Corinne MAUFRAIS[1,3], Christophe D'ENFERT[1], Melanie LEGRAND[1#]

[1] Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, France
[2] Université Paris Diderot, Sorbonne Paris Cité, Paris, France
[3] Hub de Bioinformatique et Biostatistique, Département de Biologie Computationnelle, USR 3756 IP CNRS, Institut Pasteur, Paris, France

Running title: Transformation-induced genome rearrangements in *Candida albicans*

Abstract

Most of our knowledge relating to molecular mechanisms of human fungal pathogenesis in *Candida albicans* relies on reverse genetics approaches, requiring strain engineering. DNA-mediated transformation of *C. albicans* has been described as highly mutagenic potentially accentuated by its genome plasticity, including the acquisition of genomic rearrangements notably upon exposure to stress. The advent of CRISPR-Cas9 has vastly accelerated the process of genetically modifying strains especially in diploid, such as *C. albicans*, and polyploid organisms. The effects of unleashing this nuclease within the genome of *C. albicans* are unknown, though several studies in other organisms report Cas9-associated toxicity and off-target DNA breaks. Upon the construction of a *C. albicans* strain collection, we took the opportunity to compare strains which were constructed using CRISPR-Cas9-free and CRISPR-Cas9-dependent transformation strategies, by quantifying and describing transformation-induced loss-of-heterozygosity and hyperploidy events. Our analysis of 57 strains highlights the mutagenic effects of transformation in *C. albicans*, regardless of the transformation protocol, but also underscores interesting differences in terms of genomic changes between strains obtained using different transformation protocols. Indeed, although strains constructed using the CRISPR-Cas9-free transformation method display numerous concomitant genomic changes randomly distributed throughout their genome, the use of CRISPR-Cas9 leads to a reduced overall number of genome changes, particularly hyperploidies. Overall, in addition to facilitating strain construction by reducing the number of transformation steps, the CRISPR-Cas9-dependent transformation strategy in *C. albicans* appears to limit transformation-associated genome changes.

Importance

Genome editing is essential to nearly all research studies aimed at gaining insights in the molecular mechanisms underlying various biological processes, including in the opportunistic pathogen *Candida albicans*. The adaptation of the CRISPR-Cas9 system greatly facilitates genome engineering in many organisms. However, our understanding of the effects of CRISPR-Cas9 technology on the biology of *C. albicans* is unknown. In this study, we sought out to compare the extent of transformation-induced genomic changes within strains engineered using CRISPR-Cas9-free and CRISPR-Cas9-dependent transformation methods. CRISPR-Cas9-dependent transformation allows to simultaneously target both homologs and, importantly, appears less mutagenic in *C. albicans*, as strains engineered using CRISPR-Cas9 display an overall decrease in concomitant genomic changes.

Introduction

Candida albicans, often found as a commensal yeast of the human gastro-intestinal tract, is also a notorious clinical fungal pathogen. It ranks as the 4th leading cause of nosocomial infections, principally due to the increasing number of immune-compromised individuals (Brown *et al.* 2012). Nowadays, *C. albicans* is arguably the pathogenic yeast that is used to the largest extent in order to investigate at the molecular level fungal pathogenesis. As a consequence, many molecular resources have been generated to facilitate the study of *C. albicans* biology, not restricted to but including gene knockout and gene-overexpression strain collections (Homann *et al.* 2009; Legrand *et al.* 2018; Schoeters *et al.* 2018), all implying *C. albicans* genome editing. Generation of gene knock-out or gene fusions with reporter genes has principally relied on homologous recombination at target loci using PCR-generated cassettes with 70 to 100 nucleotides of homology to the targeted genomic locus and a lithium acetate-heat shock protocol (Gola *et al.* 2003) or electroporation protocol (De Backer *et al.* 1999). These strategies allow single allele disruption, which can quickly become cumbersome in a diploid organism such as *C. albicans* since multiple transformation rounds are required for construction of null mutants. An inconvenience that has also emerged over time is the occurrence of non-specific genome rearrangements in the course of transformation, as several studies reported transformation-associated genome changes, which sometimes impact the phenotypic traits of the engineered strains (Sanglard *et al.* 1997; Chen *et al.* 2004; Selmecki *et al.* 2005; Dunkel and Morschhäuser 2011; Ciudad *et al.* 2016). The occurrence of loss-of-heterozygosity (LOH) events and aneuploidies have been previously described, and can occur on chromosomes other than those targeted by genetic manipulations (Hughes *et al.* 2000; Bouchonville *et al.* 2009; Arbour *et al.* 2009; Abbey *et al.* 2011). Interestingly, while short-time exposure of *C. albicans* to heat during the transformation heat shock process tends to increase aneuploidy events, longer-time exposure to a milder heat favors appearance of LOH events (Bouchonville *et al.* 2009). Previous work has also shown that aneuploid strains are particularly susceptible to undergo genomic changes upon transformation (Bouchonville *et al.* 2009). As a safeguard, researchers have been advised to test the ploidy of their mutants at various genome locations using tools such as the qPCR ploidy screen (Arbour *et al.* 2009).

The discovery and adaptation of the CRISPR-Cas9 system has revolutionized genome editing in a multitude of organisms. Initially described in *Streptococcus pyogenes*, the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) and its endonuclease Cas9 act as a bacterial defense system against non-self DNA. A single guide RNA (sgRNA) directs the cleaving activity of Cas9 to a specific genomic site. This sgRNA is composed of a recognition motif of 20 bp directly followed by a protospacer-adjacent motif sequence (PAM),

composed of the nucleotides NGG (Jinek *et al.* 2012). Bacterial protection is achieved by recognition and induction of DNA double-strand break (DSB) in the non-self DNA (Barrangou *et al.* 2007). The CRISPR-Cas9 technology is the latest of several customizable DNA-binding nucleases that have been engineered to facilitate the introduction of DNA fragments at a specific genomic location thanks to homologous recombination (HR) upon DNA repair of the induced DNA-DSB (Wood *et al.* 2011). Its ability to target DSB during genetic manipulations stimulates HR-mediated recombination at a specific locus, therefore enhancing transformation efficiency as compared to classical transformation approaches.

The CRISPR-Cas9 technology has been successfully implemented in mammalian cells, plants, bacteria and fungi, including *C. albicans,* and continues to be adapted to an increasing number of organisms. In diploid organisms, such as *C. albicans,* this technique has facilitated the construction of null mutants, where both alleles may be modified at once (Vyas *et al.* 2015). CRISPR-Cas9 has also been successfully used to generate multiple simultaneous knockouts using a unique sgRNA to target multiple genes in *C. albicans* (personal communication from S. Bachellier-Bassi). However, due to its novelty, the repercussions of the active CRISPR-Cas9 system on the biology of *C. albicans* are not yet fully understood. In *Saccharomyces cerevisiae*, the integration and constitutive expression of Cas9 has been associated with cell toxicity, lowered fitness of strains as illustrated by a slower growth rate. Additionally, toxicity of constitutive Cas9 expression has also been associated with a lower transformant yield (DiCarlo *et al.* 2013). To limit these effects, a transient system has been developed in *C. albicans* circumventing constitutive Cas9 and sgRNA expression (Min *et al.* 2016). Another issue regards the report of off-target DSBs by Cas9, in mammalian cells (Fu *et al.* 2013; Tsai *et al.* 2015; Banakar *et al.* 2019). Overall, our understanding of the effects of CRISPR-Cas9 technology in *C. albicans* is limited, especially in terms of its impact of the integrity of the genome.

We had initially set out to build a collection of isogenic strains possessing an LOH reporter system allowing detection of LOH events throughout the genome of *C. albicans*. This system involves the insertion of two fluorescent marker genes in a neutral genomic region on both homologs of a given chromosome, allowing spontaneous LOH events to be detected by monitoring the loss of one of the fluorescent markers using flow cytometry. While validating our collection of transformed strains, whole genome sequencing revealed the presence of genome changes. Here, we present a descriptive study of the genome changes in our collection of *C. albicans* transformed strains, comparing strains constructed using CRISPR-Cas9-free and -dependent methods. With a total of 57 sequenced strains, our results illustrate

and highlight the important mutagenic properties of each transformation protocol both in terms of quantification and nature of those genome changes.

Overall, the democratization of whole genome sequencing allowed us to address the extent of “unwanted” genome changes during strain construction in *C. albicans* as we show that the CRISPR-Cas9-free transformation method was a high inducer of genome changes with often multiple genome change events within a single constructed strain as compared to the CRISPR-Cas9-dependent transformation method, which results in less frequent concomitant genome changes.

Results

Strain construction and phenotyping

We have previously established a LOH reporter system for *C. albicans* whereby the *GFP* gene is integrated at a given locus on one chromosome and the *BFP* gene is integrated on the homologous chromosome at the same locus (Loll-Krippleber *et al.* 2015). LOH at this locus leads to the loss of either the *BFP* or *GFP* gene and the frequency at which LOH events arise can be quantified using FACS analysis. In an attempt to explore variation in LOH frequency on the 8 *C. albicans* chromosomes, we undertook to construct a collection of isogenic strains, each one possessing the BFP/GFP LOH reporter system at a distinct genomic locus (Table S1). In addition to the LOH reporter system, a unique barcode sequence was also integrated in each strain in order to permit pool experiments. All strains derive from SN148, a *C. albicans* laboratory reference strain displaying arginine, histidine, uridine and leucine auxotrophies. This strain was sequentially transformed in order to integrate (i) the BFP/GFP LOH reporter system, (ii) a unique barcode sequence and (iii) the leucine marker rendering the strains prototroph. CRISPR-Cas9-free and CRISPR-Cas9-dependent transformation methods were used to generate a total 57 strains, 30 and 27 strains respectively (Figure 1). Both methods are based on homology-directed recombination with exogenous DNA, prepared using PCR and directed to the target genomic locus by the two flanking homology regions (100 bp). Additionally, both CRISPR-Cas9-free and -dependent transformation protocols use lithium acetate/PEG and heat-shock treatments to transform the yeast cells. Transformation by electroporation was not considered since it was shown to be associated with frequent ectopic integrations of free DNA (De Backer *et al.* 1999; Bachellier-Bassi and d’Enfert 2015).

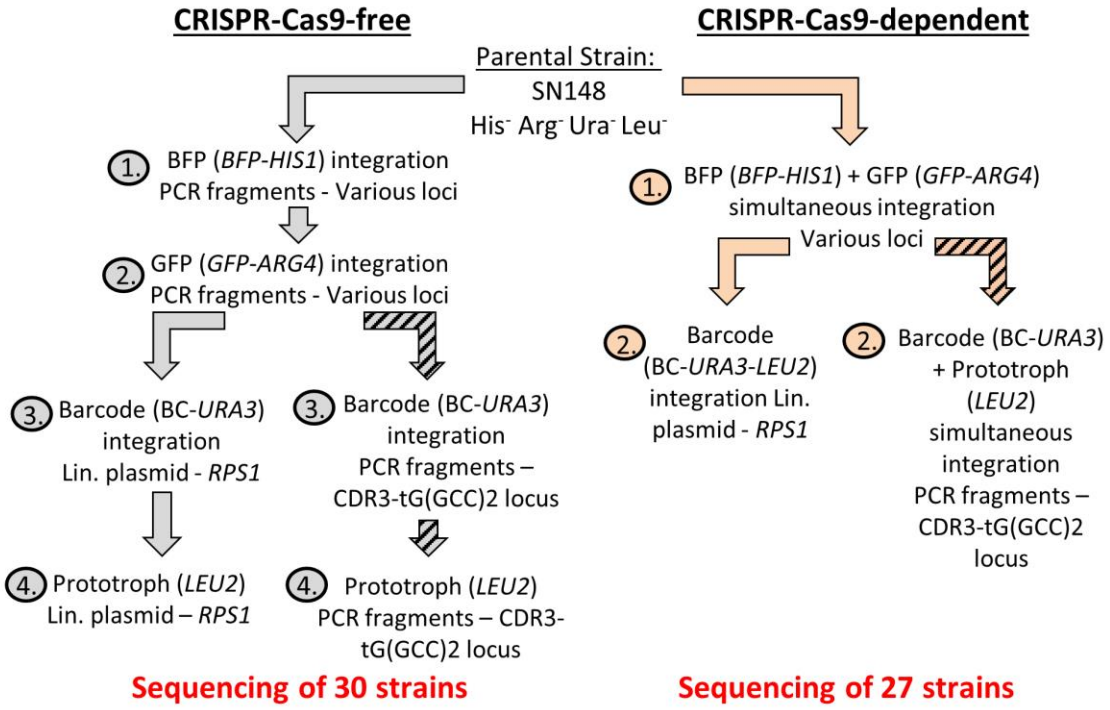

Figure 1: CRISPR-Cas9-free and -dependent transformation strategies used for strain construction.
Derived from the *C. albicans* reference strain SC5314, the parental SN148 strain was sequentially transformed using CRISPR-Cas9-free (gray) or CRISPR-Cas9-dependent (orange) strategies in order to integrate (i) the BFP/GFP LOH reporter system, (ii) a unique barcode sequence and (iii) the leucine marker rendering the strains prototroph. Strains constructed using the CRISPR-Cas9-free strategy underwent four transformation steps, as opposed to two using the CRISPR-Cas9-dependent strategy. The barcode sequence (BC) and *LEU2* auxotrophic marker are introduced at the *RPS*1 locus on Chr1. The BC and *LEU*2 marker are introduced at another locus for the strains carrying the LOH reporter system on Chr1 (in the intergenic sequence between the *CDR3* and tG(GCC)2 loci), as indicated by the striped arrows.

All engineered strains underwent basic phenotyping to assess if the transformation process drastically impacted their behavior. Functionality of auxotrophy markers associated with integration cassettes was tested by spot assay, where all strains demonstrated a capacity to grow on media with dropout amino acids of histidine, arginine, uridine or leucine. Moreover, functionality of fluorescence proteins, BFP and GFP, was validated by fluorescence microscopy and flow-cytometry. No strain exhibited any difference in colony morphology on YPD or SD media as compared to the parental strain SN148. Additionally, the doubling time of all constructed strains was measured in YPD at 30°C (Figure 2). We observed that the parental SN148 strain possesses a longer doubling time, which is most likely due to the uridine auxotrophy (Kirsch and Whitney 1991). Among the strains displaying genomic changes as highlighted in yellow in Figure 2, some have increased doubling time as compared to the average doubling time obtained for all 57 strains analyzed in this study.

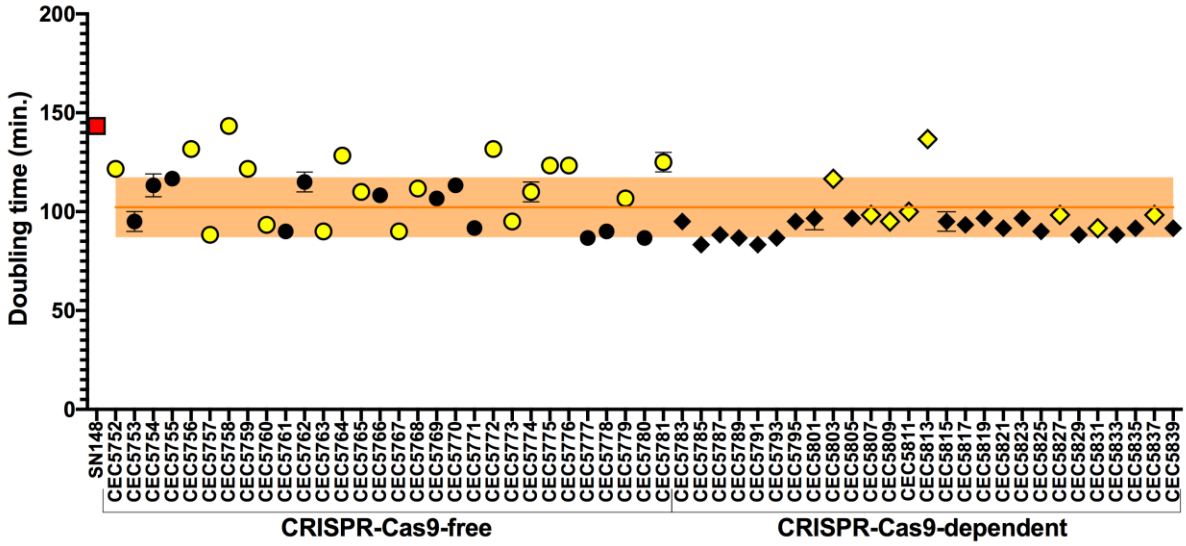

Figure 2: Doubling times in YPD at 30°C.
Each data point is representative of the average doubling time (min.) (n=8) of the parental (red square, SN148) and constructed strains (circles: CRISPR-Cas9-free or rhombus: CRISPR-Cas9-dependent) with error bars (+/-SD). Average doubling time of all constructed strains (57 strains) is represented by the orange line (+/-SD). Strains displaying at least one genomic change are identified in yellow while, black color indicate strains which are free of transformation-acquired genomic changes.

Genome sequencing

To ensure that the strains did not acquire major genome changes during transformation, we seized the opportunity to perform whole genome sequencing. For the sake of time and resources, we took advantage of a pipeline available for rapid genome sequencing of our 57 strains and that was calibrated to give a 50x coverage for a bacterial genome (4.6 Mb). Based on *C. albicans* genome size, a 15x coverage was expected, which would still be sufficient to allow detection of LOH and aneuploidies. An average sequence depth of 25.5 +/- 17.2 (ranging from 7.44x to 73.12x) was obtained upon sequencing of the 57 *C. albicans* strains (Table S4). Post-sequence cleanup and SNP calling were conducted, and allele ratios at heterozygous sites (ABHet) were evaluated in order to allow identification of genome changes, in particular aneuploidy and LOH. Despite the low sequencing depth, ABHet plots were successfully generated by plotting the allele balance at heterozygous positions across the genome (Figure S1), and used to visually identify large scale LOH events as represented by an absence (or weak presence) of ABHet values. Although less striking in ABHet plots (Figure S1) due to limited sequencing depth, hyperploidy events (aneuploidy with a chromosome number that is more than the diploid number) were identified by plotting the distribution of ABHet values and calculating the mean ABHet value per chromosome, as hyperploid chromosomes would shift away from the 0.5 ABHet value, which represents a 0.5 allelic ratio (50/50 heterozygous nucleotide ratio at the given position). Thus, this method allowed us to differentiate ABHet values corresponding to trisomy (0.33 or 0.66) or tetrasomy

(0.25 or 0.75) (Figure 3A, Table S5). The genomic changes identified throughout the 57 strains are graphically summarized in Figure 3B.

As illustrated by elevated standard deviations (Table S4), an uneven sequencing depth was obtained across our genomes, probably resulting from library preparations being optimized for bacterial genomes. This prevented us from using sequencing depth data to identify or characterize genome changes and therefore we could not differentiate monosomies from diploid whole-chromosome LOH events. As we solely relied on ABHet analysis to study transformation-induced genome changes, we do acknowledge that we may be underestimating the occurrence of aneuploidy events, notably events resulting in a balance of both haplotypes such as balanced tetraploidy (events 2xHapA 2xHapB) which would also be represented as a 0.5 ratio in the ABHet graphs. Nevertheless, this methodology allowed us to efficiently identify numerous genome changes generated during the transformation process.

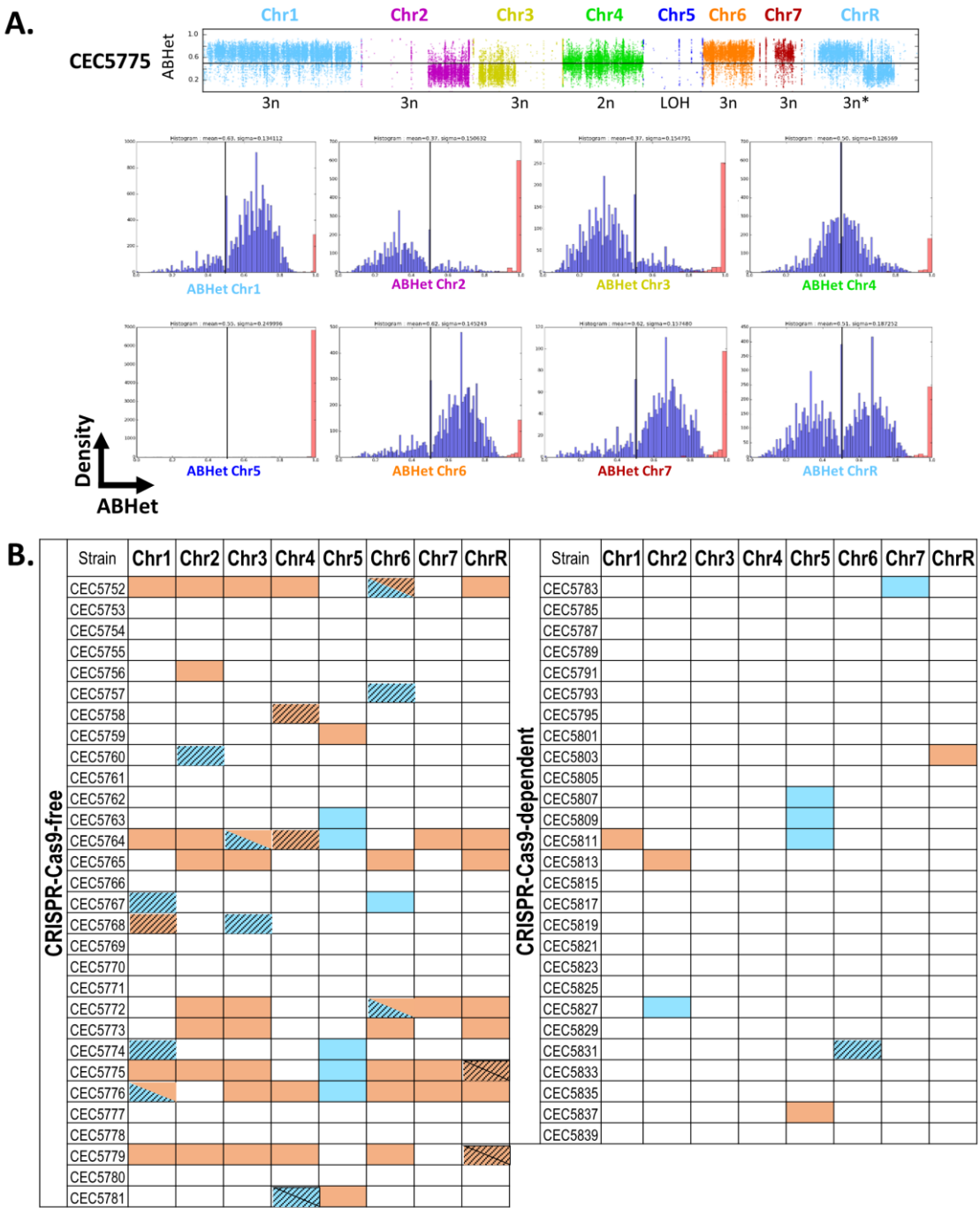

Figure 3: Identification of transformation-induced genome changes within the 57 C. albicans sequenced strains.

A. Determining the average ABHet values per chromosome for strain CEC5775. The upper panel shows the plots of allele balances for the eight chromosomes of strain CEC5775. Below, histograms illustrate the distribution of ABHet values across a given chromosome, where black vertical bar represents a 0.5 ABHet value (heterozygous diploid). Blue and red color represent respectively the ABHet and ABHom values. Data interpretation for each chromosome: Chr1, 6 and 7: trisomy (1xHapA, 2xHapB) - Chr2 and 3: trisomy (2xHapA, 1xHapB) - Chr4: disomy (1xHapA, 1xHapB), Chr5: LOH - ChrR: recombination event localized in proximity of the centromere + trisomy (Left arm: 1xHapA, 2xHapB, Right arm: 2xHapA, 1xHapB). Additional LOH events have previously been described in parental strains SN148 (LOH on Chr2 (Loll-Krippleber *et al.* 2015)) and in SC5314 (LOH on Chr3 and Chr7 (Abbey *et al.* 2011)). **B.** Summary of the genomic changes identified across the eight chromosomes for all 57 *C. albicans* sequenced strains, using the strategy presented in panel A. The

plots showing the allele balance at heterozygous positions and the mean ABHet values per chromosome for each strain can be found in Figure S1 and Table S5, respectively. LOH are indicated in blue while aneuploidies in orange. Genomic changes events impacting whole-chromosomes are identified by solid colours while those partially impacting chromosomes are identified by a striped pattern.

Both CRISPR-Cas9-free and CRISPR-Cas9-dependent transformations trigger unwanted GCRs but to different extents

Amongst the 30 strains constructed using the CRISPR-Cas9-free method, a total of 65 genomic changes were identified (Figure 3B). Only 40% of the strains (12/30) display no obvious genomic changes while the other 60% (18/30) have at least one genomic change, of which half (9/18 strains) possess three or more genomic changes (Figure 4A-B). Of interest, 10/18 strains display at least one rearrangement on the targeted chromosome, sometimes accompanied by additional changes on other chromosomes. The remaining 8/18 strains display changes only on chromosome(s) non-targeted by the transformation process (Figure 4D). At the chromosome scale, a quarter of the sequenced chromosomes (60/240) display at least one genomic change, including five chromosomes with two identifiable genomic changes (Figure 4A, Table S6). Although our analysis did not reveal any obvious enrichment of genomic changes on a given chromosome, we did notice slightly higher and lower amounts of changes on Chr4 and Chr7, respectively (Figure 4C). Overall, strains constructed using the CRISPR-Cas9-free display an average of 0.54 genomic changes per strain per transformation (Figure 5A).

Because of the extensive mutagenic effect of CRISPR-Cas9-free transformation, the CRISPR-Cas9-dependent method, via transient Cas9 expression, was favored to generate the second part of our collection as it reduces the number of sequential transformations (Figure 1) by targeting DNA-DSB on both homologs of the loci of interest. A total of 10 genomic changes, distributed amongst nine strains (33.3%), were identified within the 27 strains engineered with this method (Figure 4A-B). While one third of the strains (3/9) display at least one change on the targeted chromosome, at times accompanied by additional changes on other chromosomes, the others display changes only on non-targeted chromosomes (Figure 4D). At the chromosome scale, only 4.6% of the sequenced chromosomes demonstrated genomic changes (10/216) with no chromosomes displaying multiple changes (Figure 4A, Table S6). These genomic changes were distributed amongst most chromosomes, apart from Chr3 and Chr4 that, in our hands, were not affected by changes even when targeted. In contrast, Chr5 seems to exhibit the highest amount of genomic changes (Figure 4C). We estimated that strains constructed using the transient CRISPR-Cas9 display an average of 0.19 genomic changes per strain per transformation (Figure 5A).

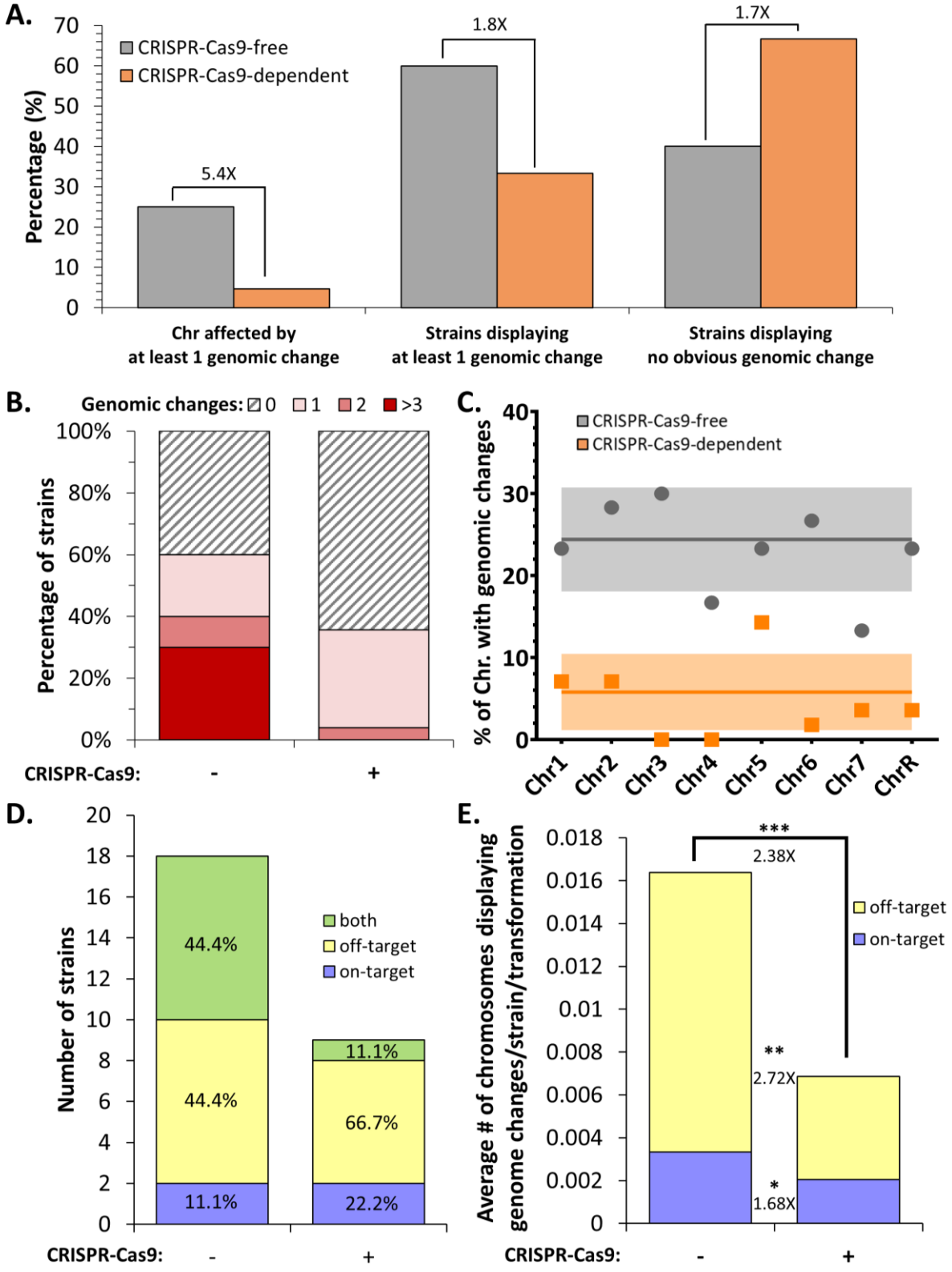

Figure 4: Quantification of genome changes identified within C. albicans strains engineered using two transformation methods.

A. Percentage of chromosomes and strains impacted by transformation-associated genomic changes as well as percentage of genomic change-free strains. **B.** Studying concurrent genomic changes within strains. **C.** Percentage of each chromosome affected by at least one genomic change. **D.** Representation of the number of strains displaying genomic change(s) on on-target, off-target or both

types of chromosomes. Fraction of strains within genomic change-displaying strains indicated by percentages. **E.** Frequency of transformation-associated genomic changes tabulated as an average number of chromosomes displaying genome changes per strain per transformation. Differences in genome change frequencies between strains constructed using two transformation strategies are represented as fold changes, T-test ($p \leq 0.05$).

Aneuploidies are overrepresented in strains engineered using the CRISPR-Cas9-free method

Genome sequence analysis allowed us to investigate in more details the nature of the genomic changes observed in the strains engineered using either the CRISPR-Cas9-free or -dependent methods. The 75 large scale genomic changes identified above were categorized into 2 major types: LOH and hyperploidy. Additionally, we also report on the size of the genome changes, spanning partially or entirely a chromosome. Hyperploidies were defined as regions possessing heterozygous SNPs with an allele balance value greater or inferior to 0.5. For comparison purposes, we took into account the fact that the strain did not go through the same number of transformations in both methods and presented the data in terms of number of changes/strain/transformation (Figure 5A).

With the CRISPR-Cas9-free transformation strategy, we observed on average 2.86 times more aneuploidy events (0.40 events) than LOH events (0.14 events) per strain per transformation (T-test, p-value = 0.0279). Additionally, both types of genome changes were identified, in decreasing order of abundance; full chromosome hyperploidies (0.35 events/strain and transformation), partial chromosomal LOH (0.09 events), partial chromosomal hyperploidies (0.05 events) and full chromosome LOH (0.05 events) (Figure 5A).

In contrast, strains transformed with the CRISPR-Cas9-dependent transformation strategy display a comparable average number of hyperploidy (0.07 events) and LOH (0.11 events) per strain per transformation (T-test, p-value=0.493). In decreasing order of abundance, we detected full chromosome LOH (0.09 events), full chromosome hyperploidies (0.06 events), partial chromosomal LOH (0.02 events) and partial chromosomal hyperploidy (0.02 events) (Figure 5A). Our detailed analysis did not reveal any obvious link between the nature of the genomic changes and a specific chromosome, though Chr5 and Chr6 seem to be implicated more frequently in LOH events in comparison to the other chromosomes (Figure 5B).

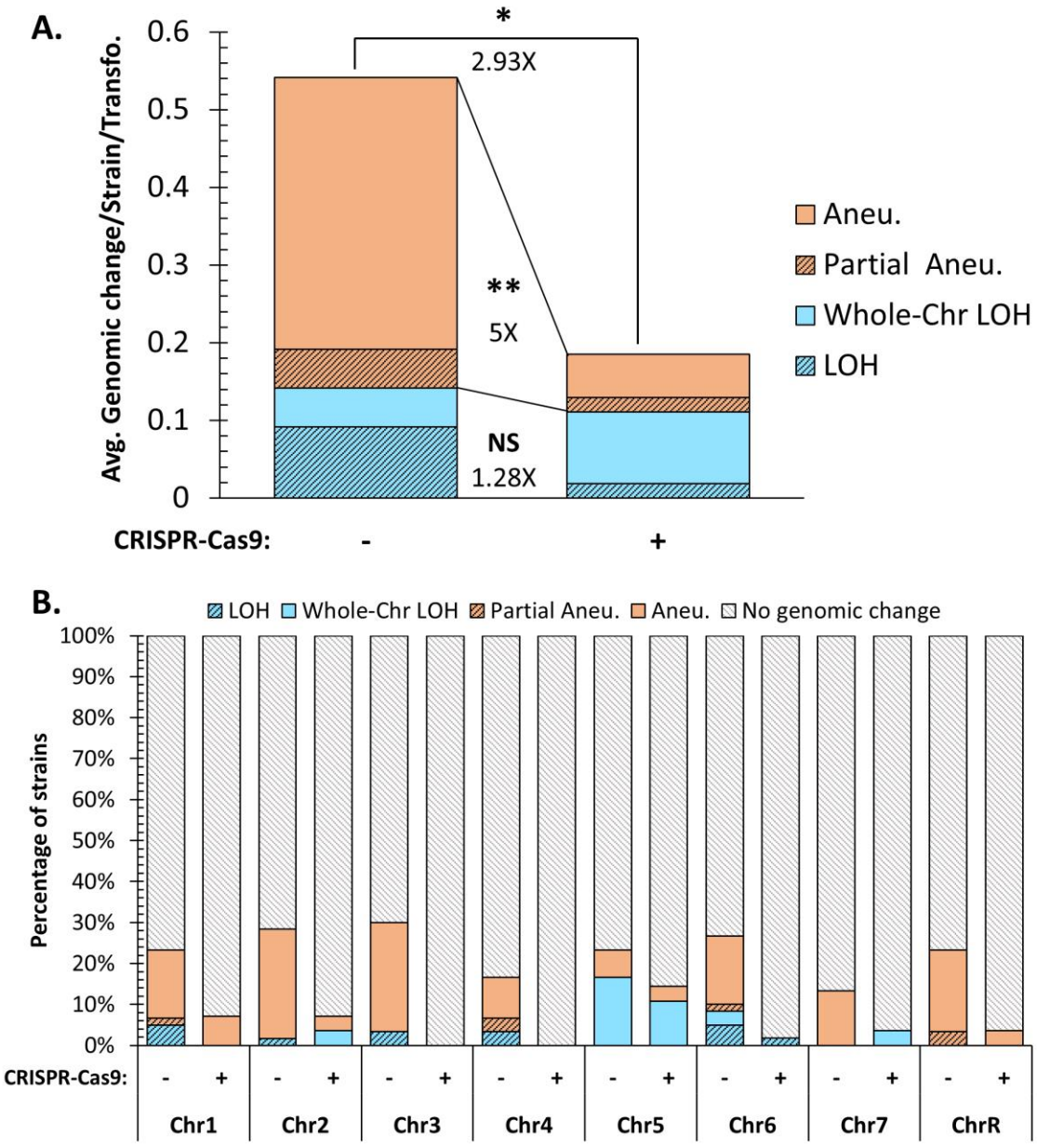

Figure 5: Nature of genomic changes identified within sequenced C. albicans strains.
A. Frequency of transformation-associated genomic changes tabulated as an average number of genome changes per strain per transformation. Differences in genomic change frequencies between the two transformation methods are represented as fold changes, T-test ($p \geq 0.05$). **B.** Percentage of strains displaying each type of genomic change per chromosome and transformation strategy.

Discussion

The construction of a *C. albicans* strain collection, aimed to study genome-wide LOH dynamics, allowed us to retrospectively compare the mutagenic landscapes of strains engineered using two transformation strategies: CRISPR-Cas9-free vs. CRISPR-Cas9-dependent. Whole genome sequencing of 57 engineered strains of *C. albicans* permitted the identification and comparison of transformation-acquired genomic changes. Although this study was not designed for this purpose, we observed during the retrospective analysis that *C. albicans* strains engineered using the CRISPR-Cas9-dependent transformation method display significantly less concomitant genomic changes, notably in terms of hyperploidy events, as compared to the strains constructed using the CRISPR-Cas9-free transformation strategy.

Similarly to transformation protocols of other yeast species, *C. albicans* competent cells are prepared and suspended in a lithium acetate solution with single-stranded carrier DNA (often single-stranded salmon sperm DNA) and the exogenous DNA, intended to integrate the yeast genome. The exogenous DNA usually consists of linearized plasmid DNA with free ends having homology with the yeast genome, or PCR-amplified integration cassettes (Gola *et al.* 2003), which are targeted to a specific locus by two homology regions (70 bp – 100 bp), flanking the DNA of interest associated to a selection marker. These homology regions on the free DNA allow strand invasion at the genomic homology site, ultimately resulting in the integration of the repair template through recombination events. Studies have shown that DNA-DSBs in *C. albicans* are predominantly repaired via HR, and only rarely by non-homologous end joining (NHEJ) (Andaluz *et al.* 2001; Ciudad *et al.* 2004; Legrand *et al.* 2007; Feri *et al.* 2016; Vyas *et al.* 2018). This facilitates genetic manipulations of both laboratory and clinical isolates as efficient genetic modifications are not restricted to NHEJ-deficient strains (as it is the case in numerous fungi (Krappmann *et al.* 2006; Goins *et al.* 2006; Choquer *et al.* 2008)). However, this technique allows the targeting of a single homologue at one locus, i.e. one integration per transformation process, rendering certain strain constructions cumbersome. The development of molecular tools permitting locus specific DNA-DSBs has largely facilitated genome editing in numerous eukaryotic organisms, by means of zinc-finger nucleases (ZFN) or TALEN which recognize a specific DNA sequence and mediate DNA-DSBs by the FokI nuclease (Wood *et al.* 2011). Nevertheless, the most recent and popular targeted genetic modification system is CRISPR-Cas9 because it is highly versatile, and its implementation is less laborious than previously mentioned nuclease. A sgRNA guides the Cas9 nuclease to its target sequence where it induces a DNA break. Promoting DNA-DSBs at the target locus increases HR, resulting in a higher transformation

efficiency, and allows the integration of exogenous DNA at the target locus on multiple homologs, facilitating genetic engineering in diploid and polyploid organisms.

We observed that both CRISPR-Cas9-free and -dependent transformations provoke unwanted genomic changes in *C. albicans*. Our data clearly showed that the transformation process itself is mutagenic as, by reducing in half the transformation steps required to engineer our strains, we approximately doubled the number of strains deprived of unwanted transformation-associated genomic changes (Figure 4A). Despite differences in the number of transformation steps and constructs (use of one vs two barcode/prototrophy constructs, see material and methods section), both transformation strategies produced genomic changes, although to different extents, as *C. albicans* strains constructed using the CRISPR-Cas9-dependent transformation process displayed obvious reduction in concomitant genomic changes (Figure 4). Indeed, we observed that 5.4x fewer chromosomes were affected by undesirable genomic changes amongst the strains engineered with the CRISPR-Cas9-dependent method. Strains constructed using the CRISPR-Cas9-free strategy displayed multiple coexisting genomic changes, with 40% of them carrying more than 1 change (up to 8 events) in contrast to 3.7% of strains engineered using the CRISPR-Cas9-dependent method (Figure 4B). The global mutagenic frequency (average genomic change/strain/transformation) is significantly higher (2.93x, p-value = 0.0145) when using the CRISPR-Cas9-free strategy as compared to the CRISPR-Cas9-dependent method (Figure 5A). These results suggest that strains engineered by CRISPR-Cas9-free transformations probably derive from cells which endured strong perturbation in genomic stability thus, generating numerous and concomitant genomic changes. In contrast, directing DNA-DSBs with the CRISPR-Cas9-dependent method favors HR-mediated repair and integration of exogenous DNA at the target locus, hence yielding more transformants independently of overall genome perturbation. From a mechanistic perspective, the free linear DNA (repair template) may attempt multiple strand invasions during homology search, leading to a series of abortive Holliday junctions, perhaps altering the overall chromosomal stability, and consequently resulting in large LOH and/or aneuploidies on on- and/or off-target chromosomes. On the other hand, global mutagenic reduction in strains resulting from CRISPR-Cas9 transformation may potentially be associated with the activation of the DNA damage response upon DNA-DSB, leading to a global response yielding a cellular environment conductive to repair; changes in the cell cycle, chromosome mobility, transcription and nuclear dNTP levels (Sirbu and Cortez 2013).

In other organisms, it has been shown that CRISPR-Cas9 can induce off-target DNA-DSBs which may have grave implications (Fu *et al.* 2013; Banakar *et al.* 2019). To the best of our knowledge, this has not been studied in *C. albicans.* The genomic changes identified within

our strain collection are distributed quasi equally between the 8 chromosomes and LOH and/or hyperploidies are predominantly found on off-target chromosomes in strains derived from both strategies even though strains engineered using the CRISPR-Cas9-free strategy are always more impacted by genomic changes (Figure 4E, Figure 5B). Unfortunately, as genomic changes more frequently involved off-target chromosomes (2.38x) (Figure 4E), we cannot revoke the possibility that the latter result from off-target Cas9 DNA-DSBs in strains that were generated using the CRISPR-Cas9-dependent strategy. Heat shock is a well-known source of DNA breaks and has been associated with appearance of genomic changes in *C. albicans* (Forche *et al.* 2011) and could explain the changes observed in our strain set, which experienced a heat shock during the transformation protocol (15 mins, 44°C). Further investigations need to be conducted in order to properly address the potential off-target activity of Cas9 in *C. albicans*.

The genome of *C. albicans* is highly tolerant to genomic rearrangement events which often arise upon exposure to various stresses, e.g. heat, fluconazole or oxidative stress (Selmecki *et al.* 2005; Bouchonville *et al.* 2009; Abbey *et al.* 2011; Forche *et al.* 2011). As transformation protocols involve inflicting stress, it comes as no surprise that we and previous studies revealed the presence of LOH and aneuploidy events in various laboratory strains. Indeed, microarray experiments and haplotype mapping have highlighted the mutagenic effect of transformation in multiple laboratory backgrounds (RM1000, CAI-4 and BWP17), between strain stocks and their derivatives (Ahmad *et al.* 2008; Bouchonville *et al.* 2009; Arbour *et al.* 2009; Abbey *et al.* 2011). For instance, transformation of SC5314 to generate the uracil auxotrophic CAI-4 strain (Fonzi and Irwin 1993) has been linked to trisomy of Chr1 (Chen *et al.* 2004) or Chr2 (Selmecki *et al.* 2005), while *HIS*1 disruption in CAI-4 resulted in deletion of telomere proximal genes of Chr5 in its derivative BWP17 (Pla *et al.* 1995; Selmecki *et al.* 2005). The same was true among our collection of 57 transformed strains for which we mainly observe two sorts of genomic changes: (i) LOH and (ii) hyperploidies (defined in our study as a chromosome or portion of chromosome present in more than two copies). In our collection, transformation-induced LOH events appear at a comparable rate of 0.14 and 0.11 LOH events/strain/transformation in strains constructed using CRISPR-Cas9-free and -dependent methods respectively (1.28x fold change, T-test, p-value = 0.5715). In contrast, strains obtained using the CRISPR-Cas9-dependent strategy display a 5-fold decrease (T-test, p-value = 0.0086) in hyperploidy events as compared to the strains constructed using the CRISPR-Cas9-free method (Figure 5A). Frequent chromosome trisomies were observed in our strain set (Table S5-S6). Chromosome trisomies have often been associated to rapid adaptation to stress conditions, such as demonstrated by an increased fitness in presence of fluconazole upon Chr5 aneuploidy (Selmecki *et al.* 2006b, 2009). Additionally, genotypic and

phenotypic diversification, namely Chr5 and Chr6 trisomies, have been described upon exposure of *C. albicans* to the oral niche leading to a low-virulence phenotype, suggesting an adaptation resulting in a commensal-like phenotype (Forche *et al.* 2018, 2019).

Altogether, transformation protocols limiting the number of sequential transformation steps should be favored in order to minimize acquisition of unwanted transformation-associated genomic changes. Additionally, we want to highlight the importance of generating, testing, and analyzing multiple transformants as, at best, a third of transformed strains undergo genomic changes (Figure 4A-B). Though ideal, routine whole genome sequencing of transformed strains is not necessarily feasible. Alternatives have been proposed by others such as a q-PCR-based assay to monitor copy number variations throughout the genome of *C. albicans* (Arbour *et al.* 2009). However, this technique is limited to the identification of aneuploidies and does not permit the detection of equality important LOH events. The SNP-RFLP technique, allowing to monitor the allelic status for multiple genomic sites, (Forche *et al.* 2009b) can be complementary to detect LOH events. Additionally, re-complementation experiments are key in reverse genetics, allowing to reinforce the conclusions drawn from phenotypic observations.

To conclude, this study has permitted a detailed investigation of the frequency and nature of genomic changes which occur upon transformation of *C. albicans* cells comparing transformation-induced mutagenic landscapes in strains constructed with two transformation methods: a CRISPR-Cas9-free and a CRISPR-Cas9-dependent strategy. By only investigating large genomic rearrangements, we highlight that *C. albicans* transformation is highly mutagenic and recognize that we may be vastly underestimating this mutagenic effect, e.g. we did not investigate the presence of indels and SNPs. Nevertheless, the CRISPR-Cas9-dependent strategy seems to reduce transformation-associated concomitant genomic changes, especially with regards to hyperploidy events.

Materials and methods

Strains and culture conditions

Yeast strains described in the study were constructed starting from *C. albicans* strain SN148 (His$^-$, Arg$^-$, Ura$^-$ and Leu$^-$) (Noble and Johnson 2005). Yeast cells were cultured on/in rich YPD medium (1% yeast extract, 2% peptone, 2% dextrose). Synthetic Defined (SD) (0.67% yeast nitrogen base without amino acids, 2% dextrose) and Synthetic Complete (SC) (0.67% yeast nitrogen base without amino acids, 2% dextrose, 0.08% drop-out mix with all the essential amino-acids) media were used for selection. Solid media were obtained by adding 2% agar.

Cloning experiments were conducted using One Shot TOP10 chemically competent *Escherichia coli* K-12 cells (ThermoFisher Scientific). *E. coli* strains were cultured on/in LB or 2YT medium, with appropriate antibiotics for selection purposes (50 µg/ml kanamycin, 50 µg/ml ticarcillin). Solid media were obtained by adding 2% agar.

All *C. albicans* strains and *E. coli* plasmids are listed in Table S1 and Table S2, respectively.

CRISPR-Cas9-free transformation

Strains constructed using the following CRISPR-Cas9-free protocol underwent four sequential heat shock and lithium acetate/PEG rounds of transformations (Walther and Wendland 2003) in order to (i-ii) integrate the BFP/GFP LOH reporter system at a distinct genomic locus, (iii) integrate a unique barcode associated to the *URA3* auxotrophic marker at the *RPS1* locus and finally (iv) integrate the *LEU2* auxotrophic marker at the *RPS1* locus rendering the strains prototroph (Figure 1). The strategy was to integrate the BFP/GFP LOH reporter system in the most telomere proximal intergenic region of ≥5 kb on each chromosome arm (Table S1). To this purpose, 120 bp primers were designed, each composed of 20 bp complementary to both the P$_{TDH3}$-*GFP-ARG4* and P$_{TDH3}$-*BFP-HIS1* cassettes and 100 bp tails possessing the complementary sequences of the targeted integration locus (Table S3). Each primer pair was used to amplify both the P$_{TDH3}$-*GFP-ARG*4 and P$_{TDH3}$-*BFP-HIS1* cassettes from plasmids pCRBluntII-P$_{TDH3}$-*GFP-ARG4* and pCRBluntII-P$_{TDH3}$-*BFP*-Cd*HIS1*, respectively. Each cassette was amplified in a total PCR volume of 500 µL, precipitated in 100% ethanol and re-suspended in 100 µL of distilled sterile water. For each transformation, competent cells were transformed with approximately 5 µg of the appropriate DNA cassette. The parental SN148 strain was initially transformed with the P$_{TDH3}$-*GFP-ARG4* cassette and then subjected to a second round of transformation with the P$_{TDH3}$-*BFP-HIS1* cassette. These two transformation steps allowed the integration of the BFP/GFP LOH reporter system at a

given intergenic locus (Table S1). The resulting strains, except those possessing the BFP/GFP LOH reporter system on Chr1, were then retransformed with *Stu*I-linearized CIp10-P_{TET}-BC-*URA3* plasmids, each containing a unique barcode (BC) sequence and targeting the *C. albicans RPS1* locus on Chr1. These plasmids are derived from a private laboratory collection of Cip10-P_{TET}-BC-GTW-*URA3* plasmids (unpublished), each one possessing a unique 25-nucleotide long unique barcode sequence. The Gateway (GTW) cassette was removed by *Hind*III digestion and self-ligation, to ensure that its presence does not influence on the biology of *C. albicans*. Lastly, BFP/GFP barcoded strains were rendered prototroph by a 4th round of transformation involving the integration of the *Stu*I-linearized CIp10-*LEU2* plasmid at the *RPS1* locus. Conversely, in strains bearing the BFP/GFP LOH reporter system on Chr1, the BC-*URA3* and *LEU2* cassettes were integrated on Chr4 at the CDR3-tG(GCC)2 locus to avoid loss of these latter upon LOH. For these strains, the BC-*URA3* and *LEU2* cassettes were generated using the same PCR-amplification protocol described above for the P_{TDH3}-*GFP-ARG4* and P_{TDH3}-*BFP-HIS1* cassettes, where long tailed primers (Table S3) were used for amplification from plasmids CIp10-P_{TET}-BC-*URA3* and CIp10-*LEU2*, respectively. Throughout the strain construction process, a selective pressure was always maintained in order to ensure the selection of transformants carrying all integration cassettes. In addition, at each transformation step, junction PCRs were conducted to ensure the proper integration of cassettes using primers listed in Table S3.

CRISPR-Cas9-dependent transformation

Unlike in the CRISPR-Cas9-free transformation method, both homologs may be simultaneously targeted for cassette integration with the CRISPR-Cas9-dependent transformation protocol. Thus, by directing a DNA-DSB with a locus-specific sgRNA, the BFP/GFP LOH reporter system can be introduced with only one confrontation to heat shock and lithium acetate/PEG, rather than two treatments. Hence, only 2 transformation rounds were required for strain construction, where (i) the BFP/GFP LOH reporter system was integrated at distinct genomic loci and (ii) strains were simultaneously barcoded and rendered prototroph using both *URA3* and *LEU2* auxotrophic markers (Figure 1). A total of 11 unique 20 bp sgRNAs were designed using CHOPCHOP (Montague *et al.* 2014), targeting the same integration loci as the ones chosen in the CRISPR-Cas9-free protocol for integration of the BFP/GFP LOH reporter system (Table S3). Because the CIp10-derived BC-*URA3-LEU2* plasmids could not be targeted to the *RPS*1 locus on Chr1 in the strains bearing the BFP/GFP LOH reporter system on Chr1, an additional sgRNA was designed to target the BC-*URA3* and *LEU2* markers on Chr4 at the CDR3-tG(GCC)2 locus (Table S3) in these strains.

We used a transient CRISPR-Cas9 system (Min *et al.* 2016), which does not necessitate the genomic integration of either Cas9 or sgRNAs. The construction of sgRNAs and amplification of Cas9 cassettes from the pV1093 plasmid were conducted as described in Min et al. while *BFP-HIS1, GFP-ARG4,* BC-*URA3* and *LEU2*-bearing cassettes were constructed as previously described. SN148 cells were co-transformed with 3 µg of the P_{TDH3}-*GFP-ARG4* cassette, 3 µg of the P_{TDH3}-*BFP-HIS1* cassette, 1 µg of the Cas9 cassette and 1 µg of sgRNA using the lithium acetate/PEG transformation protocol. Transformants were then selected on SC-Arg-His medium and junction PCRs were performed in order to ensure proper integration of both cassettes at the targeted locus. Strains bearing the BFP/GFP LOH reporter system on Chr1 were then transformed using the transient CRISPR-Cas9 system targeting BC-*URA3* and *LEU2* cassette integration at the CDR3-tG(GCC)2 locus on Chr4. For the remaining strains, both BC-*URA3* and *LEU2* markers were integrated in one transformation step. We did so using the Gateway recombination system, where the *LEU2* gene was transferred from a pDONR-*LEU2* into Cip10-P_{TET}-BC-GTW-*URA3* plasmids (unpublished) by an LR reaction. The unique Cip10-P_{TET}-BC-*LEU2-URA3* plasmids were then linearized by *Stu*I and integrated at the *RPS1* locus. These transformants were selected on SD medium and junction PCRs were performed.

Strain phenotyping

All selected strains underwent basic phenotypic characterization upon validation of cassette integration at targeted loci by junction PCRs. Functionality of the auxotrophic markers was evaluated by drop tests on SC medium with the appropriate drop-out amino acid mix, based on the tested marker. Overnight saturated cultures of selected strains in liquid YPD were spotted on solid YPD, SC-His, SC-Arg, SC-Ura and SC-Leu and placed at 30°C for 24h to monitor growth. Furthermore, the functionality/intensity of both fluorescence proteins (BFP and GFP) was validated by flow cytometry (MACSQuant analyzer (Miltenyi Biotec)) and fluorescence microscopy (Olympus IX83). The colony morphology of all strains was also assessed both on solid YPD and SD media, at 30°C. Finally, doubling times were evaluated in liquid YPD media at 30°C by measuring the optical density with TECAN Infinite over a 24 period.

DNA extraction and whole genome sequencing of strains

Prototroph strains were cultured in 5 mL of liquid SD medium overnight at 30°C and DNA was extracted following the manufacturer's protocol using the QIAGEN QIAamp DNA Mini Kit. The DNA was eluted in a total volume of 100µL. The genomes were sequenced at the P2M Platform of Institut Pasteur, using the Illumina sequencing technology. Libraries were

prepared with the NEXTERA XT Sequencing kit and NextSeq500 platforms were used to generate 151 bp paired-ends reads.

Identification of gross-chromosomal rearrangements

Sequences and genomic variations were analyzed as previously described in (Ropars *et al.* 2018; Sitterle *et al.* 2019). Each set of paired-end reads was mapped against the *C. albicans* reference genome, SC5314 haplotype A and haplotype B (version A22-s07-m01-r57), using Minimap2 version (Li 2018). SAMtools version 1.9 and Picard tools version 2.8.1 (http://broadinstitute.github.io/picard) were then used to filter, sort and convert SAM files. SNPs were called using Genome Analysis Toolkit version 3.6, according to the GATK Best Practices. SNPs were filtered using the following parameters: VariantFiltration, QD < 2.0, LowQD, ReadPosRankSum < -8.0, LowRankSum, FS > 60.0, HightFS, MQRankSum < -12.5, MQRankSum, MQ < 40.0, LowMQ, HaplotypeScore > 13.0. Sequencing depths were also calculated using the Genome Analysis Toolkit (Table S1). The GATK variant filtration walker (VariantAnnotator) was used to add allele balance information to VCF files. The value of allele balance at heterozygous sites (ABHet) is a number that varies between 0 and 1. ABHet is calculated as the number of reference reads from individuals with heterozygous genotypes divided by the total number of reads from such individuals. Thus, a diploid genome will be defined by an ABHet value of 0.5. In contrast, while a triploid strain may contain either three identical alleles (allelic frequency of 1) or two identical alleles and one different allele (frequency of 0.66 and 0.33), a tetraploid strain may have allelic frequencies of 0.5 (2x2 identical alleles) or 1 (4 identical alleles) or 0.25 and 0.75 (3 identical alleles and 1 different allele). In order to obtain an average ABHet value per chromosome, we evaluated the ABHet and homozygous calls (ABHom) with the AlleleBalance annotation GATK module (Table S5). Histograms were built based on the number of SNPs with ABHet values with the matplotlib 2D graphics package (Hunter 2007). Blue and red color represent respectively the ABHet and ABHom values.

Acknowledgments

We are grateful to the members of the Institut Pasteur Mutualized Platform for Microbiology (P2M) for Illumina sequencing.

T.M, M.L., and C.D. designed experiments. T.M. performed experiments. T.M., C.M., M.L and C.D. analyzed data. T.M., M.L., and C.D. wrote the manuscript.

T.M. is the recipient of a Ph.D. fellowship from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID). We acknowledge support from the French Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases [ANR10-LABX-62-IBEID]).

Supplemental materials

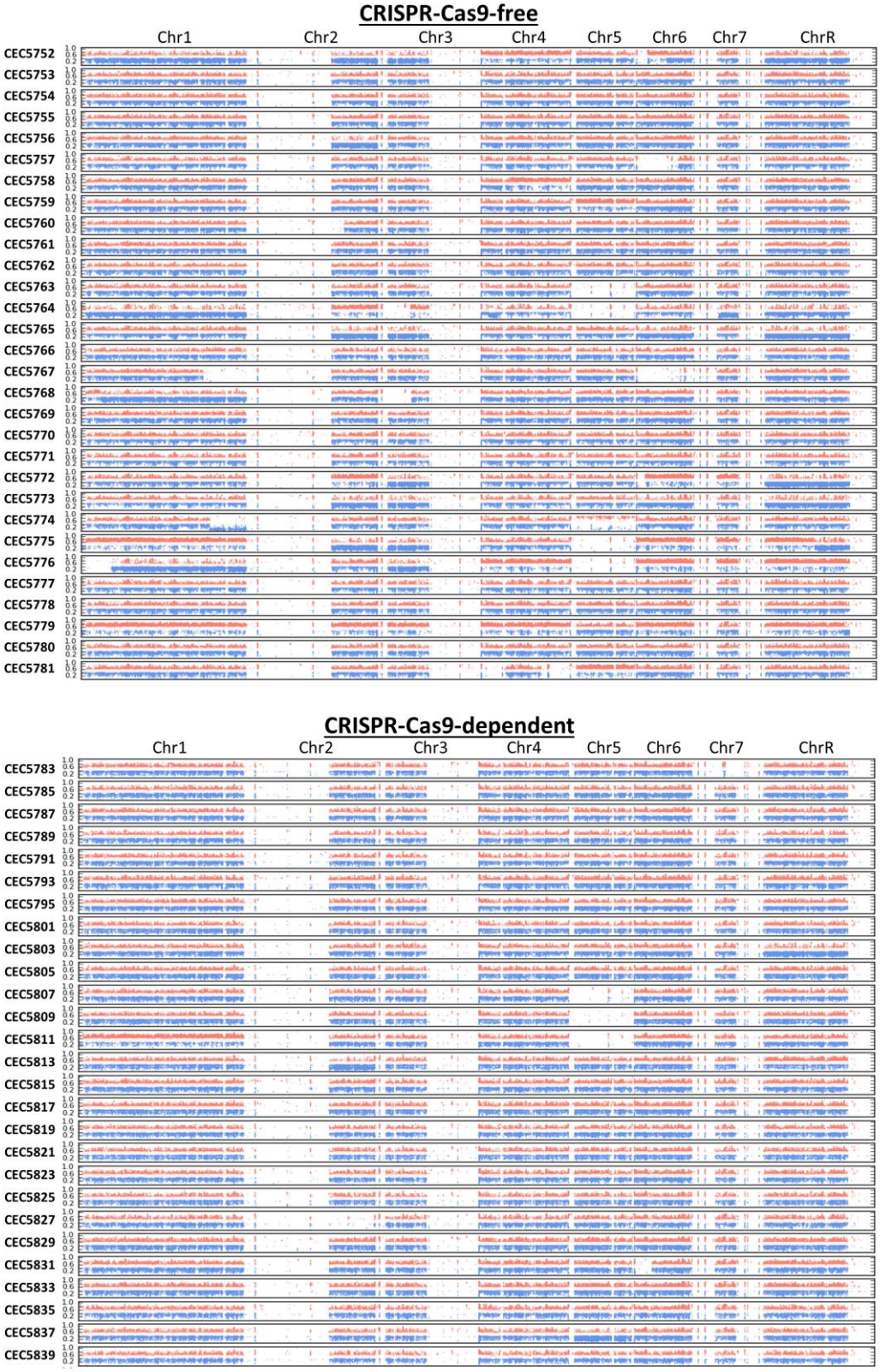

Figure S1: Allele balance at heterozygous site plots for the collection of 57 C. albicans strains. Plots of allele balance at heterozygous sites (ABHet) across the eight chromosomes for each of the 57 *C. albicans* sequenced strains. Colour representation of ABHet values: > 0.6 in salmon, <0.6 and >0.4 in white and, <0.4 in blue.

Table S1: C. albicans strains used in this study. *

Strains	Genotype	Location of BFP/GFP sys.	Location of BC-*URA3* & *LEU2**	Barcode Sequence	Transfo. Protocol	Ref. e
SN148	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434*	-	-	-	-	(1)
CEC5752	SN148 Ca22chr1_C_albicans_SC5314:144339 to 144749::P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*/Cm*LEU2*	Chr1L	Chr4L	TCTCGGTCT AGGACTAAT ACAAGCT	CRISPR-Cas9-free	This study
CEC5753	SN148 Ca22chr2_C_albicans_SC5314:252055 to 252382:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr2L	*RPS*1	TTCTCCGAG TCGTGGACC TATTACG	CRISPR-Cas9-free	This study
CEC5754	SN148 Ca22chr3_C_albicans_SC5314:129967 to 130211:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr3L	*RPS*1	TCAGCACGC CTAGCCTCT GAAGTAC	CRISPR-Cas9-free	This study
CEC5755	SN148 Ca22chr4_C_albicans_SC5314:128561 to 129559:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr4L	*RPS*1	TCTCAAGGT CGTCCTGGG TTACTGT	CRISPR-Cas9-free	This study
CEC5756	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr5L	*RPS*1	TTCGTGATA TGTGTAAGC AGACCCA	CRISPR-Cas9-free	This study
CEC5757	SN148 Ca22chr6_C_albicans_SC5314:154762 to 155045:: P_{TDH3}-*GFP*-*Cd*ARG4*/P_{TDH3}-*BFP*-*CdHIS1* *RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr6L	*RPS*1	TACGTCTCT TGCTCGATA GCCCGAT	CRISPR-Cas9-free	This study
CEC5758	SN148 Ca22chr7_C_albicans_SC5314:108458 to 108838:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr7L	*RPS*1	TCTAGTTAC CTAACTAAA GCCCGAT	CRISPR-Cas9-free	This study
CEC5759	SN148 Ca22chrR_C_albicans_SC5314:79361 to 79780:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	ChrRL	*RPS*1	TTCGTAAAT ACACTCGGG CAGCACG	CRISPR-Cas9-free	This study
CEC5760	SN148 Ca22chr1_C_albicans_SC5314:3111775 to 3111996:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*/Cm*LEU2*	Chr1R	Chr4L	TTAGGCTTA GGAACTACG CATGCCA	CRISPR-Cas9-free	This study
CEC5761	SN148 Ca22chr2_C_albicans_SC5314:200708 4 to 2007290:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr2R	*RPS*1	TGGCTACAA CAGCTAGGA TCTATAG	CRISPR-Cas9-free	This study
CEC5762	SN148 Ca22chr3_C_albicans_SC5314:172606 8 to 1726326:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr3R	*RPS*1	TTATCGATA CTGCCGTAA CAACGGA	CRISPR-Cas9-free	This study
CEC5763	SN148 Ca22chr4_C_albicans_SC5314:145279 1 to 1453180:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS*1::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr4R	*RPS*1	TACGATCTA ACGAGAGAT TGCCCTA	CRISPR-Cas9-free	This study

CEC5764	SN148 Ca22chr5_C_albicans_SC5314:1085416 to 1085845:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr5R	*RPS*1	TTCACACCG TCGACACAG AAGCGTC	CRISPR-Cas9-free	This study
CEC5765	SN148 Ca22chr7_C_albicans_SC5314:911053 to 911347:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr7R	*RPS*1	TTGTTAGCC AGGATGCCT GCCCATC	CRISPR-Cas9-free	This study
CEC5766	SN148 Ca22chrR_C_albicans_SC5314:2085021 to 2085583:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	ChrRR	*RPS*1	TATTAAGCC CAGATCACC GAGTACC	CRISPR-Cas9-free	This study
CEC5767	SN148 Ca22chr1_C_albicans_SC5314:144339 to 144749:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*/Cm*LEU2*	Chr1L Chr4L		TTCGTGACG TTGACGCTT AGGAAGT	CRISPR-Cas9-free	This study
CEC5768	SN148 Ca22chr2_C_albicans_SC5314:252055 to 252382:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr2L	*RPS*1	TTATGCAGA GGAGCGAAT TAACTTC	CRISPR-Cas9-free	This study
CEC5769	SN148 Ca22chr3_C_albicans_SC5314:129967 to 130211:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr3L	*RPS*1	TGGCTCAGT CTCGGTAAA AGGTTAC	CRISPR-Cas9-free	This study
CEC5770	SN148 Ca22chr4_C_albicans_SC5314:128561 to 129559:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr4L	*RPS*1	TCACACCTG CGCGTAATA TCGTCCG	CRISPR-Cas9-free	This study
CEC5771	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr5L	*RPS*1	TCATCATAG GGCCGTTTC CCCAAGA	CRISPR-Cas9-free	This study
CEC5772	SN148 Ca22chr6_C_albicans_SC5314:154762 to 155045:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr6L	*RPS*1	TGGCTTTAA GAGCGCTCT GGTTAGG	CRISPR-Cas9-free	This study
CEC5773	SN148 Ca22chr7_C_albicans_SC5314:108458 to 108838:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr7L	*RPS*1	TTAAAAGGT AGCCGATCC TGCCAGA	CRISPR-Cas9-free	This study
CEC5774	SN148 Ca22chrR_C_albicans_SC5314:79361 to 79780:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	ChrRL	*RPS*1	TACGAACGT CCTGACTCT TAACGGT	CRISPR-Cas9-free	This study
CEC5775	SN148 Ca22chr1_C_albicans_SC5314:3111775 to 3111996:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*/Cm*LEU2*	Chr1R Chr4L		TGGCAAAAT GCTAGAAGT ATGACAT	CRISPR-Cas9-free	This study
CEC5776	SN148 Ca22chr2_C_albicans_SC5314:2007084 to 2007290:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr2R	*RPS*1	TATGTACGT GCCGAGGC ATAGAAAC	CRISPR-Cas9-free	This study
CEC5777	SN148 Ca22chr3_C_albicans_SC5314:1726068 to 1726326:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA3*/CIp-Cm*LEU2*	Chr3R	*RPS*1	TAGGAATCT CAAGCCGGA AACTTAA	CRISPR-Cas9-free	This study

CEC5778	SN148 Ca22chr4_C_albicans_SC5314:1452791 to 1453180:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA*3/CIp-Cm*LEU2*	Chr4R	*RPS*1	TATAGCCTTAACGTGTACCGAGCAA	CRISPR-Cas9-free	This study
CEC5779	SN148 Ca22chr5_C_albicans_SC5314:1085416 to 1085845:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA*3/CIp-Cm*LEU2*	Chr5R	*RPS*1	TCAGTTGCGCGGAATGTCACATCCA	CRISPR-Cas9-free	This study
CEC5780	SN148 Ca22chr7_C_albicans_SC5314:911053 to 911347:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA*3/CIp-Cm*LEU2*	Chr7R	*RPS*1	TACGATATTCCAGGCGACGACTACC	CRISPR-Cas9-free	This study
CEC5781	SN148 Ca22chrR_C_albicans_SC5314:2085021 to 2085583:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-BC-*URA*3/CIp-Cm*LEU2*	ChrRR	*RPS*1	TAACCCCAATAGGGCGCAAACGTCT	CRISPR-Cas9-free	This study
CEC5783	SN148 Ca22chr1_C_albicans_SC5314:145484 to 145771:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*-Cm*LEU2*	Chr1L	Chr4L	TTCGTGGGAATCGCGTTATTGACAG	CRISPR-Cas9-dep.	This study
CEC5785	SN148 Ca22chr1_C_albicans_SC5314:145484 to 145771:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*-Cm*LEU2*	Chr1L	Chr4L	TTAACACGAAGCCGGTGATAATATC	CRISPR-Cas9-dep.	This study
CEC5787	SN148 Ca22chr1_C_albicans_SC5314:145484 to 145771:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*-Cm*LEU2*	Chr1L	Chr4L	TGCCACCCAACGAGACAGTTTAACA	CRISPR-Cas9-dep.	This study
CEC5789	SN148 Ca22chr1_C_albicans_SC5314:3109738 to 3110154:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*-Cm*LEU2*	Chr1R	Chr4L	TAACCTTGGGGCACACGTATGTCAC	CRISPR-Cas9-dep.	This study
CEC5791	SN148 Ca22chr1_C_albicans_SC5314:3109738 to 3110154:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*-Cm*LEU2*	Chr1R	Chr4L	TTAGGCTTAGGAACTACGCATGCCA	CRISPR-Cas9-dep.	This study
CEC5793	SN148 Ca22chr1_C_albicans_SC5314:3109738 to 3110154:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-*URA3*-Cm*LEU2*	Chr1R	Chr4L	TGGCAAAATGCTAGAAGTATGACAT	CRISPR-Cas9-dep.	This study
CEC5795	SN148 Ca22chr5_C_albicans_SC5314:1083848 to 1084278:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr5R	*RPS*1	TATAGACGTGAGTCCCCATGTCGAG	CRISPR-Cas9-dep.	This study
CEC5801	SN148 Ca22chr6_C_albicans_SC5314:1561344 to 156430:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr6L	*RPS*1	TATACGCGGGGAACCATCGACCTCA	CRISPR-Cas9-dep.	This study
CEC5803	SN148 Ca22chr6_C_albicans_SC5314:1561344 to 156430:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr6L	*RPS*1	TCAAGAGGAAACCGCGATATCTGCC	CRISPR-Cas9-dep.	This study
CEC5805	SN148 Ca22chr6_C_albicans_SC5314:1561344 to 156430:: P_{TDH3}-*GFP*-	Chr6L	*RPS*1	TTAGGGAGAGGCGCTTTAGTCCTAA	CRISPR-Cas9-dep.	This study

	Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3					
CEC5807	SN148 Ca22chr7_C_albicans_SC5314:108458 to 108838:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr7L	*RPS*1	TCACCAGAC GAAGGTAGC CTCAAGT	CRISPR-Cas9-dep.	This study
CEC5809	SN148 Ca22chr7_C_albicans_SC5314:108458 to 108838:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr7L	*RPS*1	TTAGAACGG GGCGTAAAA GTGTAAT	CRISPR-Cas9-dep.	This study
CEC5811	SN148 Ca22chr7_C_albicans_SC5314:108458 to 108838:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr7L	*RPS*1	TATAGTTTC GGGGCCGA AGAAATAA	CRISPR-Cas9-dep.	This study
CEC5813	SN148 Ca22chrR_C_albicans_SC5314:79361 to 79780:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	ChrRL	*RPS*1	TTATGGAAA TAGCGGATG ATAACCT	CRISPR-Cas9-dep.	This study
CEC5815	SN148 Ca22chrR_C_albicans_SC5314:79361 to 79780:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	ChrRL	*RPS*1	TCAAACGCC GCTACGTAC ACATCTC	CRISPR-Cas9-dep.	This study
CEC5817	SN148 Ca22chrR_C_albicans_SC5314:79361 to 79780:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	ChrRL	*RPS*1	TATTTAGAA CCGCGACA GGCCAAAA	CRISPR-Cas9-dep.	This study
CEC5819	SN148 Ca22chr2_C_albicans_SC5314:252146 to 252382:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr2L	*RPS*1	TAGGGACAG TCAGCGCTA CTTATCT	CRISPR-Cas9-dep.	This study
CEC5821	SN148 Ca22chr2_C_albicans_SC5314:252146 to 252382:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr2L	*RPS*1	TTATTTAGA CGACGATTC CCACCCT	CRISPR-Cas9-dep.	This study
CEC5823	SN148 Ca22chr2_C_albicans_SC5314:252146 to 252382:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr2L	*RPS*1	TGTGGTATC CGGCGACCT AAAGGAC	CRISPR-Cas9-dep.	This study
CEC5825	SN148 Ca22chr2_C_albicans_SC5314:200708 4 to 2007475:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr2R	*RPS*1	TAAATCGGA GGACCCCAT TCCACGT	CRISPR-Cas9-dep.	This study
CEC5827	SN148 Ca22chr2_C_albicans_SC5314:200708 4 to 2007475:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr2R	*RPS*1	TAGAGGTGA CTCGCCCGC AAGTTCT	CRISPR-Cas9-dep.	This study
CEC5829	SN148 Ca22chr2_C_albicans_SC5314:200708 4 to 2007475:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr2R	*RPS*1	TACTCGGGT TGCCGCATA TCTATCG	CRISPR-Cas9-dep.	This study
CEC5831	SN148 Ca22chr4_C_albicans_SC5314:145279 1 to 1453180:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr4R	*RPS*1	TGCAGACCG GTTCGCACT TTATCAG	CRISPR-Cas9-dep.	This study
CEC5833	SN148 Ca22chr4_C_albicans_SC5314:145279 1 to 1453180:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr4R	*RPS*1	TACACGGTG AAGTGCTAT CAGCCAT	CRISPR-Cas9-dep.	This study

CEC5835	SN148 Ca22chr4_C_albicans_SC5314:1452791 to 1453180:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr4R	*RPS*1	TTTAGTATAACGATCGATGTGCAAT	CRISPR-Cas9-dep.	This study
CEC5837	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr5L	*RPS*1	TAGGTCAACGGCGCTACGTACCCAA	CRISPR-Cas9-dep.	This study
CEC5839	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802:: P_{TDH3}-*GFP*-Cd*ARG4*/P_{TDH3}-*BFP*-Cd*HIS1* *RPS1*::CIp-*CmLEU2*-BC-*URA*3	Chr5L	*RPS*1	TCAGCCATGCTCGCGCTATATAGTA	CRISPR-Cas9-dep.	This study

*We did not test to differentiate tandem or heterozygous integrations at *RPS1* locus
(1) Noble *et al.* 2005

Table S2: Plasmids used in this study.

Plasmid	**Usage**	**Ref.**
pCRBluntII-P_{TDH3}-*GFP-ARG*4	Yeast Fluorescent Cassette	(2)*
pCRBluntII-P_{TDH3}-*BFP*-Cd*HIS*1	Yeast Fluorescent Cassette	(2)*
pV1093	CRISPR Cas9 plasmid	(3)*
CIp-P_{TET}-BC-GTW-*URA*3	Barcoding	unpublished
CIp-P_{TET}-BC-*URA*3	Barcoding	This study
CIp10-*LEU*2	Leucine prototrophy	This study
CIp-P_{TET}-BC-*LEU*2-*URA*3	Barcoding and leucine prototrophy	This study

* (2) Loll-Krippleber *et al.* 2015, (3) Vyas *et al.* 2015

Table S3: List of primers

Name		Sequence	Usage
TM50	K7_BFP-GFP_Chr1_Left_F	TACATATAATATAAATTATTTATTCAACTTAC TAACTAACTAACATTCAAACATTCTCAACAA TCAATTGTTGTTATTTATGATCATAACCATTC AACGT**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM51	K7_BFP-GFP_Chr1_Left_R	TTTATATGGGACCCGGATATCGGAAAGCAT TAGTAGTAGTACTAATTCCAATCATTGACAA TATCCGAACTCTTTGCAATAGTTGTTGATAA ATAATGCT**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM52	K7_BFP-GFP_Chr1_Right_F	CTACTGTCATACTTAACCGTGAAAGTAATCG AAATACATCTCTCATACTACAATAGTTAGGT TGTAGGTACTATTGACTTTTGGGTTCTTAAC AGGTGCA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM53	K7_BFP-GFP_Chr1_Right_R	ATGATTACGAGAAACGCCTGTGCGTTTGAA GCAGTACTGGACTTGGTGTAGGTTTACTTT TAAGAACACAATACATAGCACATGTTGGTC ACGTGAACAC**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM54	K7_BFP-GFP_Chr2_Left_F	ATCTCCTTCCTCCTCCACCTCTTAGTCTTAG TCTTCTAACCGTGATAGTCAACAATTAGTAC CATTGTTCTATTCCTTCATGACGTTATTTCTT TCCTGA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM55	K7_BFP-GFP_Chr2_Left_R	ACAACGACGACGCCGTGGACAATACACGG AGTATTATTGAACGCCGAACAAATTACTTTG AGAATGTATGTTTAAAACTTTGCAGCTGCAA CCTACAATG**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM56	K7_BFP-GFP_Chr2_Right_F	AATGGACGGGAACTGACGGAGAGAAGGAG TTCAGGACTAGTGAATAGTTTCCAGTTGAAT TCAAGTCCGGATATCCCTGTAGCAATTGAA ATGAAGAACA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM57	K7_BFP-GFP_Chr2_Right_R	AATATCATAATGATATCATTATATTGCAAAG ATTACTAGGATAAGGCATGAATTCCAATTG GCAATTTCATGATCTACGATACTTGTTATTG TTGAGGAT**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM58	K7_BFP-GFP_Chr3_Left_F	ATCAACTTTGTTGGCATAGATATATTGTTAG TACTATAGTTTAATTCCAGCCGCATCTAAAG CAGCATTCTGACAACACAAAACATCACCAA GAGTTTCA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM59	K7_BFP-GFP_Chr3_Left_R	TTAAACAAACATGCTATATGCATACACTCCC TAAATCTATATTGTTAACATAATGTAAGTTC CCATGGGGGAAAATAGGAAAGAAGACGGG GTGAATGTA**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM60	K7_BFP-GFP_Chr3_Right_F	TATACTAGGTGAATAGTACCAAATTGATTCC AAAGAACTTTACTTCAAGTGGCTATATGGTA CTGTTATTATAGCTCCAGTAACAACTTATCA TCAAATC**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM61	K7_BFP-GFP_Chr3_Right_R	TCCTGGTTGTCGTAGTAGCTGGCACTCAAC GTACTTGTCTTCAAAAATATTATTATTTAACT ACTCTATGAGTAATCGTTATTATAAGTCTTT TCATCCA**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM62	K7_BFP-GFP_Chr4_Left_F	TCCGTCAGTTAAGAAAAACCTTATTTTTAGG TTAGTACCGGTTCACTTTAACGCAAACTTAG TTCTATGGGTGGCCTAATTCACACCAAGAA ATACACAT**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM63	K7_BFP-GFP_Chr4_Left_R	CTGGCGACCTTTGCCCCTGACGCCTTGTTT GTTAATCCGTTTCAAAGTTAATTGTAGCAAC AATAACAATAATAATTAGTGATTGATTGAAT GCTAGGAC**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM64	K7_BFP-GFP_Chr4_Right_F	TATATATTTCTGGGCAATGCAGCAATTCTCG GATATCACCGAAAAAAAAGATCTTAGCGGG CACGACACGACTCTCTTGATATAAGCGAAT TTTCAGTAT**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM65	K7_BFP-GFP_Chr4_Right_R	TCTCTATACGAGTTAAGAGTAGTCTTACAAT AGTCTATAGATAGAATTTCAGACCTTTTTGT GTGGGTATTGCCGAAATTCTTTTTCCAGAA GATGACGA**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM66	K7_BFP-GFP_Chr5_Left_F	TTTATAATATCAATGGGGGAAATGTACGTCA TAAATGATATCTCTTCAATACTGTTCAGATT GTACAATTGCGAAATCTGTACGCAACTGTC ATATTACA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM67	K7_BFP-GFP_Chr5_Left_R	TTATTTCCTCTTTAGTCCAATTTTGTTATTGT TGTTATTTAATTAACTTAACTAATTAAGTAAC CACAATAATACTAGTGGTGACGACAGCGGA AGTGGA**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms

TM68	K7_BFP-GFP_Chr5_Right_F	CAATAAAGATCATCGTATAATATTATAAATG TTAGATTCACTCTTGTTGGTTTTGACCATTA AGTATTGAAACAGCAACATCATCCTCATATC AAGTGCA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM69	K7_BFP-GFP_Chr5_Right_R	ACTAGATTGATGGATTAATCGAGAGCCAAA ACAAACAAGCAAGTAAGGGGGGAGGAAGA AGGAAAGAGAGGAGTAAAAACTCTAAAACA TGATTAGACGA**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM70	K7_BFP-GFP_Chr6_Left_F	TCTGAATTAAACTTTTTCTTTCAAATAAAGTA GAGAACGGAAACTTTCCAGTAGACTTTGAT TGGATAAACCGATTCTAGATCAGTAGGAGT AAAATTCT**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM71	K7_BFP-GFP_Chr6_Left_R	ATTGTATACAGAGTTCGAGTTTCCTTGCTTG AATATTTGTAGATTGTAAGGATGTGACTACA TCTAAGACGGTAGTTGTACTAAATTGTTTGT TATATCT**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM72	K7_BFP-GFP_ChrR_Right_F	ACTAAATTTACAAATGAAAATACATGGAACC ATACATGGTAAACAATTCAGTATTGTGAAAC GTTTCCTGATAGAAGGACAACTTTACTATTC AGTAAAG**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM73	K7_BFP-GFP_ChrR_Right_R	AATTATAATTTAGAATATGGCAGTTTTAGGA GACAATATCCCTCAAGATGCTTGAGGCGGT CCAATGTCATCTTTGTTGTTCTAGATACTTC TTCACTCT**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM74	Verif_K7_BFP-GFP_C1_Left_F	TCATTCATCGTCGTCGTCGT	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM75	Verif_K7_BFP-GFP_C1_Right_F	TTGCTTGGCAGGTTAGTGGA	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM76	Verif_K7_BFP-GFP_C2_Left_F	TTGGTTCATTGGCGGTGGTA	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM77	Verif_K7_BFP-GFP_C2_Right_F	AAACATGCACCACGAGGGAT	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM78	Verif_K7_BFP-GFP_C3_Left_F	TGCCAATTGACCCTTCACCT	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM79	Verif_K7_BFP-GFP_C3_Right_R	ACTGGCCATTGCTGTTGCTA	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM80	Verif_K7_BFP-GFP_C4_Left_F	TTTCCGACAATCACCGATGC	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM81	Verif_K7_BFP-GFP_C4_Right_F	TGGGCTCAAGATTCTGCTCA	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM82	Verif_K7_BFP-GFP_C5_Left_F	TTCTTGTTTCGCCAAGCTGC	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM83	Verif_K7_BFP-GFP_C5_Right_F	GCGTGTGGAAAAACTGCTGT	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM84	Verif_K7_BFP-GFP_C6_Left_F	AAATCCCGCCTTGACTCTGA	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM85	Verif_K7_BFP-GFP_CR_Left_F	AGCCCGAAACTATCACCACT	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM86	Verif_K7_BFP-GFP_CR_Right_R	TGGAAACCATGGACAGCACA	3' verification of integration, Collection of strains bearing

			BFP/GFP System on all chromosomes both arms
TM87	Verif_K7_BFP-GFP_C1_Left_R	GACCCAGGACAAACTGAGCA	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM88	Verif_K7_BFP-GFP_C1_Right_R	TCCCATCCGGCTTTATTGCT	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM89	Verif_K7_BFP-GFP_C2_Left_R	AAATGGGATTGTGGAGGGTG	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM90	Verif_K7_BFP-GFP_C2_Right_R	TGTCGCCGAACGCAATCTAT	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM91	Verif_K7_BFP-GFP_C3_Left_R	ACTTGGAGAGATGCGTCGAC	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM92	Verif_K7_BFP-GFP_C3_Right_F	ACATTGGTTGGGGATGTGGA	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM93	Verif_K7_BFP-GFP_C4_Left_R	AAAAGGGTTAGATCGGGCCG	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM94	Verif_K7_BFP-GFP_C4_Right_R	CCAAGAAGGTTTGGGCAACT	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM95	Verif_K7_BFP-GFP_C5_Left_R	ACTTGATCACCAAACGGGCA	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM96	Verif_K7_BFP-GFP_C5_Right_R	ATGTACACGCACGTCTCCAC	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM97	Verif_K7_BFP-GFP_C6_Left_R	TTCTACAACTTCCGCTCGGC	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM98	Verif_K7_BFP-GFP_CR_Left_R	ACCCCCGTTGATCCTACTGT	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM99	Verif_K7_BFP-GFP_CR_Right_F	GTGTGTGCGGCATTTTGAGT	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM100	K7_BFP-GFP_ChrR_Left_F	TCGGATGAAATTCTCACCACACACAGGTTG CTCTTCTTCTTTGGGCAATTTATGCAAAGGT GATGGCTCACCAATTATACATTACAGATATG CACCATTC**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM101	K7_BFP-GFP_ChrR_Left_R	ATAATAACAATAGTAATGTTTTATAAAAGAA ATTGAGCATCGGCGGCCAAGGTAAATGTTA CTAGAGAGCTTTTAATGTGCGCTGTAGGTC ATCATAGGT**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM107	K7_BC-URA3_Chr1_F2	AATTCCAATGCGGTTGCAGCATCTAAACCA CGAGTAGAATTGTCCCAGCATTGAATTGAT GCCTGCACCAATGTAACTTCGGCAATGGAA AGACGTTTAC**TTGCACGGTGAAGGTTCAAC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms. For BC integration on Chr4
TM108	K7_BC-URA3_Chr1_R2	TACAAGCCAGGTAAATTAGGTGTTGCTTATA GAAACCTTCGAGTTTATGGTGATGCTATTG AAAGTGATTATCAAACAACCGTATCGAACG GGGTATTGA**TTTGACGTTGGAGTCCACGT**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms. For BC integration on Chr4
TM109	K7_Leu2onChr4_F	AATTCCAATGCGGTTGCAGCATCTAAACCA CGAGTAGAATTGTCCCAGCATTGAATTGAT	For production of transformation cassettes, Collection of strains

		GCCTGCACCAATGTAACTTCGGCAATGGAA AGACGTTTAC**GACGTTGGAGTCCACGTTCT**	bearing BFP/GFP System on all chromosomes both arms. For LEU2 integration on Chr4
TM110	K7_Leu2onChr4_R	TACAAGCCAGGTAAATTAGGTGTTGCTTATA GAAACCTTCGAGTTTATGGTGATGCTATTG AAAGTGATTATCAAACAACCGTATCGAACG GGGTATTGA**CACGGTGAAGGTTCAACCGA**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms. For LEU2 integration on Chr4
TM111	BG-Collec-verifChr4-TG(GCC)2_F	GCCATGGCAATATTACCCGC	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms. For BC & LEU2 integration on Chr4
TM112	BG-Collec-verifChr4-TG(GCC)2_R	GGCATCACAAGCAGAAGGAC	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms. For BC & LEU2 integration on Chr4
TM117	K7_BGcol_C1L_cas9_F	TAGTGTGATTTCTGAAACACATTATCTGATA TAACTTATAATAATAATCTGTCTATCTCTTTC TATATTTTTTAGTTTGTGATTTGCGTTTTTTC CCACC**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms with CRISPR cas9
TM118	K7_BGcol_C1L_cas9_R	GTTTGTTTCTTTCGTTTGTGATATCCAAAAA TAGGGAAAAAAGCAGCAATAAACTCTTCTTT ATTTTAACTTTCTGTTATTAATTTTATTCTAG ATCCTT**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms with CRISPR cas9
TM119	K7_BGcol_C1R_cas9_F	AATTCACAGCTAGGCTGGTTACGAAGATTT GTCCATAAGTTTTAGTGAGAGGTTATTTTTA AAGATTATTTCTGCTTGCTTTAAAGATTAAA TCTCGCGT**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms with CRISPR cas9
TM120	K7_BGcol_C1R_cas9_R	CGTAATTGAAACTAAAATAGACAAAAATTAT CTCAAAATTGTAACAATTGTAAAAAAAAATA ATATGGCCAAACGGGAAACCAAAGTTGCTA ACCAACTA**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms with CRISPR cas9
TM121	K7_BGcol_C5R_cas9_F	GTAATTTATTTAGGACAAATTCATAATTCTCT ACAAAAATGAGATATCTGTACGTATATAAAG ATTGTAGAAGATCTTAAATGTAAGTAAAAGT CAGTGG**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms with CRISPR cas9
TM122	K7_BGcol_C5R_cas9_R	GTTTTCCTTTTTAATCTTTCTATAACCTTAAA GTTTAGTTTAGTTTTTTTTTTTTCTCTTACCC ATTATCGTTGTATAGACGAAACATTAAACTA TTGTG**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms with CRISPR cas9
TM123	K7_BGcol_C6L_cas9_F	GAAGCCAGTGCTATATAATCTGTGGTGATG ACGTAGTAGCAAGATAAGCGATGAAAGAAG TTGAAATTGTGACAGGTATGCTGGCTGATT TGCAAGGCTA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms with CRISPR cas9
TM127	sgRNA_F_Chr7L	TGGCGATGCTACTGATAAGAGTTTTAGAGC TAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM128	SNR52_R_Chr7L	TCTTATCAGTAGCATCGCCACAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM129	sgRNA_F_ChrRL	CAAGAAACAGGGGCAAGAAGGTTTTAGAGC TAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM130	SNR52_R_ChrRL	CTTCTTGCCCCTGTTTCTTGCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM131	sgRNA_F_Chr1L	CAAGATTAGTCTAACAACCCGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM132	SNR52_R_Chr1L	GGGTTGTTAGACTAATCTTGCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM133	sgRNA_F_Chr1R	ATACAAAGAAGGATTGGGAGGTTTTAGAGC TAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence

TM134	SNR52_R_Chr1R	CTCCCAATCCTTCTTTGTATCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM135	sgRNA_F_Chr5R	ATTGGTTCTAGATTCTACCGGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM136	SNR52_R_Chr5R	CGGTAGAATCTAGAACCAATCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM151	New-sgRNA-Chr6L-F	TGTATGTGATTTATGCAGGAGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM152	New-SNR52-Chr6L-R	TCCTGCATAAATCACATACACAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM153	sgRNA-Chr2L-F	GGAAAGAAATAACGTCATGAGTTTTAGAGC TAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM154	SNR52-Chr2L-R	TCATGACGTTATTTCTTTCCCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM155	sgRNA-Chr2R-F	TTGCAAAGATTACTAGGATAGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM156	SNR52-Chr2R-R	TATCCTAGTAATCTTTGCAACAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM157	sgRNA-Chr3L-F	CATAATGTAAGTTCCCATGGGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM158	SNR52-Chr3L-R	CCATGGGAACTTACATTATGCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM159	sgRNA-Chr4R-F	AAATATAATCATTATACGAAGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM160	SNR52-Chr4R-R	TTCGTATAATGATTATATTTCAAATTAAAAAT AGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM161	sgRNA-Chr4L-BC-F	AGTTGTGCTAGGAAGACCAGGTTTTAGAGC TAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM162	SNR52-Chr4L-BC-R	CTGGTCTTCCTAGCACAACTCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM163	newK7_BGcol_C6L_cas9_R	ATCCACCTTAATTTTGAGTTGGTTTCTTCTT TCAATTATTACTAACCATTACTATTCCTCGG CTTCACTATAATAATAATAATATTACAACTTG ATAAAC**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM164	newK7_BGcol_C2L_cas9_F	TGTTTTAGATTCTAACTTAAACCTCCTTACT CATTACTCTTTCATTTTCTATCATTTCTTGGA TGTGACTAAAAATCTTACCACCACCGCCAT CTCCTTC**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM165	newK7_BGcol_C2R_cas9_R	AGATGTGAACACTTGGGACCACTACTTTGT GTATTTTCGTGAGTAATCGGCACTGATCCC GAATTATTCGTACAAAGGCTTACTTGTTATG AGCATCCAC**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM166	newK7_BGcol_C3L_cas9_R	ACAATAAATAGTAATATTATTACATTACAATA AAATAAAATGAAACAGGTCAAGAAAAAAATA GTACTCTAAAATGTACATGAATTATTAGCTG ATTAAT**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM167	newVerif_K7_BG_C6L_F	ATTCTCAACCGAGCTTCGCA	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM168	newVerif_K7_BG_C6L_R	AAGAAGGGGAATTGCAGTATGT	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM171	sgRNA-Chr5L-F	TTATGCATATAATATATCTCGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence

TM172	SNR52-Chr5L-R	GAGATATATTATATGCATAACAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM173	SNR52_F1	AAGAAAGAAAGAAAACCAGGAGTGAA	Construction of guide RNA primers for CRISPR cas9, Forward primer for amplification of SNR52 promoter
TM174	sgRNA_R2	ACAAATATTTAAACTCGGGACCTGG	Construction of guide RNA primers for CRISPR cas9, Reverse primer for amplification of sgRNA scaffold
TM175	sgRNA_N3	GCAGCTCAGTGATTAAGAGTAAAGATGG	Construction of guide RNA primers for CRISPR cas9, Forward and reverse nested primers for third round PCR for construction of sgRNA expression cassette
TM176	SNR52_N3	GCGGCCGCAAGTGATTAGACT	Construction of guide RNA primers for CRISPR cas9, Forward and reverse nested primers for third round PCR for construction of sgRNA expression cassette
TM177	CaCas9/for	ATCTCATTAGATTTGGAACTTGTGGGTT	Forward and reverse primers for amplification of *CaCas9* cassette
TM178	CaCas9/rev	TTCGAGCGTCCCAAAACCTTCT	Forward and reverse primers for amplification of *CaCas9* cassette
TM115	GFP-end-F	CCAATTGGTGATGGTCC	Sequence GFP end junction 3' for integ. Verif in CIp30
TM106	GFP_Rev	CGGAGACAGAAAATTTGTGACC	Used to validate 5' junction when integrating pTDH3-ARG4-GFP cassettes, internal oligo.
RLO066	TDH3p-BFP-R	TCTTTAATCAATTCAGACATTTTAATAAAGCT TCTGCAGG	Used to validate 5' junction when integrating pTDH3-HIS1-BFP cassettes, internal oligo.
AF44	Leu2_down	GCTACTGAAGTTGGTGACGCGATTGT	Used to validate integrating of LEU2 cassettes.
ML42	CdHIS1-detect-R	TGCCCTTCTACCTGGAGTAATGGT	Used to validate 3' junction when integrating pTDH3-HIS1-BFP cassettes, internal oligo.
CipUL	CipUL	ATACTACTGAAAATTTCCTGACTTTC	Used to validate integrating at *RPS*1 locus.
CIpUR	CIpUR	ATTACTATTTACAATCAAAGGTGGTC	Used to validate integrating at *RPS*1 locus.

Table S4: Sequencing depth of C. albicans strains

	Strain	Cov	SD	Avg.	SD
CRISPR-Cas9-free Transformation Protocol	CEC5752	25.37	46.4	36.5	16.5
	CEC5753	27.64	52.5		
	CEC5754	30.84	66.6		
	CEC5755	26.04	63.8		
	CEC5756	28.41	62.8		
	CEC5757	50.91	85.8		
	CEC5758	21.26	54.8		
	CEC5759	27.93	56.9		
	CEC5760	17.97	45.1		
	CEC5761	13.13	30		
	CEC5762	34.86	64.6		
	CEC5763	36.32	67.5		
	CEC5764	73.12	94.4		
	CEC5765	28.61	76.4		
	CEC5766	10.07	22.6		
	CEC5767	21.42	47.1		
	CEC5768	16.27	45.4		
	CEC5769	30.53	50.5		
	CEC5770	48.51	84.1		
	CEC5771	65.63	84.8		
	CEC5772	58.31	80.9		
	CEC5773	52.15	85.8		
	CEC5774	42.86	76		
	CEC5775	36.24	73.7		
	CEC5776	48.05	79.2		
	CEC5777	70.57	84.2		
	CEC5778	36.17	72.7		
	CEC5779	36.86	66.9		
	CEC5780	29.67	65.7		
	CEC5781	49.31	75		

	Strain	Cov	SD	Avg.	SD
CRISPR-Cas9-independent Transformation Protocol	CEC5783	9.12	20.6	13.4	6.4
	CEC5785	10.69	26.6		
	CEC5787	13.75	35.2		
	CEC5789	11.78	28.5		
	CEC5791	10.35	24.9		
	CEC5793	26.84	62.3		
	CEC5795	8.56	15.3		
	CEC5801	8.99	19.5		
	CEC5803	12.67	32.8		
	CEC5805	10.66	25.4		
	CEC5807	13.77	38.9		
	CEC5809	9.46	20.2		
	CEC5811	8.95	16		
	CEC5813	9.7	22.7		
	CEC5815	14.67	40.5		
	CEC5817	13.92	37.5		
	CEC5819	7.66	13.1		
	CEC5821	7.44	12.8		
	CEC5823	10.3	24.5		
	CEC5825	8.72	18.4		
	CEC5827	25.09	33.9		
	CEC5829	22.8	43.5		
	CEC5831	18.85	41.5		
	CEC5833	21.16	47.3		
	CEC5835	29.88	36.8		
	CEC5837	7.96	13.6		
	CEC5839	7.76	15.4		

Table S5: Analysis of average ABHet values per chromosome. *

	Strain	Chromosome							
		1	2	3	4	5	6	7	R
CRISPR-Cas9-free	CEC5752	0.41	0.37	0.37	0.59	0.49	0.53	0.53	0.41
	CEC5753	0.49	0.49	0.49	0.49	0.49	0.5	0.53	0.49
	CEC5754	0.49	0.5	0.49	0.51	0.5	0.5	0.53	0.5
	CEC5755	0.49	0.5	0.5	0.5	0.5	0.5	0.52	0.5
	CEC5756	0.49	0.37	0.5	0.5	0.49	0.5	0.52	0.5
	CEC5757	0.5	0.51	0.5	0.5	0.51	0.52	0.54	0.51
	CEC5758	0.49	0.5	0.49	0.59	0.49	0.5	0.52	0.5
	CEC5759	0.49	0.49	0.49	0.49	0.63	0.49	0.52	0.5
	CEC5760	0.48	0.49	0.49	0.48	0.48	0.49	0.52	0.49
	CEC5761	0.47	0.48	0.48	0.47	0.48	0.48	0.5	0.48
	CEC5762	0.49	0.49	0.48	0.49	0.5	0.49	0.53	0.5
	CEC5763	0.49	0.5	0.48	0.5	0.56	0.5	0.53	0.5
	CEC5764	0.36	0.64	0.63	0.48	0.61	0.52	0.4	0.43
	CEC5765	0.5	0.37	0.37	0.5	0.5	0.43	0.53	0.37
	CEC5766	0.47	0.48	0.47	0.47	0.46	0.48	0.48	0.48
	CEC5767	0.48	0.49	0.49	0.48	0.49	0.49	0.52	0.49
	CEC5768	0.41	0.49	0.49	0.48	0.49	0.49	0.52	0.49
	CEC5769	0.49	0.49	0.49	0.49	0.5	0.5	0.52	0.49
	CEC5770	0.5	0.51	0.5	0.5	0.51	0.51	0.54	0.5
	CEC5771	0.5	0.51	0.51	0.5	0.51	0.51	0.53	0.51
	CEC5772	0.5	0.62	0.38	0.5	0.5	0.63	0.62	0.37
	CEC5773	0.51	0.39	0.39	0.51	0.5	0.44	0.54	0.39
	CEC5774	0.47	0.5	0.5	0.5	0.52	0.5	0.54	0.5
	CEC5775	0.63	0.37	0.37	0.5	0.55	0.62	0.62	0.51
	CEC5776	0.36	0.52	0.37	0.62	0.55	0.62	0.64	0.63
	CEC5777	0.5	0.51	0.51	0.51	0.51	0.51	0.54	0.51
	CEC5778	0.49	0.5	0.5	0.5	0.51	0.51	0.53	0.5
	CEC5779	0.61	0.61	0.6	0.61	0.5	0.61	0.53	0.59
	CEC5780	0.49	0.5	0.49	0.5	0.5	0.51	0.53	0.5
	CEC5781	0.49	0.5	0.5	0.5	0.64	0.51	0.53	0.5

	Strain	Chromosome							
		1	2	3	4	5	6	7	R
CRISPR-Cas9-independent	CEC5783	0.47	0.48	0.48	0.47	0.48	0.48	0.57	0.47
	CEC5785	0.47	0.47	0.48	0.47	0.47	0.47	0.49	0.47
	CEC5787	0.47	0.48	0.48	0.48	0.48	0.49	0.5	0.48
	CEC5789	0.47	0.47	0.48	0.47	0.47	0.49	0.48	0.47
	CEC5791	0.46	0.47	0.48	0.47	0.47	0.47	0.48	0.47
	CEC5793	0.49	0.49	0.5	0.49	0.49	0.5	0.52	0.5
	CEC5795	0.46	0.46	0.47	0.46	0.46	0.48	0.49	0.47
	CEC5801	0.46	0.46	0.47	0.46	0.47	0.48	0.47	0.47
	CEC5803	0.47	0.47	0.48	0.47	0.48	0.48	0.49	0.38
	CEC5805	0.46	0.47	0.48	0.47	0.47	0.47	0.49	0.48
	CEC5807	0.47	0.47	0.48	0.48	0.53	0.49	0.51	0.49
	CEC5809	0.46	0.47	0.48	0.46	0.54	0.47	0.47	0.47
	CEC5811	0.58	0.46	0.47	0.46	0.53	0.49	0.49	0.47
	CEC5813	0.47	0.34	0.48	0.47	0.47	0.48	0.5	0.52
	CEC5815	0.48	0.49	0.49	0.48	0.48	0.49	0.52	0.48
	CEC5817	0.48	0.48	0.49	0.48	0.47	0.49	0.49	0.49
	CEC5819	0.46	0.46	0.48	0.46	0.46	0.46	0.48	0.46
	CEC5821	0.46	0.47	0.48	0.46	0.47	0.46	0.46	0.47
	CEC5823	0.47	0.48	0.48	0.47	0.47	0.47	0.5	0.48
	CEC5825	0.46	0.46	0.48	0.47	0.47	0.47	0.46	0.47
	CEC5827	0.46	0.5	0.46	0.46	0.48	0.5	0.49	0.48
	CEC5829	0.47	0.48	0.49	0.48	0.49	0.49	0.52	0.49
	CEC5831	0.48	0.48	0.48	0.48	0.48	0.49	0.52	0.49
	CEC5833	0.48	0.49	0.49	0.48	0.49	0.49	0.51	0.49
	CEC5835	0.48	0.49	0.46	0.48	0.48	0.49	0.52	0.48
	CEC5837	0.45	0.45	0.47	0.46	0.36	0.47	0.48	0.46
	CEC5839	0.46	0.47	0.47	0.46	0.46	0.47	0.48	0.46

4n AAAB	3n AAB	2n AB	3n ABB	4n ABBB
<0.29	0.3-0.44	0.45-0.55	0.56-0.69	0.7-1

*in red font are indicated LOH events

Table S6: Identification and characterization of genomic changes in the collection of 57 sequenced C. albicans strains.

Available: https://doi.org/10.1128/mSphere.00620-20

Chapter 5: How *C. albicans* chromosomes respond to stress exposure.

Contains: Article in preparation, to be submitted summer 2020

Multiple stochastic parameters influence genome dynamics in a heterozygous diploid eukaryotic model

Timea MARTON[1,2], Christophe D'ENFERT[1], Mélanie LEGRAND[1]

[1] Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, France
[2] Université Paris Diderot, Sorbonne Paris Cité, Paris, France

Abstract

The heterozygous diploid genome of *Candida albicans* displays frequent genomic rearrangements, in particular loss-of-heterozygosity (LOH) events, which can be seen on all eight chromosomes and affect both laboratory and clinical strains. LOH, which are often the consequence of DNA damage repair, can be observed upon stresses reminiscent of the host environment, and result in homozygous regions of various sizes depending on the molecular mechanisms at their origins. LOH are frequent and ubiquitous, though little is known on how they are regulated in *C. albicans*. In diploid *Saccharomyces cerevisiae*, LOH facilitate the passage of recessive beneficial mutations through Haldane's sieve allowing rapid evolutionary adaptation. This also appears to be true in *C. albicans*, where the full potential of an adaptive mutation is often only observed upon LOH, as recently illustrated in the case of antifungal resistance and niche adaptation. To understand the genome-wide dynamics of LOH events in *C. albicans*, we constructed a collection of 15 strains, each one carrying a LOH reporter system on a different chromosome arm. This system involves the insertion of two fluorescent marker genes in a neutral genomic region on both homologs, allowing spontaneous LOH events to be detected by monitoring the loss of one of the fluorescent marker using flow cytometry. Using this collection, we observed significant LOH frequency differences between genomic loci in standard laboratory growth conditions; however, we further demonstrated that comparable heterogeneity was also observed for a given genomic locus between independent strains. Additionally, upon exposure to stress, three outcomes could be observed in *C. albicans*, where individual strains displayed increases, decreases or no effect of stress in terms of LOH frequency. Our results argue against a general stress response triggering overall genome instability. Indeed, we showed that heterogeneity of LOH frequency in *C. albicans* is present at various levels, inter-strain, intra-strain, and inter-chromosomes, suggesting that LOH events may occur stochastically within a cell, though genetic background potentially impacts genome stability in terms of LOH throughout the genome in both basal and stress conditions. This heterogeneity in terms of genome stability may serve as an important adaptive strategy for predominantly clonal human opportunistic pathogen *C. albicans,* by quickly generating a wide-spectrum of genetic variation combinations potentially permitting subsistence in a rapidly evolving environment.

Introduction

The genome of the human opportunistic fungal pathogen *C. albicans* is highly tolerant to genomic rearrangements, in particular loss of heterozygosity (LOH) events. This type of genomic rearrangements is ubiquitously found across sequenced chromosomes of both clinical and laboratory strains. LOH are often the result of DNA double-strand breaks (DSBs) as these DNA damages are preferentially repaired through homologous recombination (HR)-mediated repair mechanisms in *C. albicans* (Legrand *et al.* 2007; Feri *et al.* 2016; Vyas *et al.* 2018). LOH size is often dictated by the type of HR mechanism used, with homozygosis that can be limited to a given locus, or span across a chromosomal arm or the entire chromosome (Malkova *et al.* 2000). By increasing the number of DNA breaks, using nucleases or exposure to DNA damaging agents mimicking stress, the frequency of LOH has been shown to increase in *C. albicans* (Forche *et al.* 2011; Feri *et al.* 2016; Vyas *et al.* 2018). Additionally, genomic instability resulting from perturbations of the DNA repair machinery also augments the occurrence of LOH (Ciudad *et al.* 2004, 2020; Legrand *et al.* 2007; Andersen *et al.* 2008; Loll-Krippleber *et al.* 2015; Tutaj *et al.* 2019). Interestingly, Forche *et al.* have shown that various *in vitro* stress conditions increase the frequency of LOH and influence the molecular mechanisms behind LOH occurrence (Forche *et al.* 2011).

Certainly, LOH play an important role in the biology of *C. albicans* as (i) they have been shown to occur *in vivo*, both during murine intravenous infection (Forche *et al.* 2009a) and, in gastrointestinal (Ene *et al.* 2018) and oropharyngeal (Forche *et al.* 2018) colonization models and (ii) they are required at the mating-type like (*MTL*) locus to make sterile *MTL* heterozygous cells into mating competent *MTL*a/a or *MTL*α/α homozygous cells. In terms of their role in adaptation, LOH have been associated to adaptive phenotypes such as antifungal resistance (Coste *et al.* 2007) and increased or reduced virulence (Tso *et al.* 2018; Liang *et al.* 2019). In diploid organisms, LOH may contribute to rapid evolutionary adaptation by revealing recessive beneficial alleles (Gerstein *et al.* 2014; James *et al.* 2019).On the other hand, these genomic rearrangement events may also unveil recessive deleterious or lethal alleles (Feri *et al.* 2016; Marton *et al.* 2019). Therefore, within a predominantly clonal population, such genomic rearrangements contribute significantly to the purging of detrimental alleles and to the acquisition of genetic variation, thus rendering the understanding of LOH events all the more interesting.

In *C. albicans*, LOH events have mostly been described as markers of genomic instability, having for consequences the potential to generate new phenotypic traits. For instance, the increased sensitivity to the alkylating agent methyl methanesulfonate (MMS) that

characterizes some derivatives of the reference strain SC5314 has been explained by an LOH event on the right arm of chromosome 3 (Chr3R), encompassing the *MBP1* locus, which encodes a component of a transcription complex involved in regulating G1/S genes important for DNA repair (Ciudad *et al.* 2016). Likewise, a growth defect when proteins are the only available nitrogen source was also associated to an LOH event comprising one of the *SAP2* alleles (secreted aspartic protease) on the left arm of ChrR (Dunkel and Morschhäuser 2011). More recently, due to the convenience of whole-genome sequencing, evolutionary studies have highlighted the frequent occurrence of small LOH events over short time scales (Ene *et al.* 2018). Besides genome sequencing, two major tools are commonly used with regards to monitoring the occurrence of LOH events in *C. albicans*. The first one, known as the *GAL1*/*URA3* system, uses the native *C. albicans GAL1* gene (located on the left arm of Chr1), encoding for a galactokinase implicated in galactose metabolism and the commonly used *URA3* auxotrophic marker, encoding orotidine-5'-phosphate decarboxylase involved in the biosynthesis of the nucleoside uridine. An artificial heterozygous locus is generated at the native *GAL1* locus by the replacement of one of the *GAL1* alleles with the *URA3* gene. LOH events at this locus are monitored by growing cells on media containing 2-deoxygalactose (2-DG) or 5-fluoroorotic acid (5-FOA), for *GAL1* and *URA3* counter-selection, respectively (Boeke *et al.* 1984; Gorman *et al.* 1992). With this system, evaluation of bidirectional LOH events (allowing to monitor loss of both haplotypes) is restricted to the endogenous *GAL1* locus. Nevertheless, unidirectional LOH events (monitoring loss of a single haplotype) can still be assessed at various loci throughout the genome by generating heterozygous *URA3* loci (Forche *et al.* 2011). The second system is a fluorescence-based system, known as the BFP/GFP LOH reporter system, and is engineered by integrating an artificial fluorescent heterozygous locus, where one of the chromosome homolog possesses the BFP-encoding gene and the other the GFP-encoding gene. Heterozygous cells at the locus of interest are therefore bifluorescent and LOH occurring at this locus can easily be monitored by tracking the loss of either the GFP or BFP signals using flow cytometry. Unlike the *GAL1/URA3* system, there is no locus restriction associated to the use of this system and high-throughput screening can easily be implemented with this fluorescence-based system. The BFP/GFP LOH reporter system has already allowed LOH screens to be performed with overexpression mutant collections in order to identify key regulators of genome stability in *C. albicans* (Loll-Krippleber *et al.* 2015). Overall, previous studies relying on both systems have estimated that LOH in standard laboratory growth conditions occur at a frequency of 10^{-4} in *C. albicans* (Forche *et al.* 2011; Loll-Krippleber *et al.* 2015; Feri *et al.* 2016; Marton *et al.* 2019).

Few studies have been specifically aimed to better understand LOH dynamics in *C. albicans.* In the study by Forche *et al*. 2011, the authors analyzed the appearance of

unidirectional LOH events using a collection of heterozygous KO mutants of genes located in 18 distinct genomic loci. For each locus, they replaced one of the two alleles with the *URA3* auxotrophic marker and recorded loss of *URA3* through the appearance of 5-FOA^R colonies. These experiments concluded that LOH rates in standard laboratory conditions are comparable between genomic loci and that the rates of LOH show a medium correlation with the distance between the *URA3* marker and the centromere. However, these estimations are based on loss of a single haplotype, while we now recognize that chromosome homozygosis biases for numerous chromosomes exist in *C. albicans* due to the presence of recessive lethal or deleterious alleles on certain homologs, where a given haplotype is preferentially lost (Feri *et al.* 2016; Marton *et al.* 2019). Additionally, the authors showed that different stress conditions induce augmentations of LOH at five genomic loci and that the extent of the increase correlates with the intensity of the stress (Forche *et al.* 2011). However, the coordination of LOH events is yet unclear as the overall dynamics of LOH in *C. albicans* has not been thoroughly investigated. Additionally, the mining of sequencing data from *in vivo*- and *in vitro*-evolved laboratory strains and clinical strains have highlighted that LOH break points are more frequently observed in proximity of repeat regions, suggesting the presence of hotspots for LOH throughout the genome (Ene *et al.* 2018; Wang *et al.* 2018; Todd *et al.* 2019). Altogether, it is still unknown if certain genomic loci exhibit higher LOH frequencies and what their biological impact is on *C. albicans* biology, potentially favoring adaptation.

The phenotype associated with a given genetic variant is undoubtedly influenced by the genetic background of the strain. In *C. albicans,* various observable phenotypic differences have been reported across different genetic backgrounds regarding biofilm formation, growth rates or antifungal resistance assays (Li *et al.* 2003; Hirakawa *et al.* 2015; Ropars *et al.* 2018). Recently, a study illustrated the importance of genetic background in biofilm formation, as diverse *C. albicans* genetic backgrounds exhibit clear phenotypic differences upon KO of key regulators of biofilm formation, indicating circuit diversification among members of this species (Huang *et al.* 2019). Depending on the genetic background of a strain, a mutation may be favorable in one environment but neutral or detrimental in another. Additionally, phenotypic heterogeneity has also been described upon growth of *C. albicans* in fluconazole, where population subsets display contrasting phenotypes within the same genetic background (Rosenberg *et al.* 2018). Altogether, genetic background-specific effects on phenotypic expression suggest a phenotypic continuum, which could complexify genotype-phenotype association studies (Fournier and Schacherer 2017). In *C. albicans*, such heterogeneity in terms of appearance of genomic plasticity events has not yet been described. Though during the course of our investigation, a study conducted by Gerstein and Berman suggested that genetic background also influences the appearance of aneuploidy events and impacts the

level of heterogeneity in terms of genome size within a single and between multiple genetic backgrounds during evolution experiments upon exposure to fluconazole (Gerstein and Berman 2020), highlighting variations in the potential for adaptability between genetic backgrounds. However, heterogeneity in terms of LOH frequency in *C. albicans* has not yet been investigated.

Here, we used *C. albicans* as a model organism to investigate heterogeneity in terms of genome stability in a eukaryotic diploid. We aimed to study LOH frequency across its genome in order to investigate genome-wide LOH dynamics and assess heterogeneity in LOH frequencies in both standard laboratory growth conditions and upon stress exposure. Overall, we compared LOH frequencies at multiple genomic loci as well as between multiple independent strains to assess the extent of heterogeneity in terms of LOH occurrence. The extent of heterogeneity in LOH frequency that we observed between independently constructed strains complicated the comparison of LOH frequency at multiple loci across the genome of *C. albicans* and led us to rethink the way genome instability has been addressed in the past, highlighting the fact that cautions should be taken when drawing conclusions from analysis done with single strains. In addition to heterogeneity in standard laboratory growth conditions, we also observed heterogeneity in LOH frequency in presence of stress, both in terms of the nature of the response (no effect, augmentation or diminution), and in terms of response intensity. Our study suggests that LOH events may occur stochastically within a cell, though genetic background potentially impacts genome stability in terms of LOH throughout the genome in both basal and stress conditions.

Results

We sought out to build a collection of isogenic strains, each one possessing the BFP/GFP LOH reporter system at a distinct genomic locus, as well as a unique barcode sequence, allowing to study LOH events throughout the genome of *C. albicans* and permitting pool experiments. The 14 *C. albicans* strains constituting the Blue-Green (BG) collection were selected amongst the 57 strains studied in Marton *et al.* (Marton *et al.* 2020) as they did not display any obvious transformation-induced gross chromosomal rearrangements (GCR) nor phenotypic aberrations. Our strategy relied on the integration of the BFP/GFP LOH reporter system in the most telomere proximal ≥5 kb intergenic region at both ends of each chromosome (Figure 1), in order to maximize the recovery of long-tract LOH events occurring on each chromosomal arm. The right arm of the asymmetrical Chr6 was not included in our collection as no intergenic region of ≥5 kb could be identified on this short chromosomal arm. Additionally, the left arm of Chr7, which is highly homozygous in SC5314, was also not

included in our analysis as we were unsuccessful in constructing a GCR-free strain carrying the BFP/GFP LOH reporter system on this chromosomal arm.

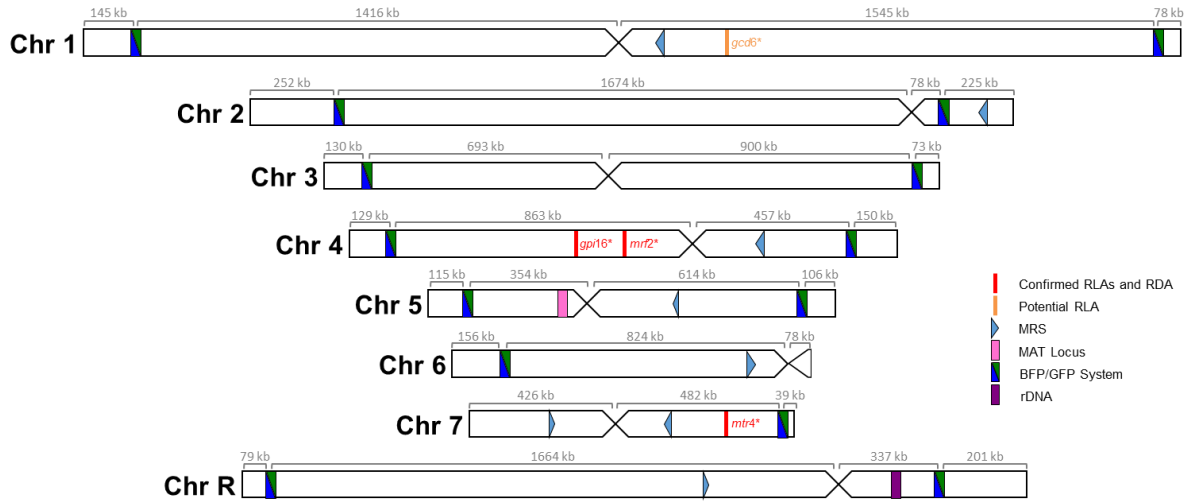

Figure 1: Blue-Green collection.
Positions of the BFP/GFP system insertions utilised to measure LOH frequencies. Distances, in kb, between the telomere and the LOH reporter system and the centromere is indicated above each chromosome. Additionally, illustrated are the locations of known recessive lethal alleles (RLA) and recessive deleterious alleles (RDA), major repeat sequences (MRS), mating-type like (*MTL*) locus and ribosomal DNA (rDNA) locus.

All strains were derived from the *C. albicans* laboratory reference strain SN148 (His^- Arg^- Ura^- Leu^-) (Noble and Johnson 2005). This strain was sequentially transformed in order to integrate (i) the BFP/GFP LOH reporter system, (ii) a unique barcode sequence and (iii) the leucine marker rendering the strains prototrophic. All engineered strains underwent basic phenotyping to assess if the transformation process drastically impacted their phenotype, as described in Marton *et al.* (Marton *et al.* 2020). These strains were shown to form round colonies on both YPD and SD media and to possess comparable doubling times (Figure S1). Additionally, the functionality of the BFP/GFP LOH reporter system was evaluated by challenging the cells with genotoxic treatments such as MMS (DNA alkylating agent) or heat shock (51°C for 90s.) and assessing fold changes in LOH frequencies. We observed, on average, a 9-fold increase in the frequency of mono-fluorescent cells upon exposure of the BG collection to stress (MMS or heat shock) (Table S1). With all strains of the BG collection displaying an augmentation in the total number of mono-fluorescent cells upon genotoxic stress, the BFP/GFP LOH reporter systems in these 14 strains were defined as being functional. Whole-genome sequencing data permitted to conclude that these strains did not acquire any large GCRs during the strain construction process Marton *et al.* (Marton *et al.* 2020).

A robust pipeline to accurately evaluate basal LOH

Fluctuation tests were performed to accurately assess LOH frequency across our collection in various growth conditions. Optimizing our set up, we initially observed high variations in terms of LOH frequency between independent cultures of a given strain in a specific condition. After validating the fact that multiple cytometry analysis from a single culture were very consistent in terms of mono-fluorescent cells frequency, we compared LOH frequency variations between multiple cultures originating from independent colonies of a given strain and between replicate cultures from a single colony preculture (parallel cultures) and saw no differences. Based on these observations, we set up an analysis matrix of LOH frequencies consisting of 24 biological replicates (24 parallel cultures from a single colony preculture), from which 10^6 cells were analyzed eight times by cytometry (technical replicates), screening a total number of 192 x 10^6 cells in order to accurately evaluate a median LOH frequency for a given strain in a specific condition, allowing statistical analyses (Figure 8, material and methods).

A genome-wide evaluation of basal LOH frequency

Using the pipeline described above, basal LOH frequencies were evaluated from 24 parallel cultures per strain in YPD at 30°C, therefore monitoring 14 distinct genomic loci of the *C. albicans* genome (Figure 2A). As shown in Figure 2A, the median basal LOH frequencies of different genomic loci appeared to be significantly different (Kruskal-Wallis Test, $p<0.0001$). The most elevated LOH frequencies were obtained for the right arm of ChrR (possessing the ribosomal DNA (rDNA) locus) and the left arm of Chr2, the chromosomal arm with the longest LOH screening distance (distance between the BFP/GFP LOH reporter system and the centromere). However, the estimated LOH frequencies, even when excluding the two outliers ChrRR and Chr2L, did not appear to be positively correlated with the distance between the BFP/GFP LOH reporter system and the centromere (Peason's R = 0.0581, p-value = 0.45) (Figure 2B). Overall, basal LOH frequencies remain low throughout the genome, with a median frequency of $1.48x10^{-5}$ LOH.

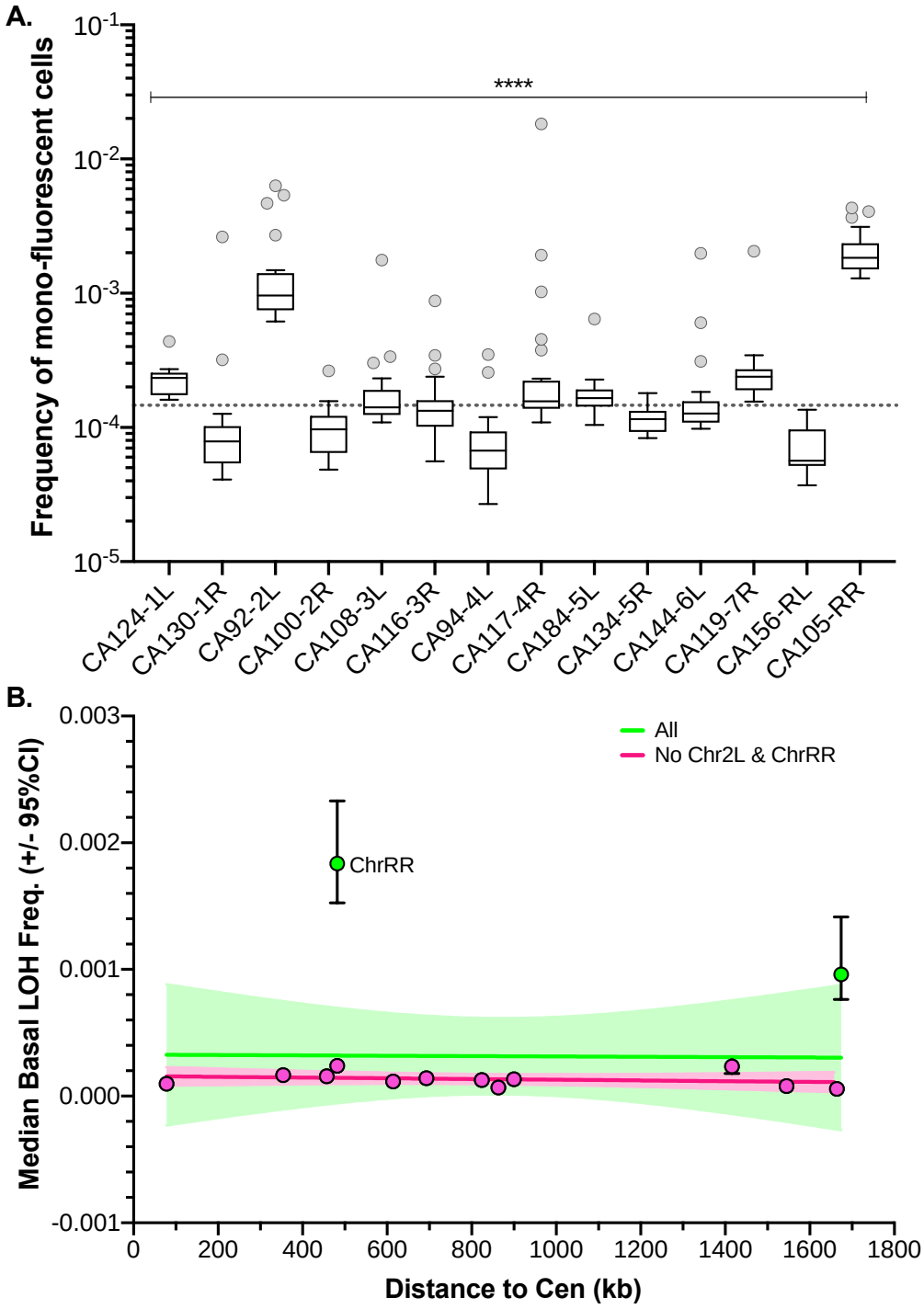

Figure 2: Median basal LOH frequencies across 14 C. albicans genomic loci.
A. Box plot and whiskers (Tukey) of median basal LOH frequencies obtained upon fluctuation analysis using the Blue-Green collection (n=24 parallel cultures). The median basal LOH frequency of the 14 genomic loci is represented by the gray dashed line. Significant differences between medians was assessed using the Kruskal-Wallis test, p<0.0001. **B.** Relationship between median basal LOH frequency (error: 95% CI) and the distance between the BFP/GFP LOH reporter system and the centromere (CEN) (kb). Linear regression of all 14 loci represented in green (Pearson's R= - 0.0149, p=0.962114) and of 13 loci (without ChrRR possession rDNA locus) represented in pink (Pearson's R= 0.3826, p=0.19696).

LOH frequency is influenced by genetic background

For certain genomic loci, we were able to retrieve two independently constructed strains and compare LOH frequencies in those pairs. With regards to basal LOH frequencies, 2/6 pairs (Chr2R and Chr4L) showed non-significant differences in terms of LOH frequency (Mann-Whitney test, p>0.05). However, the other four pairs of strains, Chr1L, Chr3L, Chr3R and Chr5L, exhibited significantly different basal LOH frequencies (Mann-Whitney test, p<0.0001) between independently constructed strains (Figure 3), despite the absence of obvious phenotypic and genomic differences.

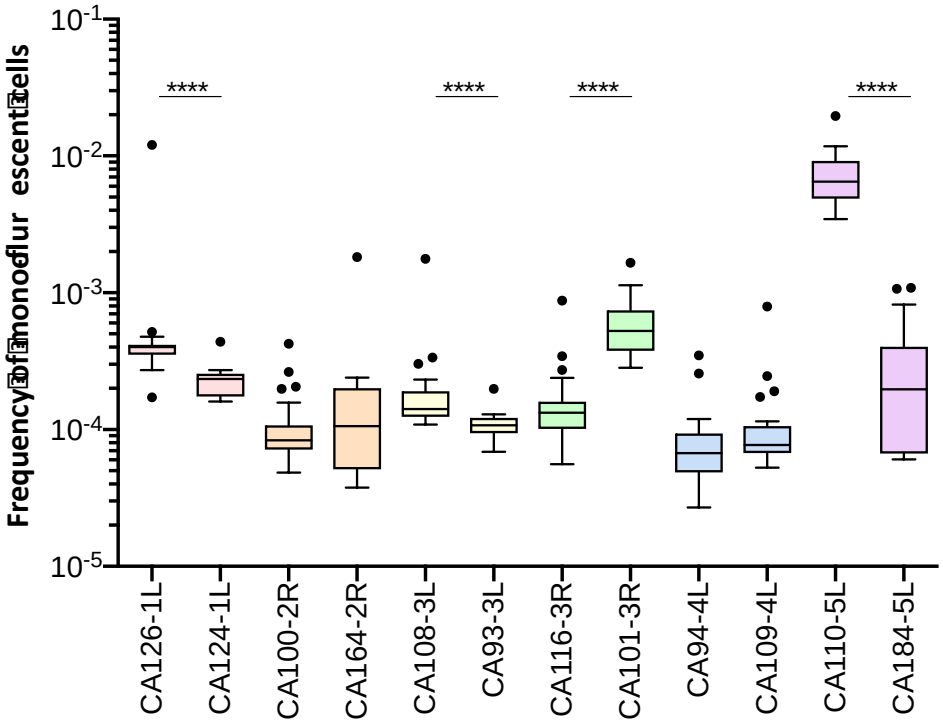

Figure 3: Basal LOH frequencies of two independently constructed strains.
Frequency of mono-fluorescence, LOH frequency, in YPD represented by box plot and whiskers (Tukey) were assessed by fluctuation analysis (n=24 parallel cultures). Frequencies of independently constructed strains were compared using the Mann-Whitney test, p<0.05 (****=p<0.0001).

These observations suggested that smaller genetic differences, potentially acquired in the process of cell transformation, could influence LOH frequency. As we know that the transformation process of *C. albicans* cells is highly mutagenic (Marton *et al.* 2020), we tested our hypothesis by rigorously evaluating heterogeneity in terms of LOH frequency between multiple independently constructed strains. We chose to construct multiple independent SN95-derivatives carrying the BFP/GFP LOH reporter system at three distinct genomic loci, (i) Chr5L (10 strains) because the highest variation was observed for CA110-5L and (ii) Chr2L (9 strains) and (iii) Chr2R (7 strains), a chromosome for which LOH frequency increase was seen for one arm (CA92-2L) but not the other (Ca100-2R). All 26 strains carry a functional BFP/GFP LOH reporter system, as illustrated by an average of a 19-fold increase in mono-fluorescent

cells upon heat shock or MMS treatment, and displayed no obvious phenotypic differences in terms of colony morphology and doubling times (Figure S1, Table S1). Basal LOH frequencies were assessed using the same pipeline as described above, with the exception of conducting 12 parallel cultures rather than 24, thus analyzing a total of 96 x 10^6 cells by cytometry for each strain. As illustrated in Figure 4A-B, the basal LOH frequencies at three different genomic locations range from 2.44x10^{-5} to 1.73x10^{-2} and display a high level of heterogeneity in basal LOH frequency between multiple transformants of a single strain (up to 424-, 69- and 102-fold change for the Chr2L, 2R and 5L loci, respectively). Figure 4C shows the heterogeneity we observed in terms of LOH frequency, represented by the coefficients of variation, a measure of variability between two different datasets. We observed that the heterogeneity between replicates of a single strain is lower than the heterogeneity observed between independent transformants (Figure 4C) and that the heterogeneity we see at the three tested genomic loci is comparable to the heterogeneity observed when comparing the 14 genomic loci within the BG collection. Our observations led us to conclude that, apart from ChrRR for which increased basal LOH frequency might be explained by the rDNA locus, the variability we see in YPD between the other loci is comparable to the intrinsic variability we see between multiple independent transformants; although slight but significant differences in median basal LOH frequencies can be seen between different genomic loci when taking into account multiple strains constructed for a given locus (Figure 4B - Kruskal-Wallis test, p=0.011). This is the case for Chr2R and Chr2L (p=0.001) and Chr2R and Chr5L (p=0.0452), for which significant differences (Mann-Whitney test, n=84-120) are obtained (Figure 4B).

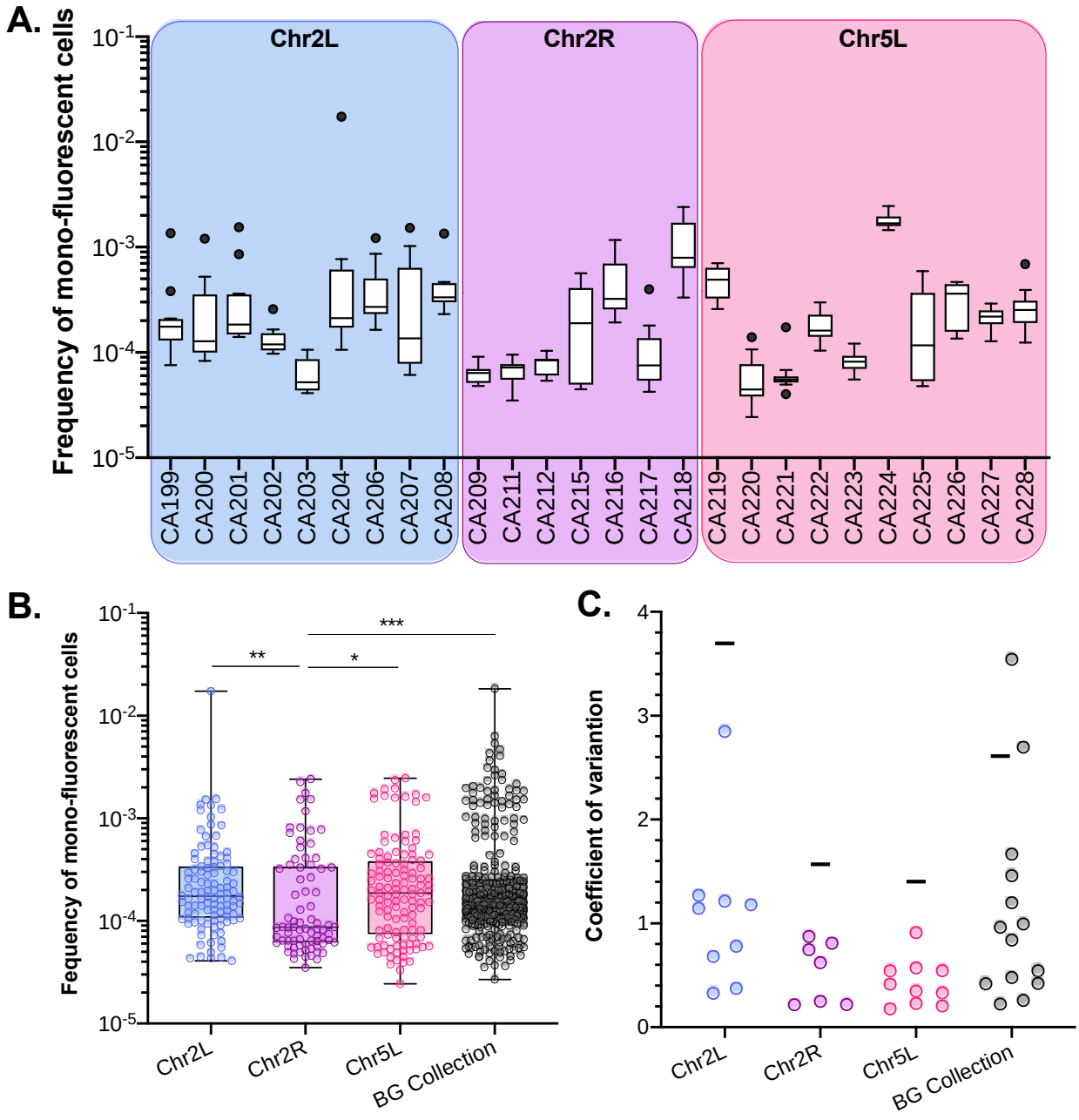

Figure 4: Inter-strain heterogeneity of basal LOH frequency.
A. Median frequencies of mono-fluorescence in YPD, basal LOH frequencies, represented as box plots and whiskers (Tukey) were assessed by fluctuation analysis (n=12 parallel cultures) for genomic loci: Chr2L (9 strains), Chr2R (7 strains) and Chr5L (10 strains). **B.** Pooled frequency of mono-fluorescence for the three loci illustrated in the top panel (Chr2L, Chr2R and Chr5L) and for the 14 strains and loci in the Blue-Green (BG) collection (Figure 1). Overall median frequencies (n=84-336 parallel cultures) were compared using the Mann-Whitney test, significant differences are shown on the graph * p<0.05, **p<0.001 and ***p<0.0001). **C.** Variation of LOH frequencies is represented by the coefficient of variation, where circles represent a given strain while horizontal lines depict the variation between transformants.

Stress induces different responses in terms of LOH frequency

As stress is known to increase LOH frequency in *C. albicans*, we sought out to investigate LOH frequency upon physiologically relevant stresses. Using the same pipeline as

described above, we evaluated LOH frequencies in the 14 *C. albicans* strains of the BG collection upon exposure to oxidative stress (4mM H_2O_2), antifungal agents (10µg/mL fluconazole) and heat (39°C) (Figure 5). Three different outcomes with regards to LOH frequency were observed upon stress exposure: (i) no effect, (ii) augmentation or (iii) reduction in LOH frequency, as compared to the basal LOH frequency assessed in YPD.

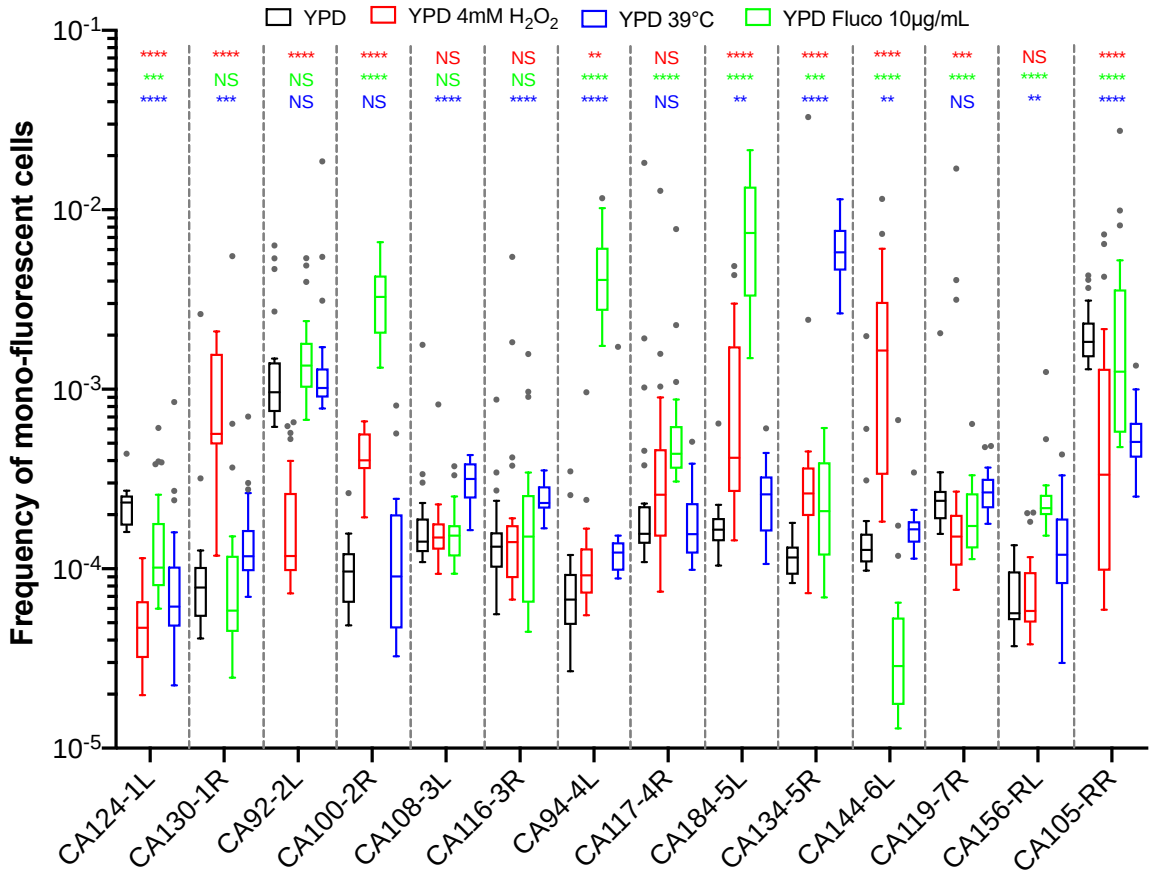

Figure 5: LOH frequency of Blue-Green collection in growth conditions mimicking physiological stresses.
Median frequencies of mono-fluorescence, LOH frequency, in YPD (black), 4mM H_2O_2 (red), 10µg/mL fluconazole (green) and growth at 39°C (blue) represented as box plots and whiskers (Tukey). LOH frequencies were assessed by fluctuation analysis (n=24 parallel cultures) for 14 strains representing fourteen genomic loci (14 strains). Differences of median LOH frequencies, between control condition (YPD) and each stress condition were evaluated using the Mann-Whitney test (NS p>0.05, **p<0.01, *** p<0.001 and ****p<0.0001).

Because of the basal LOH frequency heterogeneity described above, we questioned if the H_2O_2-induced changes in LOH frequencies we observed for certain chromosomes could be interpreted as a true stress-induced chromosome-specific instability or could just illustrate the natural variability we observe between multiple transformants of a same strain. As shown in Figure 5, strains CA92-2L and CA100-2R display opposite responses in terms of LOH frequency in response to H_20_2 treatment, with a 8-fold decrease for CA92-2L and a 4-fold increase in LOH frequency for CA100-2R. To test if these outcomes could be reproduced with

multiple independent strains for the given loci, we used the nine Chr2L and seven Chr2R independently constructed strains, as described above, to assess LOH frequencies upon exposure to hydrogen peroxide (4mM). Here again, the independently constructed strains display different responses to H_2O_2 in terms of genome stability (Figure 6A-B). In contrast to the conclusions drawn from the LOH analysis in single strains within the BG collection (Figure 5), the analysis performed with multiple transformants for a given locus allows us to conclude that only Chr2L exhibits an augmentation of median LOH frequency upon H_2O_2 exposure (Figure 6C), while nonsignificant LOH frequencies were assessed for Chr2R and the BG collection as a whole. Interestingly, while variability between replicates for a specific strain is logically less in YPD than in response to H_2O_2 treatment, the opposite is true concerning variability between multiple independent transformants (Figure 6D). The overall specific impact of the genotoxic stress seems to “mask” the intrinsic variability due to potential genetic variations between multiple independent transformants seen in YPD.

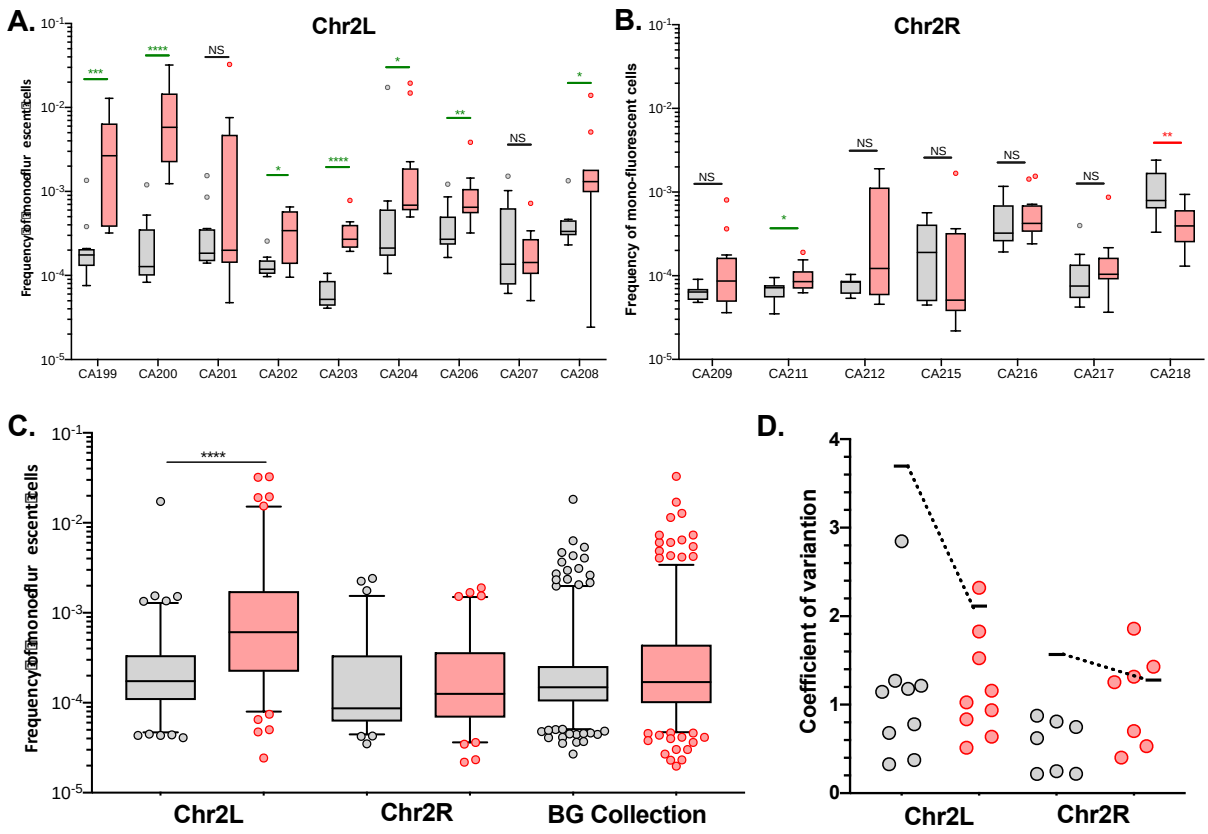

Figure 6: Effect of hydrogen peroxide of the LOH frequency of C. albicans.
Median frequencies of mono-fluorescence, LOH frequency, in YPD (gray) and 4mM H_2O_2 (red) represented as box plots and whiskers (Tukey). LOH frequencies were assessed by fluctuation analysis (n=12 parallel cultures). Differences of median LOH frequencies, between control condition (YPD) and the oxidative stress condition were evaluated using the Mann-Whitney test (NS p>0.05, *p<0.05, **p<0.01, *** p<0.001 and ****p<0.0001). Significant augmentation and reductions of LOH frequencies are identified in green and red, respectively, for panels A. and B. **A.** Representation of median LOH frequencies for 9 strains (n=12 parallel cultures) possessing the BFP/GFP LOH reporter system on Chr2L. **B.** Representation of median LOH frequencies for 7 strains (n=12 parallel cultures) possessing

the BFP/GFP LOH reporter system on Chr2R. **C.** Heterogeneity and median LOH frequency of Chr2L (9 strains), Chr2R (7 strains) and Blue-Green (BG) collection (14 strains) by pooling of multiple strains (n=84-108 parallel cultures). **D.** Heterogeneity of LOH frequencies represented by coefficients of variation. Each circle represents the coefficient of variation for a single strain (n=12) while the variation between strains is shown by a horizontal line (n=84-108).

Stress may not lead to a general response triggering overall genome instability

We then sought out to understand if the LOH increases we observed in certain strains in response to specific stressors when looking at a specific locus could be detected on different chromosomes within the same strain, which would illustrate a general response triggering overall genome instability or if, conversely, chromosomes do react independently to a specific stress. As shown in Figure.5, CA184-5L shows an increase in LOH frequency upon fluconazole treatment, while CA124-1L does not. We took advantage of the *GAL1* locus on the left arm of Chr1 to build a strain possessing two different LOH reporter systems: the BFP/GFP LOH reporter system localized on the left arm of Chr5 and the *GAL1/URA3* system on the left arm of Chr1 (Figure 7A). In YPD, we obtained median basal LOH frequencies of 4.95×10^{-5} and 2.07×10^{-4} on Chr1L and Chr5L, respectively (Figure 7B). Upon exposure to fluconazole (10 µg/mL), no changes in LOH frequency on Chr1L were detected (Mann-Whitney test, $p < 0.142$), while LOH frequency on Chr5L shows a statistically significant 52.9-fold increase (T-test, $p < 1.185 \times 10^{-9}$) (Figure 7B), demonstrating that within the same strain, chromosomes will react differently to the same genotoxic stress, with certain loci being more prone to undergo LOH than others. These differences are also observed when normalizing LOH frequencies to the distance between the LOH reporter systems and the centromeres.

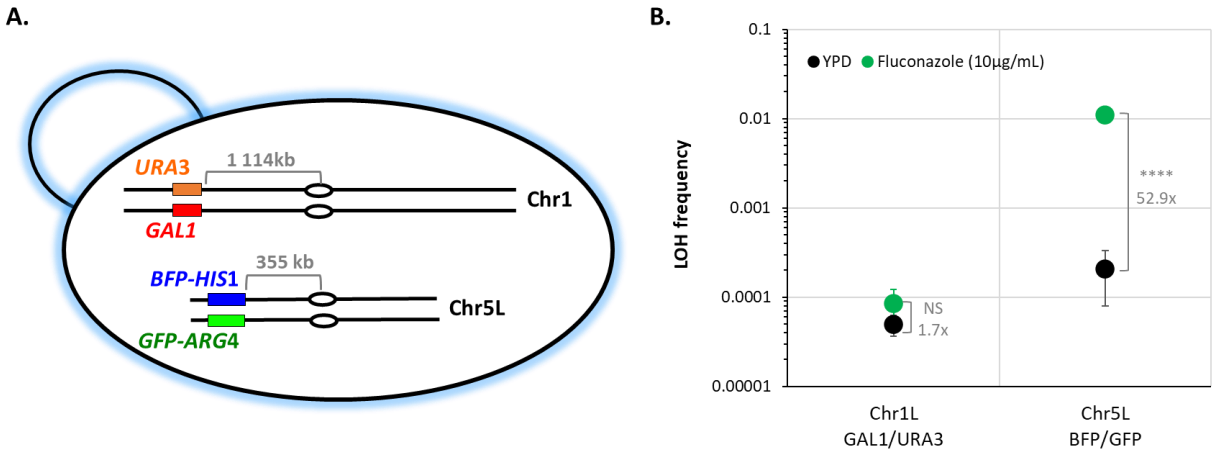

Figure 7: Assessing LOH frequency of two distinct C. albicans chromosomes within a single strain.
A. Illustration of the strain possessing the BFP/GFP LOH reporter system on Chr5L and the GAL1/URA3 system of Chr1L. Centromeres are indicated by ovals and distances between reporter systems and centromeres are indicated in gray. **B.** Median frequencies of mono-fluorescence, LOH frequency, in YPD (black) and 10µg/mL fluconazole (green), +/- standard deviation. LOH frequencies were assessed by fluctuation analysis (n=6, from two independent strains). Fold changes of median

LOH frequencies, between YPD and fluconazole conditions are indicated on the plot in gray and statistical differences were evaluated using T-test (NS p>0.05 and ****p<0.0001).

Discussion

Eukaryotic genomes are naturally dynamic (Parfrey et al. 2008), under the constant watch of specific surveillance mechanisms required to reach the fine-tuned balance between genomic stability, which ensures transmission of genetic content to the next generation, and genomic plasticity, which promotes acquisition of new genetic variants and defines the adaptation potential of microorganisms to a fluctuating environment. A particular form of chromosome instability, namely LOH, is sufficient to provide selective growth advantage and is often seen in tumors. LOH occurs when a cell that is originally heterozygous at a locus loses one of its two alleles at that locus. This phenomenon is particularly relevant in heterozygous diploid organisms where LOH allows recessive phenotypes to be expressed.

We took advantage of the human fungal pathogen *C. albicans* to investigate the dynamics of genome-wide LOH in a diploid heterozygous eukaryotic genome. By monitoring LOH frequency on each arm of the 8 chromosome pairs, our work revealed high variability in terms of genome stability not only at the locus level but also at the strain level, highlighting inter-strain, intra-strain and inter-chromosome variations, suggesting that multiple stochastic parameters influence genome dynamics in *C. albicans*.

Our initial analysis, looking at a single strain per locus, showed that the median basal LOH frequencies of different genomic loci appeared to be significantly different in absence of stress (Figure 2A). Two outliers were observed, namely ChrRR (possessing the ribosomal DNA (rDNA) locus) and Chr2L, the chromosomal arm with the longest LOH screening distance (distance between the BFP/GFP LOH reporter system and the centromere) (Figure 2). Instability of the rDNA locus is a known feature of the *C. albicans* genome (Rustchenko *et al.* 1993). This locus encompasses the 18S, 5.8S, 25S, and 5S rRNAs organized as tandem repeating units, whose number can vary from 21 to 176 copies depending on the strains and growth conditions. The variation in copy number, resulting from frequent unequal intrachromosomal recombination, translates into large-scale shifts in the rDNA locus size, ranging from 244 kb to 2200 kb. The recombinogenic properties of the rDNA locus are also illustrated by the fact that the majority of *C. albicans* clinical isolates are homozygous downstream of the rDNA locus and up to the telomere. Regarding Chr2L, it is largely accepted that the probability of a recombination event between a locus and its centromere increases with distance from the centromere. While Forche *et al*. demonstrated that LOH rates at different loci generally correlated with increased distance from the centromere in *C. albicans*

(Forche *et al.* 2011), such correlation was not reached in our hands (Figure 2B). This discrepancy can be explained by the high variability revealed by our work, in terms of LOH frequency between independent strains carrying the LOH reporter system at the same locus. Additionally, the localizations of our LOH reporter systems were designed to recover a maximum amount of distal LOH events whereas the study of Forche *et al.* used randomly distributed loci. Thus, the latter inconsistency between the two studies could also be explained by the absence of correlation between distal LOH frequency and chromosome arm length (Marton *et al., unpublished,* Chapter 1). Indeed, our analysis, extended to several independent transformants for the same locus, led us to conclude that, apart from ChrRR for which increased basal LOH frequency might be explained by the rDNA locus, the variability we see in YPD between the other loci is comparable to the intrinsic variability we see between multiple independent transformants (Figure 4B); although slight but significant differences in median basal LOH frequencies can be seen between different genomic loci when taking into account multiple strains constructed for a given locus. This is the case for Chr2R for which we observed a significant lower basal LOH frequency as compared to Chr2L and Chr5L (Figure 4B).

Stress-inducible mutagenesis mechanisms have been described in bacteria, yeast, and human cancer cells. In particular in hybrid yeast strains, positive selection on existing heterozygosity for adaptation to new environment through LOH has been documented (Smukowski Heil *et al.* 2017). Relying on LOH to generate phenotypic diversity is all the more relevant in a heterozygous diploid microorganism whose mode of reproduction is mainly clonal as it allows *C. albicans* to avoid Haldane's sieve, with recessive beneficial mutations that can be revealed and can contribute to rapid evolutionary adaptation through LOH (Gerstein *et al.* 2014). Therefore, we were interested in having a genome-wide picture of LOH dynamics in response to relevant stresses. Although we were able to conclude that Chr2L is more prone to undergo LOH as compared to Chr2R in response to oxidative treatment (Figure 6C), our data especially highlighted the great natural variability that can be observed again between multiple transformants of a same strain (Figure 6D). It was striking to see independent transformants displaying opposite response in terms of LOH frequency upon exposure to H_2O_2 with some showing a decrease while others show an increase in LOH frequency at a single genomic locus (Figure 6A-B). Remarkably, our results demonstrate that the stress-induced genome instability we observed for a specific chromosome is not the consequence of a general response that triggers overall genome instability. Indeed, we showed that within the same strain, Chr1L and Chr5L will react differently to fluconazole treatment, with Chr5L being more prone to undergo LOH than Chr1L (Figure 7). Altogether, our results revealed a high heterogeneity between multiple transformants in terms of genome stability and highlighted the fact that conclusions drawn from genome analysis performed on a single strain must be

tempered by appropriate cautions. Knowing now that data from two independent transformants can go in opposite directions, only analysis from multiple transformants should be performed to confidently conclude on a trend in terms of LOH frequency at a specific locus.

This elevated variability that seems to be omnipresent in *C. albicans* can be seen at various levels. Of course, by the very definition of a fluctuation analysis, variations in terms of LOH frequency between parallel cultures from a unique pre-culture were anticipated based on the timing of appearance of the new mutation in a single cell during growth of the population. Also, inter-chromosome variability in terms of genome stability in response to stress was already suggested by Forche *et al.* (Forche *et al.* 2011) when they showed that fold changes for LOH rates in response to heat, oxidative stress and antifungals, were highly variable between strains with a LOH reporter system at five locations in the genome. Inter-strain variability in additional aspects of *C. albicans* biology were also suggested in other studies such as the work by Huang *et al*. (Huang *et al.* 2019) on differences in regulation of biofilm formation in five clinical isolates. Altogether, other genomic features or other parameters seem to influence LOH occurrence in *C. albicans*. These parameters are not yet fully comprehended but several scenarios can be considered. (i) We know that the process of transformation in *C. albicans* is highly mutagenic with 33% to 60% of transformed strains displaying gross chromosomal rearrangements, depending on the protocol(Marton *et al.* 2020). It would be interesting to perform deep whole-genome sequencing of supposedly isogenic independent transformants to see if the phenotypic variations observed in terms of LOH frequency is a fair representation of transformation-induced genomic differences. (ii) Cells could also differ in terms of level of gene expression as it has been reported in *S. cerevisiae*, in particular expression noise of genes affecting HR activity and responsible for cell-to-cell heterogeneity in terms of HR rate (Liu *et al.* 2019), which could thus impact LOH frequency. (iii) Variations in epigenetic states between strains could also explain LOH frequency heterogeneity as it has been shown that chromatin structure influences the occurrence of recombination events (reviewed in (Buscaino 2019)). Therefore, comparison of methylation profiles between strains could bring light on the causes of LOH frequency heterogeneity. (iv) It has been shown in cancer genomes that LOH preferentially occurs in early replicating regions, where interference between replication and transcription machineries can lead to the formation of double-strand breaks (Pedersen and De 2013)**.** Based on the observations made by Koren *et al.* (Koren *et al.* 2014) that DNA replication timing is shaped by genetic polymorphisms in human cells, it would be interesting to investigate whether variations in terms of DNA replication origins landscape between different strains could explain the variations in LOH frequency. (v) Studies on the relative frequency of LOH events between human chromosomes revealed that the relative frequency of LOH events per chromosome had significantly inverse correlation with

the distance between homologous chromosomes in the nucleus (Pedersen and De 2013). Eukaryotic chromosomes occupy distinct nuclear territories, such that some pairs of homologous chromosomes are closer to each other than other pairs. Little has been done regarding chromosome organization in the *C. albicans* nucleus. Variations from cell to cell could impact the evaluation of LOH frequency in a given *C. albicans* population.

Using molecular approaches to understand *C. albicans* biology, we tend to see phenotypic variability as a barrier that prevents us from carrying out conclusive statistical analysis when we should in fact embrace it as a fair representation of the potential of adaptation of this human fungal pathogen and a functional consequence of genetic variations. Our work calls into question certain statements that have been made on the action of certain genes in the genome dynamics of *C. albicans*.

Material and methods

Strains and culturing conditions

C. albicans strains described in the study are derived from the reference strains SC5314, parental strain SN148 (His^- Arg^- Ura^- Leu^-) or SN95 (His^- Arg^-) (Noble and Johnson 2005). Yeast cells were cultured on/in rich YPD medium (1% yeast extract, 2% peptone, 2% dextrose). Synthetic Defined (SD) (0.67% yeast nitrogen base without amino acids, 2% dextrose) and Synthetic Complete (SC) (0.67% yeast nitrogen base without amino acids, 2% dextrose, 0.08% drop-out mix with all the essential amino-acids) media were used for selection. Solid media were obtained by adding 2% agar.
All *C. albicans* strains are listed in Table S2.

Blue-Green Collection strain construction

The Bleu-Green (BG) Collection is composed of a series of barcoded strains, each one possessing the BFP/GFP LOH reporter system on a given chromosome, designed to assess LOH frequencies throughout the *C. albicans* genome. These strains were constructed as described in Marton *et al*.(Marton *et al.* 2020). These strains (Table S2) were selected as they do not display any obvious GCR.

Validation of strains

- *Basic phenotyping*

All strains underwent basic phenotypic characterization, additionally to the verification of the proper integration of cassettes at targeted loci by junction PCR. Functionality of auxotrophic markers was evaluated by drop tests on SC medium with appropriate drop-out amino acid,

depending on tested marker. Overnight saturated cultures in YPD of selected strains were spotted on YPD, SC-His, SC-Arg, SC-Ura and SC-Leu and placed at 30°C for 24h to observe presence and absence of growth. Furthermore, the functionality/intensity of both fluorescence proteins (BFP and GFP) was validated by flowcytometry and fluorescence microscopy. The colony morphology of all strains was also assessed on both solid YPD and SD media, at 30°C. While doubling times were evaluated in YPD media at 30°C by measuring the optical density with TECAN Infinite.

- *Assessing functionality of the LOH reporter system*

Additionally, functionality of the BFP/GFP LOH reporter system was also evaluated by exposing strains to genotoxic stress, Methyl methanesulfonate (MMS) and heat shock, and measuring the increase of mono-fluorescent populations. Homogenous BFP/GFP heterozygous precultures were obtained in SC-His-Arg medium overnight at 30°C. Sixty million cells of the preculture were inoculated in 3 mL of YPD and incubated for 3 h at 30°C, in order to obtain cells in exponential growth phase. The latter culture was divided into two (2x 1.5 mL) and cells were pelleted and resuspended in either YPD at room temperature (control) or, YPD + MMS or pre-heated YPD, stress condition. The MMS stress assay is conducted in volumes of 3 mL YPD (control) or YPD + 0.03% MMS (stress) and exposed for 30 mins. at 30°C. Then, cells are pelleted, washed with fresh YPD and are recovered in 3mL of YPD incubated at 30°C overnight. The heat shock assay is performed in 200 µL of YPD (control) or pre-heated YPD at 50°C (stress) and stress exposed cells are confronted to 51°C for 90 s.. Both heat shocked and control samples are placed on ice for 5 mins. before inoculation into a fresh 3 mL of YPD and left to recover at 30°C overnight. Following the recovery time, 10^6 cells of each sample is analysed by flowcytometry, MACSQuant analyzer (Miltenyi Biotec), where the BFP is detected with a 405-nm laser and 425- to 475-nm filters and the GFP is detected with a 488-nm laser and 500- to 550-nm filters. To determine the number of LOH affected cells, FACS profiles are analysed using FlowJo V10.1 software.

- *Identifying the delimitations of the mono-fluorescent populations on FACS profiles in different genomic loci*

The appearance of LOH events was triggered by heat shock in various strains of the Blue-Green Collection to facilitate the identification, characterization, and delimitation of the flowcytometry sorted populations, particularly the mono-fluorescent populations. Heat shocked cultures were filtered using BD Falcon™ Cell strainers in order to remove large debris. Cells were diluted in 1x PBS at a final concentration of at least $20x10^6$ cells/mL and kept on ice. The MoFlo® Astrios™ flow cytometer, located at the Cytometry platform of the

Institut Pasteur, Paris, was used to analyze and sort the cells of interest. For each sorted gate, 400 cells were recovered in 400 µL of liquid YPD medium. Sorted cells were plated on three YPD agar plates and incubated at 30°C for 48h. Recovered CFUs were counted and a subset of CFUs (n=16) from each population was characterized by spot assay and by FACS. Spot assays were conducted on YPD (control), SC-His (test presence of the *BFP-HIS*1 cassette) and SC-Arg (test presence of the G*FP-ARG*4 cassette) while fluorescence status was assessed by FACS (10^4 events per sample). Overall, all major populations identifiable in the FACS profiles were characterized, permitting the identification and the delimitation of mono-fluorescent populations (mono-BFP, mono-GFP) and double fluorescent population (BFP/GFP).

Fluctuation assay to determine LOH frequency

LOH frequencies at different genomic loci and in different growth conditions were evaluated using the following fluctuation assay. A preculture was established by inoculating a single colony in liquid SC-His-Arg medium and incubated overnight at 30 °C, assuring the starting population is homogenous at the BFP/GFP locus. One hundred cells of the preculture were then inoculated into multiple parallel cultures of 1 mL (YPD or YPD + stressor) in Deep-well 96 plates and incubated at 30°C (or 39°C) until stationary phase was reached (48h growth and 72h growth for fluconazole condition). For each parallel culture, 8 x10^6 cells were analyzed by FACS, 8 FACS analysis each assessing fluorescence of 10^6 yeast cells, using the MACSQuant analyzer (Miltenyi Biotec), where the BFP is detected with a 405-nm laser and 425- to 475-nm filters and the GFP is detected with a 488-nm laser and 500- to 550-nm filters. A total of 192 (24 parallel cultures) or 96 (12 parallel cultures) FACS output files per strain and condition were analyzed using the FlowJo V10.1 software. The gates to determine the LOH frequencies were selected according to the delimitations of the mono-fluorescent populations (as identified in the above section) conserved throughout sample analysis. Total number of mono-BFP and mono-GFP cells, from the 8 FACS profiles obtained per parallel culture, are added to determine a total number of cells which have undergone LOH. The absolute mono-fluorescence frequency for a given strain and condition represents the median (+/-SD) LOH frequency of 24 or 12 parallel cultures. Figure 8 is a schematic representation of analysis matrix permitting to determine an LOH frequency for a strain in a condition.

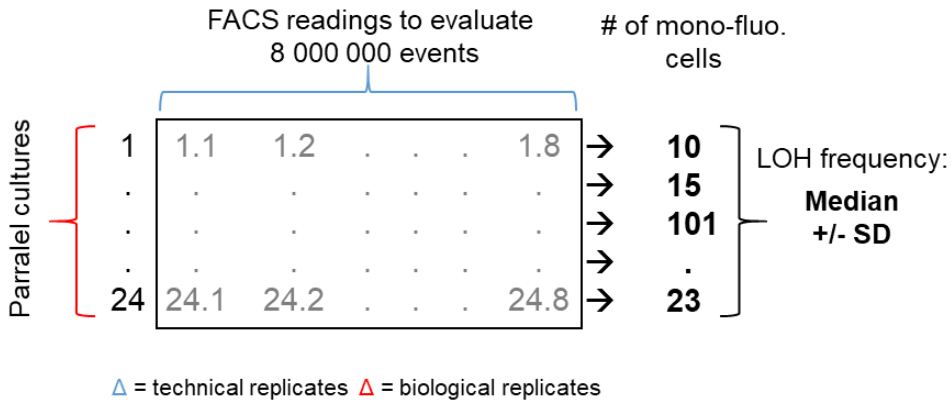

Figure 8: Fluctuation analysis allowing to determine an accurate LOH frequency.
For a given strain in one growth condition, 24 parallel cultures are generated. For each parallel culture, 8 FACS analyses of 10^6 cells is conducted for a total sampling size of $8x10^6$ cells. A total number of mono-fluorescent cells is calculated per parallel culture. The median LOH frequency (+/- SD) is assessed using the 24 parallel cultures.

Strain construction for assessing LOH frequency heterogeneity

Using the transient CRISPR-cas9 transformation protocol (Min *et al.* 2016), both homologues were simultaneous targeted for transformation cassette integration. Thus, by directing a DNA DSB with a sgRNA the BFP/GFP LOH reporter system can be engineered in only one transformation round. The BFP/GFP LOH reporter system was integrated in the most telomere proximal intergeneic region of ≥ 5 kb on Chr2L, Chr2R or Chr5L, at the same loci selected in Marton *et al.* (Marton *et al.* 2020) using the same sgRNA and repair templates. The sgRNA and cas9 cassettes utilized were conducted as described in Min *et al.* (Min *et al.* 2016), from the pV1093 plasmid. While repair template cassettes were constructed using 120 bp primers, composed of 20 bp complementary to both the P_{TDH3}-*GFP-ARG4* and P_{TDH3}-*BFP-HIS1* cassettes and 100 bp tails possessing the complementary sequences of the targeted integration locus (Table S2). Each primer pair was utilized to amplify both the P_{TDH3}-*GFP-ARG4* and P_{TDH3}-*BFP-HIS1* cassettes from plasmid pCRBluntII-P_{TDH3}-*GFP-ARG4* and plasmid pCRBluntII-P_{TDH3}-*BFP*-Cd*HIS1*, respectively. Each cassette was amplified in a total PCR volume of 500 µL, precipitated in 100% ethanol and re-suspended in 100 µL of distilled sterile water. The SN95 parental strain, possessing both arginine and histidine auxotrophies, was co-transformed with 3 µg of P_{TDH3}-*GFP-ARG4* cassette, 3 µg of P_{TDH3}-*BFP-HIS1* cassettes, 1 µg of Cas9 cassette and 1 µg of sgRNA using the Lithium Acetate/PEG transformation protocol. The transformants were selected on SD medium and junction PCRs were performed in order to ensure proper integration of both cassettes at the targeted locus. Additionally, these strains underwent basic phenotyping and the functionality of the LOH reporter system was validated as described in above sections.

Stresses conditions

Stress conditions and concentrations were selected based on literature and laboratory routine use to induce genomic instability; oxidative stress (H_2O_2) 4mM, fluconazole 10µg/mL and heat (39°C). The fitness of each strain was evaluated by each strain by doubling time, YPD media with stressor at 30°C by measuring the optical density with TECAN Infinite.

Two reporter system strain

A strain possessing two different LOH reporter systems, BFP/GFP (Chr5L) and *GAL1/URA3* (Chr1L) systems, was constructed in order to simultaneous evaluate LOH frequencies in two district genomic loci. The CA183 strain (Table S3) (Ura^- Leu^-) possessing the BFP/GFP LOH reporter system on the left arm of Chr5 was transformed using the *URA3* knock-out cassette targeting the *GAL1* locus on the left arm of Chr1. Repair template was constructed by PCR amplification from plasmid pFA-*URA3*, as described for the integration of the BFP/GFP LOH reporter, using primer in Table S3 and classical Lithium Acetate/PEG transformation protocol (Walther and Wendland 2003). The transformants were selected on SD+leu medium, junction PCRs were performed in order to ensure proper integration of both cassettes at the targeted locus and strains underwent basic phenotyping (as described in above sections).

The resulting strain is heterozygous at the *GAL1* locus, *GAL1/URA3*, and heterozygous on the Chr5L, *BFP-HIS1*/*GFP-ARG4*, thus, cells which undergo an LOH event loose either the *GAL1* or *URA3* alleles or *BFP-HIS1* or *GFP-ARG4*, respectively. Parallel cultures (2x three 1 mL cultures) were generated to assess LOH frequencies of Chr5L and Chr1L. LOH frequency of Chr5L was evaluated using the same protocol as the FACS fluctuation assay previously described, total of 48 FACS profiles. While, the same parallel cultures were also used for counter selection of CFUs having lost of either *GAL1* or *URA3* on media containing 2-deoxygalactose (2-DG) or 5-fluorrotic acid (5-FOA), respectively. The 5-FOA resistant (5-FOA^R) CFUs are indicative of individuals which have undergone an LOH and have lost the *URA3* marker while 2-DG resistant (2-DG^R) CFUs are indicative of individuals which have lost the *GAL1* gene. LOH frequency at this locus on Chr1L is calculated by the tabulation of 5-FOA^R and 2-DG^R CFUs.

Acknowledgments

We are grateful to Gilles Fischer, Nicolas Agier and Stéphane Delmas for their advice and help with the setup of the fluctuation assay. We also thank to Pierre-Henri Commere for help with the cell sorting experiments.

T.M., M.L., and C.D. designed experiments. T.M. performed experiments. T.M., M.L. and C.D. analyzed data. T.M., M.L., and C.D. wrote the manuscript.

T.M. is the recipient of a PhD. fellowship from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID). We acknowledge support from the French Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases [ANR10-LABX-62-IBEID]).

Supplementary material

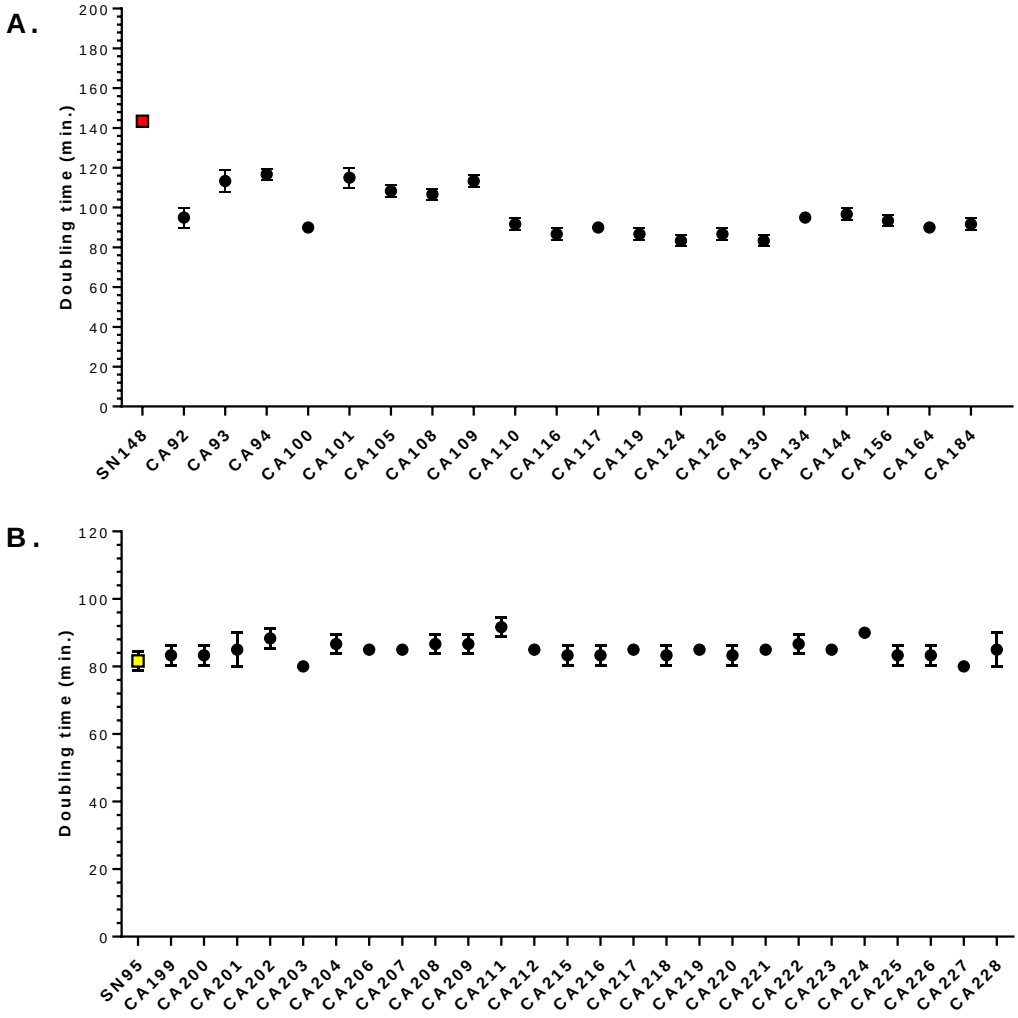

Figure S1: Doubling time of C. albicans strains used within this study.
Average doubling time (mins.) evaluated from optical density 24h kinetic at 30°C (n= 3), error bars represent standard deviation. **A.** Strains within the Blue-Green (BG) Collection and replicates. Parental strains SN148 is indicated by a red square. **B.** Strains used for heterogeneity experiments. Parental strains SN95 is indicated by a yellow square.

Table S1: Fold changes of monofluorescent cells upon stress exposure permitting to validate the functionality of the BFP/GFP LOH reporter systems within C. albicans strains used in this study. *

Strain	Loc. BFP/GFP Sys. (Chr)	Condition	Percentage of Population		Frequency			Fold Changes
			B+	G+	B+	G+	Total	
CA124	1L	YPD	0.0061	0.0293	6.10E-05	2.93E-04	3.54E-04	4.30
		HS	0.0092	0.1430	9.20E-05	1.43E-03	1.52E-03	
CA126	1L	YPD	0.0451	0.0042	4.51E-04	4.20E-05	4.93E-04	5.78
		HS	0.0516	0.2335	5.16E-04	2.34E-03	2.85E-03	
CA130	1R	YPD	0.0560	0.0070	5.60E-04	7.00E-05	6.30E-04	14.71
		HS	0.6376	0.2891	6.38E-03	2.89E-03	9.27E-03	
CA92	2L	YPD	0.0480	0.0640	4.80E-04	6.40E-04	1.12E-03	3.21
		HS	0.1400	0.2200	1.40E-03	2.20E-03	3.60E-03	
CA100	2R	YPD	0.0270	0.0290	2.70E-04	2.90E-04	5.60E-04	2.21
		HS	0.0860	0.0380	8.60E-04	3.80E-04	1.24E-03	
CA164	2R	YPD	0.0398	0.0239	3.98E-04	2.39E-04	6.37E-04	15.20
		HS	0.5908	0.3773	5.91E-03	3.77E-03	9.68E-03	
CA93	3L	YPD	0.1900	0.0740	1.90E-03	7.40E-04	2.64E-03	2.84
		HS	0.6400	0.1100	6.40E-03	1.10E-03	7.50E-03	
CA108	3L	YPD	0.0390	0.0015	3.90E-04	1.50E-05	4.05E-04	11.36
		HS	0.4200	0.0400	4.20E-03	4.00E-04	4.60E-03	
CA101	3R	YPD	0.0110	0.0540	1.10E-04	5.40E-04	6.50E-04	2.08
		HS	0.0350	0.1000	3.50E-04	1.00E-03	1.35E-03	
CA116	3R	YPD	0.0140	0.0025	1.40E-04	2.50E-05	1.65E-04	9.58
		HS	0.1400	0.0180	1.40E-03	1.80E-04	1.58E-03	
CA94	4L	YPD	0.0008	0.0002	8.00E-06	1.70E-06	9.70E-06	8.97
		HS	0.0053	0.0034	5.30E-05	3.40E-05	8.70E-05	
CA109	4L	YPD	0.0005	0.0002	4.90E-06	1.60E-06	6.50E-06	13.38
		HS	0.0064	0.0023	6.40E-05	2.30E-05	8.70E-05	
CA117	4R	YPD	0.0420	0.0160	4.20E-04	1.60E-04	5.80E-04	3.10
		HS	0.0600	0.1200	6.00E-04	1.20E-03	1.80E-03	
CA110	5L	YPD	0.0340	0.0004	3.40E-04	4.00E-06	3.44E-04	7.44
		HS	0.2400	0.0160	2.40E-03	1.60E-04	2.56E-03	
CA184	5L	YPD	0.1261	0.0017	1.26E-03	1.70E-05	1.28E-03	5.81
		HS	0.5992	0.1432	5.99E-03	1.43E-03	7.42E-03	
CA134	5R	YPD	0.0479	0.0185	4.79E-04	1.85E-04	6.64E-04	21.71
		HS	0.8467	0.5950	8.47E-03	5.95E-03	1.44E-02	
CA144	6L	YPD	0.0431	0.0116	4.31E-04	1.16E-04	5.47E-04	19.67
		HS	0.8178	0.2581	8.18E-03	2.58E-03	1.08E-02	
CA119	7R	YPD	0.0180	0.0006	1.80E-04	6.00E-06	1.86E-04	7.74
		HS	0.1300	0.0140	1.30E-03	1.40E-04	1.44E-03	
CA156	RL	YPD	0.0171	0.0036	1.71E-04	3.60E-05	2.07E-04	7.76
		HS	0.0425	0.1181	4.25E-04	1.18E-03	1.61E-03	
CA105	RR	YPD	0.0590	0.0070	5.90E-04	7.00E-05	6.60E-04	7.12
		HS	0.4100	0.0600	4.10E-03	6.00E-04	4.70E-03	
CA199	2L	YPD	0.0074	0.0002	7.40E-05	2.00E-06	7.60E-05	43.80
		HS	0.3298	0.0031	3.30E-03	3.10E-05	3.33E-03	
CA200	2L	YPD	0.0101	0.0009	1.01E-04	9.00E-06	1.10E-04	14.47
		MMS	0.1466	0.0126	1.47E-03	1.26E-04	1.59E-03	
CA201	2L	YPD	0.0424	0.0031	4.24E-04	3.10E-05	4.55E-04	3.24
		HS	0.0686	0.0788	6.86E-04	7.88E-04	1.47E-03	
CA202	2L	YPD	0.0025	0.0038	2.50E-05	3.80E-05	6.30E-05	13.90
		HS	0.0116	0.0760	1.16E-04	7.60E-04	8.76E-04	
CA203	2L	YPD	0.0189	0.0003	1.89E-04	3.00E-06	1.92E-04	5.14
		MMS	0.0907	0.0079	9.07E-04	7.90E-05	9.86E-04	
CA204	2L	YPD	0.0105	0.2076	1.05E-04	2.08E-03	2.18E-03	3.14
		HS	0.0066	0.6786	6.60E-05	6.79E-03	6.85E-03	
CA206	2L	YPD	0.0150	0.0005	1.50E-04	5.00E-06	1.55E-04	16.50
		MMS	0.2545	0.0012	2.55E-03	1.20E-05	2.56E-03	
CA207	2L	YPD	0.0078	0.0028	7.80E-05	2.80E-05	1.06E-04	5.68
		MMS	0.0534	0.0068	5.34E-04	6.80E-05	6.02E-04	
CA208	2L	YPD	0.0060	0.0092	6.00E-05	9.20E-05	1.52E-04	22.05
		HS	0.3173	0.0178	3.17E-03	1.78E-04	3.35E-03	

CA209	**2R**	**YPD**	0.0048	0.0071	4.80E-05	7.10E-05	1.19E-04	3.17
		MMS	0.0102	0.0275	1.02E-04	2.75E-04	3.77E-04	
CA211	**2R**	**YPD**	0.0287	0.0005	2.87E-04	5.00E-06	2.92E-04	2.17
		HS	0.0562	0.0072	5.62E-04	7.20E-05	6.34E-04	
CA212	**2R**	**YPD**	0.0145	0.0001	1.45E-04	1.00E-06	1.46E-04	3.01
		HS	0.0419	0.0020	4.19E-04	2.00E-05	4.39E-04	
CA215	**2R**	**YPD**	0.0117	0.0245	1.17E-04	2.45E-04	3.62E-04	8.88
		HS	0.1603	0.1611	1.60E-03	1.61E-03	3.21E-03	
CA216	**2R**	**YPD**	0.0221	0.0007	2.21E-04	7.00E-06	2.28E-04	19.67
		MMS	0.4436	0.0049	4.44E-03	4.90E-05	4.49E-03	
CA217	**2R**	**YPD**	0.0035	0.0005	3.50E-05	5.00E-06	4.00E-05	103.05
		HS	0.2967	0.1155	2.97E-03	1.16E-03	4.12E-03	
CA218	**2R**	**YPD**	0.0050	0.0035	5.00E-05	3.50E-05	8.50E-05	9.00
		MMS	0.0608	0.0157	6.08E-04	1.57E-04	7.65E-04	
CA219	**5L**	**YPD**	0.0056	0.0004	5.60E-05	4.00E-06	6.00E-05	27.10
		MMS	0.1620	0.0006	1.62E-03	6.00E-06	1.63E-03	
CA220	**5L**	**YPD**	0.0011	0.0021	1.10E-05	2.10E-05	3.20E-05	124.69
		HS	0.2129	0.1861	2.13E-03	1.86E-03	3.99E-03	
CA221	**5L**	**YPD**	0.0049	0.0079	4.90E-05	7.90E-05	1.28E-04	3.94
		MMS	0.0323	0.0181	3.23E-04	1.81E-04	5.04E-04	
CA222	**5L**	**YPD**	0.0087	0.0023	8.70E-05	2.30E-05	1.10E-04	29.23
		HS	0.0906	0.2309	9.06E-04	2.31E-03	3.22E-03	
CA223	**5L**	**YPD**	0.0050	0.0006	5.00E-05	6.00E-06	5.60E-05	7.38
		MMS	0.0371	0.0042	3.71E-04	4.20E-05	4.13E-04	
CA224	**5L**	**YPD**	0.0083	0.0029	8.30E-05	2.90E-05	1.12E-04	3.71
		HS	0.0408	0.0008	4.08E-04	8.00E-06	4.16E-04	
CA225	**5L**	**YPD**	0.0078	0.0186	7.80E-05	1.86E-04	2.64E-04	5.49
		HS	0.1383	0.0067	1.38E-03	6.70E-05	1.45E-03	
CA226	**5L**	**YPD**	0.0055	0.0111	5.50E-05	1.11E-04	1.66E-04	8.39
		MMS	0.1268	0.0125	1.27E-03	1.25E-04	1.39E-03	
CA227	**5L**	**YPD**	0.0067	0.0165	6.70E-05	1.65E-04	2.32E-04	2.25
		MMS	0.0325	0.0198	3.25E-04	1.98E-04	5.23E-04	
CA228	**5L**	**YPD**	0.0386	0.0008	3.86E-04	8.00E-06	3.94E-04	6.68
		MMS	0.2613	0.0017	2.61E-03	1.70E-05	2.63E-03	

* Conditions; YPD-control, HS-Heat shock, or MMS-methyl methanesulfonate
B+: BFP monofluorescent population, G+: GFP monofluorescent population

Table S2: List of strains and plasmids used throughout this study.

Strain		Genotype	Location of BFP-*HIS1* /GFP-*ARG4*	Ref. *
C. albicans	**SN148**	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434*	-	(1)
	CA92	SN148 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-Cm*LEU2*	Chr2L	(2)
	CA93	SN148 Ca22chr3_C_albicans_SC5314:129967 to 130211::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr3L	(2)
	CA94	SN148 Ca22chr4_C_albicans_SC5314:128561 to 129559::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr4L	(2)
	CA100	SN148 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr2R	(2)
	CA101	SN148 Ca22chr3_C_albicans_SC5314:1726068 to 1726326::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr3R	(2)
	CA105	SN148 Ca22chrR_C_albicans_SC5314:2085021 to 2085583::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	ChrRR	(2)
	CA108	SN148 Ca22chr3_C_albicans_SC5314:129967 to 130211::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr3L	(2)
	CA109	SN148 Ca22chr4_C_albicans_SC5314:128561 to 129559::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr4L	(2)
	CA110	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr5L	(2)
	CA116	SN148 Ca22chr3_C_albicans_SC5314:1726068 to 1726326::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr3R	(2)
	CA117	SN148 Ca22chr4_C_albicans_SC5314:1452791 to 1453180::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr4R	(2)
	CA119	SN148 Ca22chr7_C_albicans_SC5314:911053 to 911347::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1 RPS1*::Clp-BC-*URA3*/Clp-*CmLEU2*	Chr7R	(2)
	CA124	SN148 Ca22chr1_C_albicans_SC5314:145484 to 145771::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-URA3-CmLEU2	Chr1L	(2)
	CA126	SN148 Ca22chr1_C_albicans_SC5314:145484 to 145771::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-URA3-CmLEU2	Chr1L	(2)
	CA130	SN148 Ca22chr1_C_albicans_SC5314:3109738 to 3110154::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr4_C_albicans_SC5314:776480 to 777156::BC-URA3-CmLEU2	Chr1R	(2)
	CA134	SN148 Ca22chr5_C_albicans_SC5314:1083848 to 1084278::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 *RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr5R	(2)
	CA144	SN148 Ca22chr6_C_albicans_SC5314:156134 to 156430::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1_RPS1/RPS1::Clp-CmLEU2-BC-URA3	Chr6L	(2)
	CA156	SN148 Ca22chrR_C_albicans_SC5314:79361 to 79780::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1_RPS1/RPS1::Clp-CmLEU2-BC-URA3	ChrRL	(2)
	CA164	SN148 Ca22chr2_C_albicans_SC5314:2007084 to 2007475::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1_RPS1/RPS1::Clp-CmLEU2-BC-URA3	Chr2R	(2)
	CA184	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1_RPS1/RPS1::Clp-CmLEU2-BC-URA3	Chr5L	(2)
	SN95	*arg4Δ/arg4Δ his1Δ/his1Δ URA3/ura3Δ::λimm434 IRO1/iro1Δ::λimm434*	-	(1)
	CA199	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA200	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA201	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA202	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA203	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA204	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA206	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA207	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA208	SN95 Ca22chr2_C_albicans_SC5314:252055 to 252382::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2L	This study
	CA209	SN95 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2R	This study
	CA211	SN95 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2R	This study
	CA212	SN95 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::Clp-*CmLEU2*-BC-*URA3*	Chr2R	This study

	CA215	SN95 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr2R	This study
	CA216	SN95 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr2R	This study
	CA217	SN95 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr2R	This study
	CA218	SN95 Ca22chr2_C_albicans_SC5314:2007084 to 2007290::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr2R	This study
	CA219	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA220	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA221	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA222	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA223	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA224	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA225	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA226	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA227	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA228	SN95 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_RPS1/RPS1*::ClpX-*CmLEU2*-BC-*URA3*	Chr5L	This study
	CA183	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1*	Chr5L	This study
	CA229	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_GAL1/gal1Δ*::*URA3*	Chr5L	This study
	CA230	SN148 Ca22chr5_C_albicans_SC5314:115379 to 115802::P_{TDH3}-*GFP-CdARG4*/P_{TDH3}-*BFP-CdHIS1_GAL1/gal1Δ*::*URA3*	Chr5L	This study
E. coli		pCRBluntII-P_{TDH3}-*GFP-ARG*4	-	(3)
		pCRBluntII-P_{TDH3}-*BFP*-Cd*HIS*1	-	(3)
		pV1093	-	(4)
		pFA-*URA3*	-	(5)

* (1) Noble *et al.* 2005, (2) Marton *et al.* 2020, (3) Loll-Krippleber *et al.* 2015, (4) Vyas *et al*. 2015, (5) Gola *et al.* 2003

Table S3: Primers used throughout this study.

Primer	Name	Sequence	Usage
TM54	K7_BFP-GFP_Chr2_Left_F	ATCTCCTTCCTCCTCCACCTCTTAGTCTTAG TCTTCTAACCGTGATAGTCAACAATTAGTAC CATTGTTCTATTCCTTCATGACGTTATTTCTT TCCTGA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM55	K7_BFP-GFP_Chr2_Left_R	ACAACGACGACGCCGTGGACAATACACGG AGTATTATTGAACGCCGAACAAATTACTTTG AGAATGTATGTTTAAAACTTTGCAGCTGCAA CCTACAATG**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM66	K7_BFP-GFP_Chr5_Left_F	TTTATAATATCAATGGGGGAAATGTACGTCA TAAATGATATCTCTTCAATACTGTTCAGATT GTACAATTGCGAAATCTGTACGCAACTGTC ATATTACA**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM67	K7_BFP-GFP_Chr5_Left_R	TTATTTCCTCTTTAGTCCAATTTTGTTATTGT TGTTATTTAATTAACTTAACTAATTAAGTAAC CACAATAATACTAGTGGTGACGACAGCGGA AGTGGA**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM76	Verif_K7_BFP-GFP_C2_Left_F	TTGGTTCATTGGCGGTGGTA	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM77	Verif_K7_BFP-GFP_C2_Right_F	AAACATGCACCACGAGGGAT	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM82	Verif_K7_BFP-GFP_C5_Left_F	TTCTTGTTTCGCCAAGCTGC	5' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM89	Verif_K7_BFP-GFP_C2_Left_R	AAATGGGATTGTGGAGGGTG	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM90	Verif_K7_BFP-GFP_C2_Right_R	TGTCGCCGAACGCAATCTAT	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM95	Verif_K7_BFP-GFP_C5_Left_R	ACTTGATCACCAAACGGGCA	3' verification of integration, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM153	sgRNA-Chr2L-F	GGAAAGAAATAACGTCATGAGTTTTAGAGC TAGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM154	SNR52-Chr2L-R	TCATGACGTTATTTCTTTCCCAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM155	sgRNA-Chr2R-F	TTGCAAAGATTACTAGGATAGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM156	SNR52-Chr2R-R	TATCCTAGTAATCTTTGCAACAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM164	newK7_BGcol_C2L_cas9_F	TGTTTTAGATTCTAACTTAAACCTCCTTACT CATTACTCTTTCATTTTCTATCATTTCTTGGA TGTGACTAAAAATCTTACCACCACCGCCAT CTCCTTC**CAGGAAACAGCTATGACC**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM165	newK7_BGcol_C2R_cas9_R	AGATGTGAACACTTGGGACCACTACTTTGT GTATTTTCGTGAGTAATCGGCACTGATCCC GAATTATTCGTACAAAGGCTTACTTGTTATG AGCATCCAC**GTTTTCCCAGTCACGACG**	For production of transformation cassettes, Collection of strains bearing BFP/GFP System on all chromosomes both arms
TM171	sgRNA-Chr5L-F	TTATGCATATAATATATCTCGTTTTAGAGCT AGAAATAGCAAGTTAAA	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM172	SNR52-Chr5L-R	GAGATATATTATATGCATAACAAATTAAAAA TAGTTTACGCAAGTC	Guide RNA primers for CRISPR cas9, in red is the target sequence
TM173	SNR52_F1	AAGAAAGAAAGAAAACCAGGAGTGAA	Construction of guide RNA primers for CRISPR cas9, Forward primer for amplification of SNR52 promoter
TM174	sgRNA_R2	ACAAATATTTAAACTCGGGACCTGG	Construction of guide RNA primers for CRISPR cas9, Reverse primer for amplification of sgRNA scaffold

TM175	sgRNA_N3	GCAGCTCAGTGATTAAGAGTAAAGATGG	Construction of guide RNA primers for CRISPR cas9, Forward and reverse nested primers for third round PCR for construction of sgRNA expression cassette
TM176	SNR52_N3	GCGGCCGCAAGTGATTAGACT	Construction of guide RNA primers for CRISPR cas9, Forward and reverse nested primers for third round PCR for construction of sgRNA expression cassette
TM177	CaCas9/for	ATCTCATTAGATTTGGAACTTGTGGGTT	Forward and reverse primers for amplification of *CaCas9* cassette
TM178	CaCas9/rev	TTCGAGCGTCCCAAAACCTTCT	Forward and reverse primers for amplification of *CaCas9* cassette
TM115	GFP-end-F	CCAATTGGTGATGGTCC	GFP integration verification
TM106	GFP_Rev	CGGAGACAGAAAATTTGTGACC	Used to validate 5' junction when integrating pTDH3-ARG4-GFP cassettes, internal oligo.
RLO066	TDH3p-BFP-R	TCTTTAATCAATTCAGACATTTTAATAAAGCT TCTGCAGG	Used to validate 5' junction when integrating pTDH3-HIS1-BFP cassettes, internal oligo.
ML42	CdHIS1-detect-R	TGCCCTTCTACCTGGAGTAATGGT	Used to validate 3' junction when integrating pTDH3-HIS1-BFP cassettes, internal oligo.
CipUL	CipUL	ATACTACTGAAAATTTCCTGACTTTC	Used to validate integrating at *RPS*1 locus.
CIpUR	CIpUR	ATTACTATTTACAATCAAAGGTGGTC	Used to validate integrating at *RPS*1 locus.
ML	CaGAL1_KO_pFA_F	TGTTTGCATGTGATTTGTTTTGACCATATG GTAGTTGCGATATTCGTCGCCCTATCTATTT TTGCAACAGACTAATACCCAACTCACTCACT GTTATCTGAAGCTTCGTACGCTGACGGT	For production of transformation cassettes targeting *GAL1* locus
ML	CaGAL1_KO_pFA_R	ACACATACAAACCGATTTTTATAAGAAAGA GTTATACCATGTCAGTTCCTACGTTTGATGA TTTATCATTTTATTCCAATGATCAAGAATTAA CTAAATTCTGATATCATCGATGAATTCGAG	For production of transformation cassettes targeting *GAL1* locus
ML33	CaGAL1+1848-F	GGTGATATTAGGTAGTGCTG	Verification of integration *GAL1* locus
AF45	URA3_rev_AF	GTTGTCCTAATCCATCACCT	Verification of integration *GAL1* locus

Perspectives

Throughout the course of my PhD work, I was lucky to have the liberty to explore LOH events in *C. albicans* from various perspectives as described in the five above chapters giving rise to a comprehensive study of this type of genomic rearrangement. As each chapter has been thoroughly discussed individually, this section will primarily focus on the overall impact and perspectives of each project.

- *Patterns of LOH within clinical isolates of C. albicans*

As much of our knowledge regarding LOH events within *C. albicans* derives from SNV and CNV calling data or laboratory microevolution experiments, we implemented a pipeline allowing to accurately identify LOH events and explore LOH landscapes within a collection of natural isolates. Using 54 clinical strains from the genetic cluster 1 of the population of *C. albicans*, we confirmed that LOH are pervasive, revealed that the median LOH size is roughly 1 kb and that the majority of LOH are small (≤10kb). Indeed, the potential impact of an LOH event is theoretically associated to its size, as larger LOH affect numerous alleles leading to a higher probability of revealing a new phenotype. Additionally, maintenance of heterozygosity is thought to be important in *C. albicans* as it has been correlated with strain fitness thus, the observation of predominantly short-tract LOH could be a conservative genome rearrangement strategy. Although short-tract LOH are the most abundant, this cannot be translated into a direct LOH frequency because long-tract LOH events could emerge at a similar frequency but not be maintained within the population as they may be associated with a high fitness cost. Nevertheless, numerous long-tract LOH events distributed across the genome were also detected and often spanned from a specific locus until the telomere (distal LOH). Although LOH can affect all chromosomes, distribution of LOH within Cluster 1 revealed LOH poor regions (cold areas) and LOH rich regions (hotspots). Centromeres are poor in LOH tracts while frequency of LOH increases towards the telomeres. Additionally, we noted that certain MRS impact frequency of LOH as their location coincided with abrupt increases in LOH, hinting that some of these large repeat sequences could be hotspots of recombination. Alternatively, LOH cold areas of genome (excluding centromeres) could be investigated to understand why these regions maintain heterozygosity and rarely undergo recombination. This type of analysis could potentially reveal overdominant loci, where an advantageous phenotype is only associated to heterozygosity at this locus. Distribution of overdominant loci across chromosomes was shown to participate in the maintenance of heterozygosity by limiting LOH in an asexual lineage of Cape honey bee (*Apis mellifera capensis*) (Goudie *et al.* 2014). Hence, such loci could participate in the overall maintenance of heterozygosity in *C. albicans*

although, this could also be explained by genetic shuffling occurring during the parasexual life cycle.

Restricted to Cluster 1, our analysis of patterns of LOH within *C. albicans* revealed a similar LOH landscape as described in microevolution experiments conducted *in vitro* and *in vivo* (Ene *et al.* 2018). Indeed, the logical next step of this study is expanding our analysis to other *C. albicans* clades towards potentially unveiling partners of LOH across the 18 genetic clusters of the population. This type of LOH data would also permit to eventually uncover cluster specific LOH events and permit to study the links between LOH and genetics of *C. albicans* biology. Population studies indicated no association between sampling niche and phylogenetic clustering of *C. albicans* strains. However, Forche *et al.* recently highlighted that the oral niche seems to select for trisomy of Chr5 and Chr6 in *C. albicans* (Forche *et al.* 2019). Therefore, it would be interesting to use LOH data in an attempt to identify potential niche specific patterns of LOH. However, ultimately, the associations between LOH and *C. albicans* biology would be highlighted upon genome-wide association studies (GWAS) aimed to explore genotype and phenotype associations. Studies investigating LOH distribution across genomes have been shown to be highly informative in survival predictions of human cancers (Chen *et al.* 2015; Deryusheva *et al.* 2017). The results of a GWAS study conducted using SNV, CNV and LOH data could lead to the establishment of a genetic variations repertoire, of interest to (i) better understand the pathogenesis of opportunistic pathogen *C. albicans* and (ii) to be eventually used in clinic settings to identify the phenotypic potential of a strain based on the genetic variations it harbors.

- *Repeat sequences can alleviate constraints of recessive lethal alleles on LOH*

Chapter 1 shows the pervasiveness of LOH in the *C. albicans* population. Given the relatively elevated level of heterozygosity in its diploid genome, it is reasonable that LOH in *C. albicans* could impact phenotypic properties or even viability based on the functional differences between alleles, only revealed upon loss of a dominant allele. The presence of recessive lethal alleles can impact the directionality of LOH events and result in chromosome homozygosis biases, where a given haplotype is never observed at the homozygous state. LOH facilitate the exposure of beneficial or disadvantageous recessive alleles thus, the latter homozygosis biases can result from the uncovering of recessive lethal alleles upon LOH. Recessive lethal alleles often arise from the random accumulation of non-synonymous point mutations disrupting the function of a non-dispensable protein. Such recessive alleles are often lineage specific and not shared by the entire population. In Chapter 2, we focused on the SC5314 genetic background, the *C. albicans* laboratory reference strain, and identified 13 candidate alleles possessing premature-STOP codons, of which two were confirmed as recessive lethal alleles and one as a deleterious recessive allele (Feri *et al.* 2016; Marton *et*

al. 2019). Although such alleles are dependent on genetic background, these studies highlighted the fact that LOH can be constrained by genetic content, as certain LOH events are not passed to the next generation. Furthermore, during the identification of the recessive lethal allele *mtr4*^{K880*}*,* situated on the right arm of Chr7, we confirmed that repeat sequences, i.e. MRS, are intra-chromosome recombination hotspots which could alleviate constraints of recessive lethal alleles upon LOH. This role can specifically be important upon long-tract LOH events spanning the entire chromosome arm, as hypothetical recombination between each homolog MRS would allow both alleles of centromere proximal coding regions to be maintained in the population. My work proposed a hypothesis on the unknown role of MRS, potentially explaining why these repeat sequences are being positively selected in *C. albicans* species.

- *Extension of homozygosis towards centromere upon DNA-DSB repair by break-induced replication*

LOH being the main interest of my PhD research, I was also interested in tackling the mechanistic aspect of LOH origins. It was already known that DNA-DSB are major sources of LOH but DNA repair pathways in *C. albicans* remain poorly studied and to the best of my knowledge, the fidelity of DNA-DSB repair has not been thoroughly investigated in this species. Chapter 3 was dedicated to better understand the fidelity of break-induced replication (BIR) in *C. albicans*, which often causes long-tract LOH events. We demonstrated that, similarly to *S. cerevisiae*, BIR in *C. albicans* is not very precise as it is often associated to a median 8.3 kb of additional homozygosis beyond the break-site towards the centromere. By characterizing the deletion effect of various key genes implicated in HR-dependent pathways (*POL32, MPH1, MUS81* or *SPO11*) or mismatch repair (MMR) (*MSH2*), we highlighted that these unexpected ascending LOH regions seem to result from both intrinsic features of BIR, undoubtedly due to extensive 5' resection, and potentially the participation of the MMR pathway, repairing naturally heterozygous positions. Additionally, this study permitted us to investigate the role of three genes (*POL32, MPH1* and *MUS81*), whose role on genome stability had not been previously studied in *C. albicans*. Homozygous deletion of either latter three genes lead to an increased genomic instability in *C. albicans,* as *mus81*$^{-/-}$ displayed the most drastic phenotype followed by *mph1*$^{-/-}$ and *pol32*$^{-/-}$. Notably, our characterization confirmed that *POL32* plays an important role in BIR and suggests that it is potentially dispensable for GC in *C. albicans,* contrary to its role in *S. cerevisiae.* In contrast, our characterization shows that *MUS81* is implicated in the CO in *C. albicans* and dysfunction of this gene can promote aneuploidy. Lastly, characterization *mph1*$^{-/-}$ confirmed the involvement of *MPH1* in NCO outcomes of HR-mediated repair in *C. albicans.*

Two types of BIR repair have been described in yeast, Rad51-dependent and -independent. Following 5' resection of DNA-DSB, the Rad51 protein binds to the 3' single strand DNA and forms a Rad51 nucleoprotein filament implicated in the homology search. In *S. cerevisiae*, the Rad51-independent BIR is often facilitated by a sequence called FBI (Facilitator Of BIR), such as an inverted repeat (Malkova *et al.* 2001; Downing *et al.* 2008), situated ~13 kb proximal to the break-site and ~30 kb away from the centromere. The Rad51-independent DSB repair mediated by FBI results in ascending LOH towards the centromere (Malkova *et al.* 2001). Unfortunately, at the time when we selected our candidate genes for our study of BIR fidelity in *C. albicans*, we were unaware of the above presented findings in *S. cerevisiae*. Thus, we did not investigate which type of BIR is associated with homozygosis between the break-site and the centromere because we failed to include *RAD51.* Indeed, to complete this chapter, we plan to include *RAD51* into the analysis by studying the effect of *RAD51* deletion on ascending LOH tract towards the centromere in *C. albicans*.

- *Mutagenic effects of transformation in C. albicans*

In Chapter 4, I took advantage of the genome sequencing data generated in the course of the construction of a collection of strains dedicated to study whole-genome LOH dynamics in *C. albicans* (see Chapter 5), to investigate the LOH and aneuploidy landscapes in strains constructed using CRISPR-Cas9-free or CRISPR-Cas9-dependent transformation procedures. The transformation process in *C. albicans* is known to be mutagenic, leading to both LOH and aneuploidy events. These genomic rearrangements are thought to notably arise during the heat shock procedure or upon exposure to lithium acetate (Bouchonville *et al.* 2009). The CRISPR-Cas9 system is increasingly implemented in genome modification procedures of numerous organisms including *C. albicans*, although its potential undesired side-effects are unknown. Thus, we compared the number of undesired LOH and aneuploidy events in our final construct strains, to approximate mutagenesis of CRISPR-Cas9-free and -dependent transformation approaches. Despite few limitations of this study (certain differences in constructs and the fact that strain production was not explicitly designed for this type of comparative study), we used whole-genome sequencing to show that we reduce the mutagenic effect of transformation by targeting DNA-DSB at the target locus, as the overall number of trisomy events is decreased in CRISPR-Cas9-dependent protocol. Upon DNA-DSB, the DNA-damage response is activated and the DNA machinery is recruited to the target locus increasing the frequency of HR and overall number of transformants obtained, consequently reducing the probability of selecting transformants possessing (concurrent) genomic rearrangements. Certainly, transformation of *C. albicans* can also be mediated by electroporation, with exposure of cells to an electric field to promote the passage of DNA into the cells. We have not explored the mutagenic effects of electroporation, which does not

require heat shock, with regards to the appearance of LOH and aneuploidy. Overall, this study highlights the importance of (i) extensive transformants characterization (preferably genome-wide) prior to subsequent analysis and (ii) of complementation experiments during reverse genetic studies.

- *Genome-wide LOH dynamics*

As I was becoming more aware of the complexity of the relationship between LOH occurrence and phenotypic outcomes, I was interested in investigating the whole dynamics of genome-wide LOH and the acquisition of adaptive phenotypes, which was the starting point for Chapter 5. While assessing LOH frequency in various genomic loci, it became evident that a high variability in terms of genome stability existed in *C. albicans* not only at the locus level but also at the strain level, highlighting inter-strain, intra-strain and inter-chromosome variations. Consequently, this heterogeneity could participate in broadening the phenotypic spectrum and act as an important adaptive strategy for *C. albicans*. The origins of this LOH frequency variability described in Chapter 5 remains to be explored since, as mentioned in the discussion of Chapter 5, several biological aspects may be associated to the generation of LOH heterogeneity, including (i) HR gene expression noise, (ii) variations in DNA packaging or (iii) variations in chromosome organization within the nucleus. Inevitably, future experiments would involve single cell analysis in order to (i) identify inter-cell variation and (ii) investigate their potential implications on heterogeneity in terms of genome stability. Genetic background also appears to impact genome stability although we did not evaluate the extent of genetic variations between strains required to observe an impact on LOH dynamics. Lastly, as multiple stochastic parameters seem to influence genome dynamics in *C. albicans*, it would be interesting to conduct a mathematical modeling study aimed at predicting parameters which could influence the appearance of LOH, in collaboration with a research team which is familiar with mathematical modeling systems.

Originally, we aimed to utilize the Blue-Green collection for *in vivo* experiments allowing us to assess genome-wide LOH dynamics within a murine GIT colonization model. The murine gut is a complex environment with multiple unknown parameters which could influence LOH frequency. Thus, I carried out a pilot study to explore (i) the effect of exposure to antibiotic treatments on LOH frequency, (ii) the distribution of colonization levels throughout the GIT, (iii) the growth rate of *C. albicans* in the mouse and (iv) within feces. As previously mentioned, *C. albicans* is not a natural colonizer of the mouse gut hence colonization requires the use of antibiotics, diet changes or immunosuppressor drugs. In the laboratory, we use the antibiotic-mediated GIT colonization model. Consequently, I tested the impact of various concentrations of these antibiotics (streptomycin, gentamicin, vancomycin and hygromycin)

on LOH frequency in *C. albicans* and observed no gross impact (unpublished data). However, as revealed in Chapter 5, intra-strain, inter-strain and inter-chromosome heterogeneity of LOH frequency is an important parameter which should now be taken into account. The quantification and distribution of *C. albicans* cells during GIT colonization is a crucial parameter permitting to estimate the available sampling size within a single mouse. Preliminary experiments revealed that within stably colonized mice *C. albicans* were quasi evenly distributed throughout the different sections of the GIT, independently of colonization levels. Colonization levels of *C. albicans* are commonly estimated by evaluating the quantity of CFUs per gram of feces with a maximum colonization level of roughly 10^7 CFU/g of stool. Within a mouse GIT, possessing a colonization level of 10^7 CFU/g, I approximated that roughly 10^7 *C. albicans* cells reside in the mouse. Additionally, because mice are known as coprophagic animals, I also investigated if *C. albicans* continued to evolve in excrements which could bias our evaluation of LOH frequencies within the GIT. *C. albicans* thrives well in feces, displaying a doubling time of roughly 4h at room temperature (unpublished data). Indeed, yeast cells can continue to evolve once excreted from the host. Lastly, with the hope of properly comparing *in vitro* versus *in vivo* LOH frequencies, by considering the mouse as a chemostat we estimated *C. albicans* growth rate at 0.03 generations/h (unpublished data). In the GIT, growth of *C. albicans* is 2x slower than the growth rate estimated in the murine systemic infection model (0.065 generations/h.) and 20x slower than in standard laboratory growth conditions (0.62 generations/h.) (Forche *et al.* 2009a). Overall, the major limitation of studying LOH dynamics in murine GIT model is the limited number of recoverable *C. albicans* cells. As highlighted in Chapter 5, a reliable estimation of LOH frequency requires an enormous number of *C. albicans* cells thus, an unrealistic number of fecal pellets or many animals per replicate are necessary. Additionally, our finding of elevated levels of LOH frequency variations *in vitro* gravely complexifies the estimation of LOH frequency *in vivo*. Alternatively, the effect of various GIT parameters (environmental and microbiota components) on LOH could be assessed using a simulator of the Human Intestinal Microbial Ecosystem SHIME® (Van de Wiele *et al.* 2015), a system which mimics *in vitro* the host GIT and permits higher throughput analysis. This type of study is a stepping-stone towards investigating the interaction between *C. albicans* genome stability and host and/or microbiota, notably of interest the effect of the bacterial components of the microbiota which produce genotoxins affecting host eukaryotic cells.

Throughout this work, it was recurrently mentioned that two key characteristics of the genome of *C. albicans* are its relatively elevated heterozygosity and its high tolerance for genomic instability, attributes thought to participate in its success as an opportunistic pathogen. However, it is important to have an equilibrium between genomic plasticity and

stability. This is particularly true regarding the appearance of LOH, as demonstrated throughout this thesis. Certainly, LOH can be beneficial by allowing fast adaptation to an environment through fixation of beneficial alleles or purge of disadvantageous alleles (e.g. recessive lethal) from the population. However, maintenance of heterozygosity is equally important as it correlates with an elevated strain fitness. In *C. albicans*, the appearance of LOH is influenced by multiple parameters in both natural and laboratory derived isolates. Some of these parameters remain unknown although, in this thesis, the importance and involvement of repeat sequences, notably MRS, and their recombinogenic properties were highlighted as valuable tools which mediate LOH events in *C. albicans*. Another emerging hallmark of *C. albicans* has also been highlighted in this work, namely heterogeneity of genomic instability within the population. This characteristic can be highly advantageous in a clonal population continuously exposed to an evolving environment. This strategy would permit to quickly modulate (at different extents) existing genetic variations within the population using genomic rearrangements to produce a wide-spectrum of combinations, in hopes that few or one is associated to an advantageous outcome eventually permitting perpetuity.

Bibliography

Abbey D., M. Hickman, D. Gresham, and J. Berman, 2011 High-Resolution SNP/CGH microarrays reveal the accumulation of loss of heterozygosity in commonly used *Candida albicans* strains. G3 1. https://doi.org/10.1534/g3.111.000885

Abbey D. A., J. Funt, M. N. Lurie-Weinberger, D. A. Thompson, A. Regev, *et al.*, 2014 YMAP: a pipeline for visualization of copy number variation and loss of heterozygosity in eukaryotic pathogens. Genome Med. 6: 1–16. https://doi.org/10.1186/s13073-014-0100-8

Ahmad A., M. A. Kabir, A. Kravets, E. Andaluz, G. Larriba, *et al.*, 2008 Chromosome instability and unusual features of some widely used strains of *Candida albicans*. Yeast 25: 433–448. https://doi.org/10.1002/yea.1597

Anand R. P., S. T. Lovett, and J. E. Haber, 2013 Break-induced DNA replication. Cold Spring Harb. Perspect. Biol. 5: a010397–a010397. https://doi.org/10.1101/cshperspect.a010397

Anand R. P., O. Tsaponina, P. W. Greenwell, C.-S. Lee, W. Du, *et al.*, 2014 Chromosome rearrangements via template switching between diverged repeated sequences. Genes Dev. 28: 2394–2406. https://doi.org/10.1101/gad.250258.114

Andaluz E., R. Calderone, G. Reyes, and G. Larriba, 2001 Phenotypic Analysis and Virulence of *Candida albicans LIG4* Mutants. Infect. Immun. 69: 137 LP – 147. https://doi.org/10.1128/IAI.69.01.137-147.2001

Andaluz E., T. Ciudad, and G. Larriba, 2002 An evaluation of the role of *LIG4* in genomic instability and adaptive mutagenesis in *Candida albicans*. FEMS Yeast Res. 2: 341–348. https://doi.org/10.1016/S1567-1356(02)00094-6

Andaluz E., A. Bellido, J. Gómez-Raja, A. Selmecki, K. Bouchonville, *et al.*, 2011 Rad52 function prevents chromosome loss and truncation in *Candida albicans*. Mol. Microbiol. 79: 1462–1482. https://doi.org/10.1111/j.1365-2958.2011.07532.x

Andersen M. P., Z. W. Nelson, E. D. Hetrick, and D. E. Gottschling, 2008 A genetic screen for increased loss of heterozygosity in *Saccharomyces cerevisiae*. Genetics 179: 1179–1195. https://doi.org/10.1534/genetics.108.089250

Anderson M., J. Haase, E. Yeh, and K. Bloom, 2009 Function and Assembly of DNA Looping, Clustering, and Microtubule Attachment Complexes within a Eukaryotic Kinetochore. Mol. Biol. Cell 20: 4131–4139. https://doi.org/10.1091/mbc.e09-05-0359

Anderson M. Z., J. A. Baller, K. Dulmage, L. Wigen, and J. Berman, 2012 The Three Clades of the Telomere-Associated TLO Gene Family of *Candida albicans* Have Different Splicing, Localization, and Expression. Eukaryot. Cell 11: 1268 LP – 1275. https://doi.org/10.1128/EC.00230-12

Anderson M. Z., L. J. Wigen, L. S. Burrack, and J. Berman, 2015 Real-Time Evolution of a Subtelomeric Gene Family in *Candida albicans*. Genetics 200: 907 LP – 919. https://doi.org/10.1534/genetics.115.177451

Anderson M. Z., G. J. Thomson, M. P. Hirakawa, and R. J. Bennett, 2019 A “parameiosis” drives depolyploidization and homologous recombination in *Candida albicans*. Nat. Commun. 10: 4388. https://doi.org/10.1038/s41467-019-12376-2

Arbour M., E. Epp, H. Hogues, A. Sellam, C. Lacroix, *et al.*, 2009 Widespread occurrence of chromosomal aneuploidy following the routine production of *Candida albicans* mutants. FEMS Yeast Res. 9: 1070–1077. https://doi.org/10.1111/j.1567-1364.2009.00563.x

Bachellier-Bassi S., and C. d’Enfert, 2015 Chemical Transformation of *Candida albicans* BT - Genetic Transformation Systems in Fungi, Volume 1, pp. 81–85 in edited by Berg M. A. van den, Maruthachalam K. Springer International Publishing, Cham.

Backer M. D. De, D. Maes, S. Vandoninck, M. Logghe, R. Contreras, *et al.*, 1999 Transformation of *Candida albicans* by electroporation. Yeast 15: 1609–1618. https://doi.org/10.1002/(sici)1097-0061(199911)15:15<1609::aid-yea485>3.3.co;2-p

Banakar R., A. L. Eggenberger, K. Lee, D. A. Wright, K. Murugan, *et al.*, 2019 High-frequency random DNA insertions upon co-delivery of CRISPR-Cas9 ribonucleoprotein and selectable marker plasmid in rice. Sci. Rep. 9: 19902. https://doi.org/10.1038/s41598-019-55681-y

Barrangou R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, *et al.*, 2007 CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science (80-.). 315: 1709 LP – 1712. https://doi.org/10.1126/science.1138140

Bauer N. C., A. H. Corbett, and P. W. Doetsch, 2015 The current state of eukaryotic DNA base damage and repair. Nucleic Acids Res. 43: 10083–10101. https://doi.org/10.1093/nar/gkv1136

Baum M., K. Sanyal, P. K. Mishra, N. Thaler, and J. Carbon, 2006 Formation of functional centromeric chromatin is specified epigenetically in *Candida albicans*. Proc. Natl. Acad. Sci. 103: 14877 LP – 14882. https://doi.org/10.1073/pnas.0606958103

Beekman C. N., and I. V Ene, 2020 Short-term evolution strategies for host adaptation and drug escape in human fungal pathogens. PLOS Pathog. 16: e1008519.

Bellido A., T. Ciudad, B. Hermosa, E. Andaluz, A. Forche, *et al.*, 2019 Partner Choice in Spontaneous Mitotic Recombination in Wild Type and Homologous Recombination Mutants of *Candida albicans*. G3 (Bethesda). 9: 3631–3644. https://doi.org/10.1534/g3.119.400516

Bennett R. J., and A. D. Johnson, 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

Beranek D. T., 1990 Distribution of methyl and ethyl adducts following alkylation with monofunctional alkylating agents. Mutat. Res. Mol. Mech. Mutagen. 231: 11–30. https://doi.org/https://doi.org/10.1016/0027-5107(90)90173-2

Bergmann O., S. Zdunek, A. Felker, M. Salehpour, K. Alkass, *et al.*, 2015 Dynamics of Cell Generation and Turnover in the Human Heart. Cell 161: 1566–1575. https://doi.org/10.1016/j.cell.2015.05.026

Berman J., 2016 Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. FEMS Yeast Res. 16. https://doi.org/10.1093/femsyr/fow020

Bernstein K. A., S. Granneman, A. V Lee, S. Manickam, and S. J. Baserga, 2006 Comprehensive Mutational Analysis of Yeast DEXD/H Box RNA Helicases Involved in Large Ribosomal Subunit Biogenesis. Mol. Cell. Biol. 26: 1195–1208. https://doi.org/10.1128/MCB.26.4.1195-1208.2006

Bétermier M., P. Bertrand, and B. S. Lopez, 2014 Is non-homologous end-joining really an inherently error-prone process? PLoS Genet. 10: e1004086. https://doi.org/10.1371/journal.pgen.1004086

Bhargava R., D. O. Onyango, and J. M. Stark, 2016 Regulation of Single-Strand Annealing and its Role in Genome Maintenance. Trends Genet. 32: 566–575. https://doi.org/https://doi.org/10.1016/j.tig.2016.06.007

Bhattacharyya M. K., and A. J. Lustig, 2006 Telomere dynamics in genome stability. Trends Biochem. Sci. 31: 114–122. https://doi.org/https://doi.org/10.1016/j.tibs.2005.12.001

Boeke J. D., F. LaCroute, and G. R. Fink, 1984 A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. 197: 345–346. https://doi.org/10.1007/bf00330984

Boiteux S., and S. Jinks-Robertson, 2013 DNA Repair Mechanisms and the Bypass of DNA Damage in *Saccharomyces cerevisiae*. Genetics 193: 1025 LP – 1064. https://doi.org/10.1534/genetics.112.145219

Bongomin F., S. Gago, R. O. Oladele, and D. W. Denning, 2017 Global and Multi-National Prevalence of Fungal Diseases-Estimate Precision. J. fungi (Basel, Switzerland) 3. https://doi.org/10.3390/jof3040057

Bordallo-Cardona M. Á., C. Agnelli, A. Gómez-Nuñez, C. Sánchez-Carrillo, E. Bouza, *et al.*, 2019 *MSH2* Gene Point Mutations Are Not Antifungal Resistance Markers in *Candida glabrata*. Antimicrob. Agents Chemother. 63. https://doi.org/10.1128/AAC.01876-18

Borman A. M., A. Szekely, C. J. Linton, M. D. Palmer, P. Brown, *et al.*, 2013 Epidemiology, antifungal susceptibility, and pathogenicity of *Candida africana* isolates from the United Kingdom. J. Clin. Microbiol. 51: 967–972. https://doi.org/10.1128/JCM.02816-12

Bouchonville K., A. Forche, K. E. S. Tang, A. Selmecki, and J. Berman, 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., S. Morand, and C. D'Enfert, 2002 Usefulness of Multilocus Sequence Typing for Characterization of Clinical Isolates of *Candida albicans*. J. Clin. Microbiol. 40: 1290 LP – 1297. https://doi.org/10.1128/JCM.40.4.1290-1297.2002

Boyce R. P., and P. Howard-Flanders, 1964 Release of Ultraviolet Light-Induced Thymine Dimers from DNA in *E. coli* K-12. Proc. Natl. Acad. Sci. 51: 293 LP – 300. https://doi.org/10.1073/pnas.51.2.293

Boyce K. J., Y. Wang, S. Verma, V. P. S. Shakya, C. Xue, *et al.*, 2017 Mismatch Repair of DNA Replication Errors Contributes to Microevolution in the Pathogenic Fungus *Cryptococcus neoformans.* MBio 8. https://doi.org/10.1128/mBio.00595-17

Brown G. D., D. W. Denning, N. A. R. Gow, S. M. Levitz, M. G. Netea, *et al.*, 2012 Hidden Killers: Human Fungal Infections. Sci. Transl. Med. 4: 165rv13 LP-165rv13. https://doi.org/10.1126/scitranslmed.3004404

Burrack L. S., S. E. Applen Clancey, J. M. Chacón, M. K. Gardner, and J. Berman, 2013 Monopolin recruits condensin to organize centromere DNA and repetitive DNA sequences. Mol. Biol. Cell 24: 2807–2819. https://doi.org/10.1091/mbc.e13-05-0229

Burrack L. S., H. F. Hutton, K. J. Matter, S. A. Clancey, I. Liachko, *et al.*, 2016 Neocentromeres Provide Chromosome Segregation Accuracy and Centromere Clustering to Multiple Loci along a *Candida albicans* Chromosome. PLoS Genet. 12: e1006317–e1006317. https://doi.org/10.1371/journal.pgen.1006317

Buscaino A., R. Allshire, and A. Pidoux, 2010 Building centromeres: home sweet home or a nomadic existence? Curr. Opin. Genet. Dev. 20: 118–126. https://doi.org/https://doi.org/10.1016/j.gde.2010.01.006

Buscaino A., 2019 Chromatin-Mediated Regulation of Genome Plasticity in Human Fungal Pathogens. Genes (Basel). 10: 855. https://doi.org/10.3390/genes10110855

Bussotti G., E. Gouzelou, M. Côrtes Boité, I. Kherachi, Z. Harrat, *et al.*, 2018 Leishmania Genome Dynamics during Environmental Adaptation Reveal Strain-Specific Differences in Gene Copy Number Variation, Karyotype Instability, and Telomeric Amplification. MBio 9. https://doi.org/10.1128/mBio.01399-18

Butler G., M. D. Rasmussen, M. F. Lin, M. A. S. Santos, S. Sakthikumar, *et al.*, 2009 Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 459. https://doi.org/10.1038/nature08064

Chaillot J., M. A. Cook, J. Corbeil, and A. Sellam, 2017 Genome-Wide Screen for Haploinsufficient Cell Size Genes in the Opportunistic Yeast *Candida albicans*. G3 (Bethesda). 7: 355–360. https://doi.org/10.1534/g3.116.037986

Chauvel M., A. Nesseir, V. Cabral, S. Znaidi, S. Goyard, *et al.*, 2012 A Versatile Overexpression Strategy in the Pathogenic Yeast *Candida albicans*: Identification of Regulators of Morphogenesis and Fitness. PLoS One 7: e45912.

Chen Q., A. Ijpma, and C. W. Greider, 2001 Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events. Mol. Cell. Biol. 21: 1819–1827. https://doi.org/10.1128/MCB.21.5.1819-1827.2001

Chen X., B. B. Magee, D. Dawson, P. T. Magee, and C. A. Kumamoto, 2004 Chromosome 1 trisomy compromises the virulence of *Candida albicans*. Mol. Microbiol. 51: 551–565. https://doi.org/10.1046/j.1365-2958.2003.03852.x

Chen J., M. K. Ghorai, G. Kenney, and J. Stubbe, 2008 Mechanistic studies on bleomycin-mediated DNA damage: multiple binding modes can result in double-stranded DNA cleavage. Nucleic Acids Res. 36: 3781–3790. https://doi.org/10.1093/nar/gkn302

Chen X., D. Cui, A. Papusha, X. Zhang, C.-D. Chu, *et al.*, 2012 The Fun30 nucleosome remodeller promotes resection of DNA double-strand break ends. Nature 489: 576–580. https://doi.org/10.1038/nature11355

Chen C., Y. Zhang, M. M. Loomis, M. P. Upton, P. Lohavanichbutr, *et al.*, 2015 Genome-Wide Loss of Heterozygosity and DNA Copy Number Aberration in HPV-Negative Oral Squamous Cell Carcinoma and Their Associations with Disease-Specific Survival. PLoS One 10: e0135074.

Chibana H., J. L. Beckerman, and P. T. Magee, 2000 Fine-resolution physical mapping of genomic diversity in *Candida albicans*. Genome Res. 10: 1865–1877. https://doi.org/10.1101/gr.148600

Chibana H., and P. T. Magee, 2009 The enigma of the major repeat sequence of *Candida albicans*. Future Microbiol. 4: 171–179. https://doi.org/10.2217/17460913.4.2.171

Chico L., T. Ciudad, M. Hsu, N. F. Lue, and G. Larriba, 2011 The *Candida albicans* Ku70 Modulates Telomere Length and Structure by Regulating Both Telomerase and Recombination. PLoS One 6: e23732.

Choquer M., G. Robin, P. Le Pecheur, C. Giraud, C. Levis, *et al.*, 2008 Ku70 or Ku80 deficiencies in the fungus *Botrytis cinerea* facilitate targeting of genes that are hard to knock out in a wild-type context. FEMS Microbiol. Lett. 289: 225–232. https://doi.org/10.1111/j.1574-6968.2008.01388.x

Chu W. S., B. B. Magee, and P. T. Magee, 1993 Construction of an *SfiI* macrorestriction map of the *Candida albicans* genome. J. Bacteriol. 175: 6637 LP – 6651. https://doi.org/10.1128/jb.175.20.6637-6651.1993

Chung W.-H., Z. Zhu, A. Papusha, A. Malkova, and G. Ira, 2010 Defective Resection at DNA Double-Strand Breaks Leads to De Novo Telomere Formation and Enhances Gene Targeting. PLOS Genet. 6: e1000948.

Ciudad T., E. Andaluz, O. Steinberg-Neifach, N. F. Lue, N. A. R. Gow, *et al.*, 2004 Homologous recombination in *Candida albicans*: role of CaRad52p in DNA repair, integration of linear DNA fragments and telomere length. Mol. Microbiol. 53: 1177–1194. https://doi.org/10.1111/j.1365-2958.2004.04197.x

Ciudad T., M. Hickman, A. Bellido, J. Berman, and G. Larriba, 2016 Phenotypic Consequences of a Spontaneous Loss of Heterozygosity in a Common Laboratory Strain of *Candida albicans*. Genetics 203: 1161–1176. https://doi.org/10.1534/genetics.116.189274

Ciudad T., A. Bellido, B. Hermosa, E. Andaluz, and G. Larriba, 2020 *DLH1*, the *Candida albicans* homologue of the meiosis-specific *DMC1*, is not involved in DNA repair but catalyses spontaneous interhomologue recombination and might promote non-crossover events. Cell. Microbiol. 22: e13137. https://doi.org/10.1111/cmi.13137

Coste A., V. Turner, F. Ischer, J. Morschhäuser, A. Forche, *et al.*, 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., A. Selmecki, A. Forche, D. Diogo, M.-E. Bougnoux, *et al.*, 2007 Genotypic Evolution of Azole Resistance Mechanisms in Sequential *Candida albicans* Isolates. Eukaryot. Cell 6: 1889–1904. https://doi.org/10.1128/EC.00151-07

Covo S., 2020 Genomic Instability in Fungal Plant Pathogens. Genes (Basel). 11. https://doi.org/10.3390/genes11040421

Darmon E., and D. R. F. Leach, 2014 Bacterial genome instability. Microbiol. Mol. Biol. Rev. 78: 1–39. https://doi.org/10.1128/MMBR.00035-13

Davies A. A., E. C. Friedberg, A. E. Tomkinson, R. D. Wood, and S. C. West, 1995 Role of the Rad1 and Rad10 proteins in nucleotide excision repair and recombination. J. Biol. Chem. 270: 24638–24641. https://doi.org/10.1074/jbc.270.42.24638

Davis A. P., and L. S. Symington, 2004 RAD51-dependent break-induced replication in yeast. Mol. Cell. Biol. 24: 2344–2351. https://doi.org/10.1128/mcb.24.6.2344-2351.2004

Deem A., A. Keszthelyi, T. Blackgrove, A. Vayl, B. Coffey, *et al.*, 2011 Break-Induced Replication Is Highly Inaccurate. PLOS Biol. 9: e1000594.

Dellière S., K. Healey, M. Gits-Muselli, B. Carrara, A. Barbaro, *et al.*, 2016 Fluconazole and Echinocandin Resistance of *Candida glabrata* Correlates Better with Antifungal Drug Exposure Rather than with *MSH2* Mutator Genotype in a French Cohort of Patients Harboring Low Rates of Resistance. Front. Microbiol. 7: 2038. https://doi.org/10.3389/fmicb.2016.02038

Deryusheva I. V, M. Tsyganov, E. Y. Garbukov, M. K. Ibragimova, J. G. Kzhyshkovska, *et al.*, 2017 Genome-wide association study of loss of heterozygosity and metastasis-free survival in breast cancer patients. Exp. Oncol. 39: 145–150.

Dianov G. L., N. Souza-Pinto, S. G. Nyaga, T. Thybo, T. Stevnsner, *et al.*, 2001 Base excision repair in nuclear and mitochondrial DNA. Prog. Nucleic Acid Res. Mol. Biol. 68: 285–297. https://doi.org/10.1016/s0079-6603(01)68107-8

DiCarlo J. E., J. E. Norville, P. Mali, X. Rios, J. Aach, *et al.*, 2013 Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 41: 4336–4343. https://doi.org/10.1093/nar/gkt135

Diogo D., C. Bouchier, C. d'Enfert, and M.-E. Bougnoux, 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

Doetsch P. W., and R. P. Cunningham, 1990 The enzymology of apurinic/apyrimidinic endonucleases. Mutat. Res. Repair 236: 173–201. https://doi.org/https://doi.org/10.1016/0921-8777(90)90004-O

Downing B., R. Morgan, K. VanHulle, A. Deem, and A. Malkova, 2008 Large inverted repeats in the vicinity of a single double-strand break strongly affect repair in yeast diploids lacking Rad51. Mutat. Res. 645: 9–18. https://doi.org/10.1016/j.mrfmmm.2008.07.013

Drotschmann K., A. B. Clark, H. T. Tran, M. A. Resnick, D. A. Gordenin, *et al.*, 1999 Mutator phenotypes of yeast strains heterozygous for mutations in the *MSH2* gene. Proc. Natl. Acad. Sci. U. S. A. 96: 2970–2975. https://doi.org/10.1073/pnas.96.6.2970

Dujon B., D. Sherman, G. Fischer, P. Durrens, S. Casaregola, *et al.*, 2004 Genome evolution in yeasts. Nature 430: 35–44. https://doi.org/10.1038/nature02579

Dujon B., 2010 Yeast evolutionary genomics. Nat. Rev. Genet. 11: 512–524. https://doi.org/10.1038/nrg2811

Dunkel N., and J. Morschhäuser, 2011 Loss of heterozygosity at an unlinked genomic locus is responsible for the phenotype of a *Candida albicans sap4Δ sap5Δ sap6Δ* mutant. Eukaryot. Cell 10: 54–62. https://doi.org/10.1128/EC.00281-10

Dunn M. J., G. M. Kinney, P. M. Washington, J. Berman, and M. Z. Anderson, 2018 Functional diversification accompanies gene family expansion of *MED2* homologs in *Candida albicans*. PLoS Genet. 14: e1007326. https://doi.org/10.1371/journal.pgen.1007326

Dunn M. J., and M. Z. Anderson, 2019 To Repeat or Not to Repeat: Repetitive Sequences Regulate Genome Stability in *Candida albicans*. Genes (Basel). 10: 866. https://doi.org/10.3390/genes10110866

Elango R., B. Osia, V. Harcy, E. Malc, P. A. Mieczkowski, *et al.*, 2019 Repair of base damage within break-induced replication intermediates promotes kataegis associated with chromosome rearrangements. Nucleic Acids Res. 47: 9666–9684. https://doi.org/10.1093/nar/gkz651

Ene I. V, R. A. Farrer, M. P. Hirakawa, K. Agwamba, C. A. Cuomo, *et al.*, 2018 Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. Proc. Natl. Acad. Sci.

Feng J., S. Yao, Y. Dong, J. Hu, M. Whiteway, *et al.*, 2020 Nucleotide Excision Repair Protein Rad23 Regulates Cell Virulence Independent of Rad4 in *Candida albicans*. mSphere 5. https://doi.org/10.1128/mSphere.00062-20

Feri A., R. Loll-Krippleber, P.-H. Commere, C. Maufrais, N. Sertour, *et al.*, 2016 Analysis of Repair Mechanisms following an Induced Double-Strand Break Uncovers Recessive Deleterious Alleles in the *Candida albicans* Diploid Genome, (J. Berman, Ed.). MBio 7: e01109-16. https://doi.org/10.1128/mBio.01109-16

Fischer G., E. P. C. Rocha, F. Brunet, M. Vergassola, and B. Dujon, 2006 Highly Variable Rates of Genome Rearrangements between Hemiascomycetous Yeast Lineages. PLOS Genet. 2: e32.

Flanagan P. R., J. Fletcher, H. Boyle, R. Sulea, G. P. Moran, *et al.*, 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

Fonzi W. A., and M. Y. Irwin, 1993 Isogenic strain construction and gene mapping in *Candida albicans*. Genetics 134: 717 LP – 728.

Forche A., G. May, J. Beckerman, S. Kauffman, J. Becker, *et al.*, 2003 A system for studying genetic changes in *Candida albicans* during infection. Fungal Genet. Biol. 39: 38–50. https://doi.org/https://doi.org/10.1016/S1087-1845(02)00585-6

Forche A., K. Alby, D. Schaefer, A. D. Johnson, J. Berman, *et al.*, 2008 The parasexual cycle in *Candida albicans* provides an alternative pathway to meiosis for the formation of recombinant strains. PLoS Biol 6. https://doi.org/10.1371/journal.pbio.0060110

Forche A., P. T. Magee, A. Selmecki, J. Berman, and G. May, 2009a 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., M. Steinbach, and J. Berman, 2009b Efficient and rapid identification of *Candida albicans* allelic status using SNP-RFLP. FEMS Yeast Res. 9: 1061–1069. https://doi.org/10.1111/j.1567-1364.2009.00542.x

Forche A., D. Abbey, T. Pisithkul, M. A. Weinzierl, T. Ringstrom, *et al.*, 2011 Stress Alters Rates and Types of Loss of Heterozygosity in *Candida albicans*. MBio 2: e00129-11. https://doi.org/10.1128/mBio.00129-11

Forche A., G. Cromie, A. C. Gerstein, N. V Solis, T. Pisithkul, *et al.*, 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., N. V Solis, M. Swidergall, R. Thomas, A. Guyer, *et al.*, 2019 Selection of *Candida albicans* trisomy during oropharyngeal infection results in a commensal-like phenotype. PLOS Genet. 15: e1008137.

Fortini P., B. Pascucci, E. Parlanti, M. D'Errico, V. Simonelli, *et al.*, 2003 The base excision repair: mechanisms and its relevance for cancer susceptibility. Biochimie 85: 1053–1071. https://doi.org/https://doi.org/10.1016/j.biochi.2003.11.003

Fournier T., and J. Schacherer, 2017 Genetic backgrounds and hidden trait complexity in natural populations. Curr. Opin. Genet. Dev. 47: 48–53. https://doi.org/https://doi.org/10.1016/j.gde.2017.08.009

Freire-Beneitez V., R. J. Price, D. Tarrant, J. Berman, and A. Buscaino, 2016 *Candida albicans* repetitive elements display epigenetic diversity and plasticity. Sci. Rep. 6: 22989. https://doi.org/10.1038/srep22989

Freire-Benéitez V., R. J. Price, and A. Buscaino, 2016 The Chromatin of *Candida albicans* Pericentromeres Bears Features of Both Euchromatin and Heterochromatin. Front. Microbiol. 7: 759.

Fu Y., J. A. Foden, C. Khayter, M. L. Maeder, D. Reyon, *et al.*, 2013 High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31: 822–826. https://doi.org/10.1038/nbt.2623

Garcia V., S. E. L. Phelps, S. Gray, and M. J. Neale, 2011 Bidirectional resection of DNA double-strand breaks by Mre11 and Exo1. Nature 479: 241–244. https://doi.org/10.1038/nature10515

Gentric G., and C. Desdouets, 2014 Polyploidization in Liver Tissue. Am. J. Pathol. 184: 322–331. https://doi.org/https://doi.org/10.1016/j.ajpath.2013.06.035

Gerstein A. C., H.-J. E. Chun, A. Grant, and S. P. Otto, 2006 Genomic Convergence toward Diploidy in *Saccharomyces cerevisiae*. PLOS Genet. 2: e145.

Gerstein A. C., and S. P. Otto, 2009 Ploidy and the Causes of Genomic Evolution. J. Hered. 100: 571–581. https://doi.org/10.1093/jhered/esp057

Gerstein A. C., A. Kuzmin, and S. P. Otto, 2014 Loss-of-heterozygosity facilitates passage through Haldane's sieve for *Saccharomyces cerevisiae* undergoing adaptation. Nat. Commun. 5: 3819.

Gerstein A. C., M. S. Fu, L. Mukaremera, Z. Li, K. L. Ormerod, *et al.*, 2015 Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. MBio 6: e01340-15. https://doi.org/10.1128/mBio.01340-15

Gerstein A. C., H. Lim, J. Berman, and M. A. Hickman, 2017 Ploidy tug-of-war: Evolutionary and genetic environments influence the rate of ploidy drive in a human fungal pathogen. Evolution (N. Y). 71: 1025–1038. https://doi.org/10.1111/evo.13205

Gerstein A. C., and J. Berman, 2020 *Candida albicans* Genetic Background Influences Mean and Heterogeneity of Drug Responses and Genome Stability during Evolution in Fluconazole, (A. P. Mitchell, Ed.). mSphere 5: e00480-20. https://doi.org/10.1128/mSphere.00480-20

Giaever G., A. M. Chu, L. Ni, C. Connelly, L. Riles, *et al.*, 2002 Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418: 387.

Girard P. M., and S. Boiteux, 1997 Repair of oxidized DNA bases in the yeast *Saccharomyces cerevisiae*. Biochimie 79: 559–566. https://doi.org/https://doi.org/10.1016/S0300-9084(97)82004-4

Gobbi M. De, V. Viprakasit, J. R. Hughes, C. Fisher, V. J. Buckle, *et al.*, 2006 A Regulatory SNP Causes a Human Genetic Disease by Creating a New Transcriptional Promoter. Science (80-.). 312: 1215 LP – 1217.

Goellner E. M., C. D. Putnam, W. J. Graham, C. M. Rahal, B.-Z. Li, *et al.*, 2018 Identification of Exo1-Msh2 interaction motifs in DNA mismatch repair and new Msh2-binding partners. Nat. Struct. Mol. Biol. 25: 650–659. https://doi.org/10.1038/s41594-018-0092-y

Goellner E. M., 2020 Chromatin remodeling and mismatch repair: Access and excision. DNA Repair (Amst). 85: 102733. https://doi.org/https://doi.org/10.1016/j.dnarep.2019.102733

Goins C. L., K. J. Gerik, and J. K. Lodge, 2006 Improvements to gene deletion in the fungal pathogen *Cryptococcus neoformans*: absence of Ku proteins increases homologous recombination, and co-transformation of independent DNA molecules allows rapid complementation of deletion phenotypes. Fungal Genet. Biol. 43: 531–544. https://doi.org/10.1016/j.fgb.2006.02.007

Gola S., R. Martin, A. Walther, A. Dünkler, and J. Wendland, 2003 New modules for PCR-based gene targeting in *Candida albicans*: rapid and efficient gene targeting using 100 bp of flanking homology region. Yeast 20: 1339–47. https://doi.org/10.1002/yea.1044

Gómez-Raja J., E. Andaluz, B. Magee, R. Calderone, and G. Larriba, 2008 A single SNP, G929T (Gly310Val), determines the presence of a functional and a non-functional allele of *HIS4* in *Candida albicans* SC5314: Detection of the non-functional allele in laboratory strains. Fungal Genet. Biol. 45: 527–541. https://doi.org/10.1016/j.fgb.2007.08.008

Goodwin T. J. D., and R. T. M. Poulter, 2000 Multiple LTR-Retrotransposon Families in the Asexual Yeast *Candida albicans*. Genome Res. 10: 174–191. https://doi.org/10.1101/gr.10.2.174

Gorman J. A., J. W. Gorman, and Y. Koltin, 1992 Direct selection of galactokinase-negative mutants of *Candida albicans* using 2-deoxy-galactose. Curr. Genet. 21: 203–206. https://doi.org/10.1007/bf00336842

Goudie F., M. H. Allsopp, and B. P. Oldroyd, 2014 Selection on Overdominant Genes Maintains Heterozygosity Along Multiple Chromosomes in a Clonal Lineage of Honey Bee. Evolution (N. Y). 68: 125–136. https://doi.org/10.1111/evo.12231

Gradia S., D. Subramanian, T. Wilson, S. Acharya, A. Makhov, *et al.*, 1999 hMSH2–hMSH6 Forms a Hydrolysis-Independent Sliding Clamp on Mismatched DNA. Mol. Cell 3: 255–261. https://doi.org/https://doi.org/10.1016/S1097-2765(00)80316-0

Grahl N., E. G. Demers, A. W. Crocker, and D. A. Hogan, 2017 Use of RNA-Protein Complexes for Genome Editing in Non-albicans Candida Species. mSphere 2: e00218-17. https://doi.org/10.1128/mSphere.00218-17

Guin K., Y. Chen, R. Mishra, S. R. B. M. Muzaki, B. C. Thimmappa, *et al.*, 2020 Spatial inter-centromeric interactions facilitated the emergence of evolutionary new centromeres. Elife 9. https://doi.org/10.7554/eLife.58556

Gusa A., and S. Jinks-Robertson, 2019 Mitotic Recombination and Adaptive Genomic Changes in Human Pathogenic Fungi. Genes (Basel). 10. https://doi.org/10.3390/genes10110901

Habraken Y., P. Sung, L. Prakash, and S. Prakash, 1993 Yeast excision repair gene *RAD2* encodes a single-stranded DNA endonuclease. Nature 366: 365–368. https://doi.org/10.1038/366365a0

Hanawalt P. C., 1991 Heterogeneity of DNA repair at the gene level. Mutat. Res. Mol. Mech. Mutagen. 247: 203–211. https://doi.org/https://doi.org/10.1016/0027-5107(91)90016-H

Haran J., H. Boyle, K. Hokamp, T. Yeomans, Z. Liu, *et al.*, 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

Harfe B. D., and S. Jinks-Robertson, 2000 Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in *Saccharomyces cerevisiae*. Genetics 156: 571–578.

Harrison B. D., J. Hashemi, M. Bibi, R. Pulver, D. Bavli, *et al.*, 2014 A Tetraploid Intermediate Precedes Aneuploid Formation in Yeasts Exposed to Fluconazole. PLOS Biol. 12: e1001815.

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

He X., and J. Zhang, 2005 Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. Genetics 169: 1157–1164. https://doi.org/10.1534/genetics.104.037051

Healey K. R., Y. Zhao, W. B. Perez, S. R. Lockhart, J. D. Sobel, *et al.*, 2016 Prevalent mutator genotype identified in fungal pathogen *Candida glabrata* promotes multi-drug resistance. Nat. Commun. 7: 11128. https://doi.org/10.1038/ncomms11128

Hickman M. A., G. Zeng, A. Forche, M. P. Hirakawa, D. Abbey, *et al.*, 2013 The 'obligate diploid' *Candida albicans* forms mating-competent haploids. Nature 494: 55–59. https://doi.org/10.1038/nature11865

Hickman M. A., C. Paulson, A. Dudley, and J. Berman, 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

Hilton C., D. Markie, B. Corner, E. Rikkerink, and R. Poulter, 1985 Heat shock induces chromosome loss in the yeast Candida albicans. Mol. Gen. Genet. 200: 162–168. https://doi.org/10.1007/bf00383330

Hirakawa M. P., D. A. Martinez, S. Sakthikumar, M. Z. Anderson, A. Berlin, *et al.*, 2015 Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res. 25: 413–425. https://doi.org/10.1101/gr.174623.114

Ho C. K., G. Mazón, A. F. Lam, and L. S. Symington, 2010 Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. Mol. Cell 40: 988–1000. https://doi.org/10.1016/j.molcel.2010.11.016

Homann O. R., J. Dea, S. M. Noble, and A. D. Johnson, 2009 A Phenotypic Profile of the *Candida albicans* Regulatory Network. PLOS Genet. 5: e1000783.

Hombauer H., C. S. Campbell, C. E. Smith, A. Desai, and R. D. Kolodner, 2011 Visualization of Eukaryotic DNA Mismatch Repair Reveals Distinct Recognition and Repair Intermediates. Cell 147: 1040–1053. https://doi.org/https://doi.org/10.1016/j.cell.2011.10.025

Hommel B., L. Mukaremera, R. J. B. Cordero, C. Coelho, C. A. Desjardins, *et al.*, 2018 Titan cells formation in *Cryptococcus neoformans* is finely tuned by environmental conditions and modulated by positive and negative genetic regulators. PLOS Pathog. 14: e1006982.

Hoog M. van het, T. J. Rast, M. Martchenko, S. Grindle, D. Dignard, *et al.*, 2007 Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol 8. https://doi.org/10.1186/gb-2007-8-4-r52

Hsu M., M. J. McEachern, A. T. Dandjinou, Y. Tzfati, E. Orr, *et al.*, 2007 Telomerase core components protect Candida telomeres from aberrant overhang accumulation. Proc. Natl. Acad. Sci. U. S. A. 104: 11682–11687. https://doi.org/10.1073/pnas.0700327104

Huang J. C., D. L. Svoboda, J. T. Reardon, and A. Sancar, 1992 Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond and the 6th phosphodiester bond; to the photodimer. Proc. Natl. Acad. Sci. 89: 3664 LP – 3668. https://doi.org/10.1073/pnas.89.8.3664

Huang M.-E., A.-G. Rio, M.-D. Galibert, and F. Galibert, 2002 Pol32, a subunit of *Saccharomyces cerevisiae* DNA polymerase delta, suppresses genomic deletions and is involved in the mutagenic bypass pathway. Genetics 160: 1409–1422.

Huang G., T. Srikantha, N. Sahni, S. Yi, and D. R. Soll, 2009 CO_2 Regulates White-to-Opaque Switching in *Candida albicans*. Curr. Biol. 19: 330–334. https://doi.org/https://doi.org/10.1016/j.cub.2009.01.018

Huang G., S. Yi, N. Sahni, K. J. Daniels, T. Srikantha, *et al.*, 2010 N-Acetylglucosamine Induces White to Opaque Switching, a Mating Prerequisite in *Candida albicans*. PLOS Pathog. 6: e1000806.

Huang M. Y., and A. P. Mitchell, 2017 Marker Recycling in *Candida albicans* through CRISPR-Cas9-Induced Marker Excision. mSphere 2. https://doi.org/10.1128/mSphere.00050-17

Huang M. Y., C. A. Woolford, G. May, C. J. McManus, and A. P. Mitchell, 2019 Circuit diversification in a biofilm regulatory network. PLOS Pathog. 15: e1007787.

Hughes T. R., C. J. Roberts, H. Dai, A. R. Jones, M. R. Meyer, *et al.*, 2000 Widespread aneuploidy revealed by DNA microarray expression profiling. Nat. Genet. 25: 333–337. https://doi.org/10.1038/77116

Hull C. M., and A. D. Johnson, 1999 Identification of a Mating Type-Like Locus in the Asexual Pathogenic Yeast *Candida albicans*. Science (80-.). 285: 1271 LP – 1275. https://doi.org/10.1126/science.285.5431.1271

Hull C. M., R. M. Raisner, and A. D. Johnson, 2000 Evidence for Mating of the “Asexual” Yeast *Candida albicans* in a Mammalian Host. Science (80-.). 289: 307 LP – 310. https://doi.org/10.1126/science.289.5477.307

Hunter J. D., 2007 Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 9: 90–95. https://doi.org/10.1109/MCSE.2007.55

Ibrahim A. S., B. B. Magee, D. C. Sheppard, M. Yang, S. Kauffman, *et al.*, 2005 Effects of ploidy and mating type on virulence of *Candida albicans*. Infect. Immun. 73: 7366–7374. https://doi.org/10.1128/IAI.73.11.7366-7374.2005

Ira G., A. Malkova, G. Liberi, M. Foiani, and J. E. Haber, 2003 Srs2 and Sgs1–Top3 Suppress Crossovers during Double-Strand Break Repair in Yeast. Cell 115: 401–411. https://doi.org/https://doi.org/10.1016/S0092-8674(03)00886-9

Iwaguchi S.-I., T. Kanbe, T. Tohne, P. T. Magee, and T. Suzuki, 2000 High-frequency occurrence of chromosome translocation in a mutant strain of *Candida albicans* by a suppressor mutation of ploidy shift. Yeast 16: 411–422. https://doi.org/10.1002/(SICI)1097-0061(20000330)16:5<411::AID-YEA532>3.0.CO;2-N

Iwaguchi S. I., M. Sato, B. B. Magee, P. T. Magee, K. Makimura, *et al.*, 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

Iwaguchi S.-I., M. Suzuki, N. Sakai, Y. Nakagawa, P. T. Magee, *et al.*, 2004 Chromosome translocation induced by the insertion of the URA blaster into the major repeat sequence (MRS) in *Candida albicans*. Yeast 21: 619–634. https://doi.org/10.1002/yea.1116

Jain S., N. Sugawara, J. Lydeard, M. Vaze, N. Tanguy Le Gac, *et al.*, 2009 A recombination execution checkpoint regulates the choice of homologous recombination pathway during DNA double-strand break repair. Genes Dev. 23: 291–303. https://doi.org/10.1101/gad.1751209

Jain S., N. Sugawara, A. Mehta, T. Ryu, and J. E. Haber, 2016 Sgs1 and Mph1 Helicases Enforce the Recombination Execution Checkpoint During DNA Double-Strand Break Repair in *Saccharomyces cerevisiae*. Genetics 203: 667–675. https://doi.org/10.1534/genetics.115.184317

James T. Y., L. A. Michelotti, A. D. Glasco, R. A. Clemons, R. A. Powers, *et al.*, 2019 Adaptation by Loss of Heterozygosity in *Saccharomyces cerevisiae* Clones Under Divergent Selection. Genetics genetics.302411.2019. https://doi.org/10.1534/genetics.119.302411

Janbon G., F. Sherman, and E. Rustchenko, 1998 Monosomy of a specific chromosome determines sorbose utilization: A novel regulatory mechanism in *Candida albicans*. Proc. Natl. Acad. Sci. 95: 5150 LP – 5155. https://doi.org/10.1073/pnas.95.9.5150

Jinek M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, *et al.*, 2012 A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337: 816–821. https://doi.org/10.1126/science.1225829

Jones T., N. A. Federspiel, H. Chibana, J. Dungan, S. Kalman, *et al.*, 2004 The diploid genome sequence of *Candida albicans*. Proc Natl Acad Sci U S A 101. https://doi.org/10.1073/pnas.0401648101

Kadyrova L. Y., E. R. Blanko, and F. A. Kadyrov, 2011 CAF-I-dependent control of degradation of the discontinuous strands during mismatch repair. Proc. Natl. Acad. Sci. 108: 2753 LP – 2758. https://doi.org/10.1073/pnas.1015914108

Kantidze O. L., A. K. Velichko, A. V Luzhin, and S. V Razin, 2016 Heat Stress-Induced DNA Damage. Acta Naturae 8: 75–78.

Ketel C., H. S. W. Wang, M. McClellan, K. Bouchonville, A. Selmecki, *et al.*, 2009 Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. PLoS Genet. 5: e1000400. https://doi.org/10.1371/journal.pgen.1000400

Kirsch D. R., and R. R. Whitney, 1991 Pathogenicity of *Candida albicans* auxotrophic mutants in experimental infections. Infect. Immun. 59: 3297–3300.

Ko H.-C., T.-Y. Hsiao, C.-T. Chen, and Y.-L. Yang, 2013 *Candida albicans ENO1* null mutants exhibit altered drug susceptibility, hyphal formation, and virulence. J. Microbiol. 51: 345–351. https://doi.org/10.1007/s12275-013-2577-z

Koc A., L. J. Wheeler, C. K. Mathews, and G. F. Merrill, 2004 Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. J. Biol. Chem. 279: 223–230. https://doi.org/10.1074/jbc.M303952200

Koren A., R. E. Handsaker, N. Kamitaki, R. Karlić, S. Ghosh, *et al.*, 2014 Genetic variation in human DNA replication timing. Cell 159: 1015–1026. https://doi.org/10.1016/j.cell.2014.10.025

Krappmann S., C. Sasse, and G. H. Braus, 2006 Gene targeting in *Aspergillus fumigatus* by homologous recombination is facilitated in a nonhomologous end-joining-deficient genetic background. Eukaryot. Cell 5: 212–215. https://doi.org/10.1128/EC.5.1.212-215.2006

Legrand M., C. L. Chan, P. A. Jauert, and D. T. Kirkpatrick, 2007 Role of DNA mismatch repair and double-strand break repair in genome stability and antifungal drug resistance in *Candida albicans*. Eukaryot. Cell 6: 2194–2205. https://doi.org/10.1128/EC.00299-07

Legrand M., C. L. Chan, P. A. Jauert, and D. T. Kirkpatrick, 2008 Analysis of base excision and nucleotide excision repair in *Candida albicans.* Microbiology 154: 2446–2456. https://doi.org/https://doi.org/10.1099/mic.0.2008/017616-0

Legrand M., S. Bachellier-Bassi, K. K. Lee, Y. Chaudhari, H. Tournu, *et al.*, 2018 Generating genomic platforms to study *Candida albicans* pathogenesis. Nucleic Acids Res. gky594–gky594.

Legrand M., P. Jaitly, A. Feri, C. D'Enfert, and K. Sanyal, 2019 *Candida albicans*: An Emerging Yeast Model to Study Eukaryotic Genome Plasticity. Trends Genet. 35: 292–307. https://doi.org/https://doi.org/10.1016/j.tig.2019.01.005

Legras J.-L., V. Galeote, F. Bigey, C. Camarasa, S. Marsit, *et al.*, 2018 Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. Mol. Biol. Evol. 35: 1712–1727. https://doi.org/10.1093/molbev/msy066

Lemos B. R., A. C. Kaplan, J. E. Bae, A. E. Ferrazzoli, J. Kuo, *et al.*, 2018 CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. Proc. Natl. Acad. Sci. 115: E2040 LP-E2047. https://doi.org/10.1073/pnas.1716855115

Lephart P. R., H. Chibana, and P. T. Magee, 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

Lephart P. R., and P. T. Magee, 2006 Effect of the major repeat sequence on mitotic recombination in *Candida albicans*. Genetics 174: 1737–1744. https://doi.org/10.1534/genetics.106.063271

Li X., Z. Yan, and J. Xu, 2003 Quantitative variation of biofilms among strains in natural populations of *Candida albicans*. Microbiology 149: 353–362. https://doi.org/10.1099/mic.0.25932-0

Li F., L. Tian, L. Gu, and G.-M. Li, 2009 Evidence that nucleosomes inhibit mismatch repair in eukaryotic cells. J. Biol. Chem. 284: 33056–33061. https://doi.org/10.1074/jbc.M109.049874

Li S., 2015 Transcription coupled nucleotide excision repair in the yeast *Saccharomyces cerevisiae*: The ambiguous role of Rad26. DNA Repair (Amst). 36: 43–48. https://doi.org/10.1016/j.dnarep.2015.09.006

Li H., 2018 Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Liang S., M. Hitomi, Y. H. Hu, Y. Liu, and A. M. Tartakoff, 1996 A DEAD-box-family protein is required for nucleocytoplasmic transport of yeast mRNA. Mol. Cell. Biol. 16: 5139–5146.

Liang S.-H., M. Z. Anderson, M. P. Hirakawa, J. M. Wang, C. Frazer, *et al.*, 2019 Hemizygosity Enables a Mutational Transition Governing Fungal Virulence and Commensalism. Cell Host Microbe 25: 418-431.e6. https://doi.org/https://doi.org/10.1016/j.chom.2019.01.005

Liang S.-H., and J. R. Bennett, 2019 The Impact of Gene Dosage and Heterozygosity on the Diploid Pathobiont *Candida albicans*. J. Fungi 6.

Lin C., L. Yang, and M. G. Rosenfeld, 2012 Molecular Logic Underlying Chromosomal Translocations, Random or Non-Random?, pp. 241–279 in edited by Tew K. D., Fisher P. B. B. T.-A. in C. R. Academic Press.

Liu D., G. Keijzers, and L. J. Rasmussen, 2017 DNA mismatch repair and its many roles in eukaryotic cells. Mutat. Res. 773: 174–187. https://doi.org/10.1016/j.mrrev.2017.07.001

Liu J., J.-M. François, and J.-P. Capp, 2019 Gene Expression Noise Produces Cell-to-Cell Heterogeneity in Eukaryotic Homologous Recombination Rate. Front. Genet. 10: 475. https://doi.org/10.3389/fgene.2019.00475

Lockhart S. R., C. Pujol, K. J. Daniels, M. G. Miller, A. D. Johnson, *et al.*, 2002 In *Candida albicans*, white-opaque switchers are homozygous for mating type. Genetics 162: 737–745.

Lockhart S. R., K. J. Daniels, R. Zhao, D. Wessels, and D. R. Soll, 2003 Cell biology of mating in *Candida albicans*. Eukaryot. Cell 2: 49–61. https://doi.org/10.1128/ec.2.1.49-61.2003

Loll-Krippleber R., C. d'Enfert, A. Feri, D. Diogo, A. Perin, *et al.*, 2014 A study of the DNA damage checkpoint in *Candida albicans*: uncoupling of the functions of Rad53 in DNA repair, cell cycle regulation and genotoxic stress-induced polarized growth. Mol. Microbiol. 91: 452–471. https://doi.org/10.1111/mmi.12471

Loll-Krippleber R., A. Feri, M. Nguyen, C. Maufrais, J. Yansouni, *et al.*, 2015 A FACS-Optimized Screen Identifies Regulators of Genome Stability in *Candida albicans*. Eukaryot. Cell 14: 311–322. https://doi.org/10.1128/EC.00286-14

Lundblad V., and E. H. Blackburn, 1993 An alternative pathway for yeast telomere maintenance rescues est1− senescence. Cell 73: 347–360. https://doi.org/https://doi.org/10.1016/0092-8674(93)90234-H

Lydeard J. R., S. Jain, M. Yamaguchi, and J. E. Haber, 2007 Break-induced replication and telomerase-independent telomere maintenance require Pol32. Nature 448: 820–823. https://doi.org/10.1038/nature06047

Ma J.-L., E. M. Kim, J. E. Haber, and S. E. Lee, 2003 Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. Mol. Cell. Biol. 23: 8820–8828. https://doi.org/10.1128/mcb.23.23.8820-8828.2003

Magee B. B., and P. T. Magee, 2000 Induction of Mating in *Candida albicans* by Construction of MTLa and MTLα Strains. Science (80-.). 289: 310 LP – 313. https://doi.org/10.1126/science.289.5477.310

Makarova A. V, and P. M. Burgers, 2015 Eukaryotic DNA polymerase ζ. DNA Repair (Amst). 29: 47–55. https://doi.org/https://doi.org/10.1016/j.dnarep.2015.02.012

Mali P., K. M. Esvelt, and G. M. Church, 2013 Cas9 as a versatile tool for engineering biology. Nat. Methods 10: 957–963. https://doi.org/10.1038/nmeth.2649

Malkova A., E. L. Ivanov, and J. E. Haber, 1996 Double-strand break repair in the absence of *RAD51* in yeast: a possible role for break-induced DNA replication. Proc. Natl. Acad. Sci. U. S. A. 93: 7131–7136. https://doi.org/10.1073/pnas.93.14.7131

Malkova A., F. Klein, W.-Y. Leung, and J. E. Haber, 2000 HO endonuclease-induced recombination in yeast meiosis resembles Spo11-induced events. Proc. Natl. Acad. Sci. U. S. A. 97: 14500–14505.

Malkova A., L. Signon, C. B. Schaefer, M. L. Naylor, J. F. Theis, *et al.*, 2001 *RAD51*-independent break-induced replication to repair a broken chromosome depends on a distant enhancer site. Genes Dev. 15: 1055–1060. https://doi.org/10.1101/gad.875901

Malkova A., M. L. Naylor, M. Yamaguchi, G. Ira, and J. E. Haber, 2005 *RAD51*-dependent break-induced replication differs in kinetics and checkpoint responses from *RAD51*-mediated gene conversion. Mol. Cell. Biol. 25: 933–944. https://doi.org/10.1128/MCB.25.3.933-944.2005

Mao P., and J. J. Wyrick, 2019 Organization of DNA damage, excision repair, and mutagenesis in chromatin: A genomic perspective. DNA Repair (Amst). 81: 102645. https://doi.org/https://doi.org/10.1016/j.dnarep.2019.102645

Marton T., A. Feri, P.-H. Commere, C. Maufrais, C. d'Enfert, *et al.*, 2019 Identification of Recessive Lethal Alleles in the Diploid Genome of a *Candida albicans* Laboratory Strain Unveils a Potential Role of Repetitive Sequences in Buffering Their Deleterious Impact. mSphere 4: e00709-18. https://doi.org/10.1128/mSphere.00709-18

Marton T., C. Maufrais, C. d'Enfert, and M. Legrand, 2020 Use of CRISPR-Cas9 To Target Homologous Recombination Limits Transformation-Induced Genomic Changes in *Candida albicans*, (A. P. Mitchell, Ed.). mSphere 5: e00620-20. https://doi.org/10.1128/mSphere.00620-20

Maslov A. Y., and J. Vijg, 2009 Genome instability, cancer and aging. Biochim. Biophys. Acta 1790: 963–969. https://doi.org/10.1016/j.bbagen.2009.03.020

McEachern M. J., and J. B. Hicks, 1993 Unusually large telomeric repeats in the yeast *Candida albicans*. Mol. Cell. Biol. 13: 551–560. https://doi.org/10.1128/mcb.13.1.551

McPherson J. P., B. Lemmers, R. Chahwan, A. Pamidi, E. Migon, *et al.*, 2004 Involvement of mammalian Mus81 in genome integrity and tumor suppression. Science 304: 1822–1826. https://doi.org/10.1126/science.1094557

McVey M., and S. E. Lee, 2008 MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet. 24: 529–538. https://doi.org/10.1016/j.tig.2008.08.007

Mehta A., A. Beach, and J. E. Haber, 2017 Homology Requirements and Competition between Gene Conversion and Break-Induced Replication during Double-Strand Break Repair. Mol. Cell 65: 515-526.e3. https://doi.org/10.1016/j.molcel.2016.12.003

Mendillo M. L., D. J. Mazur, and R. D. Kolodner, 2005 Analysis of the interaction between the *Saccharomyces cerevisiae MSH2-MSH6* and *MLH1-PMS1* complexes with DNA using a reversible DNA end-blocking system. J. Biol. Chem. 280: 22245–22257. https://doi.org/10.1074/jbc.M407545200

Miller M. G., and A. D. Johnson, 2002 White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell 110: 293–302. https://doi.org/10.1016/s0092-8674(02)00837-1

Mimitou E. P., and L. S. Symington, 2008 Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. Nature 455: 770–774. https://doi.org/10.1038/nature07312

Min K., Y. Ichikawa, C. A. Woolford, and A. P. Mitchell, 2016 *Candida albicans* Gene Deletion with a Transient CRISPR-Cas9 System, (M. J. Imperiale, Ed.). mSphere 1: e00130-16. https://doi.org/10.1128/mSphere.00130-16

Mishra P. K., M. Baum, and J. Carbon, 2007 Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. Mol. Genet. Genomics 278: 455–465. https://doi.org/10.1007/s00438-007-0263-8

Mitra S., J. Gómez-Raja, G. Larriba, D. D. Dubey, and K. Sanyal, 2014 Rad51–Rad52 Mediated Maintenance of Centromeric Chromatin in *Candida albicans*. PLoS Genet. 10: e1004344–e1004344. https://doi.org/10.1371/journal.pgen.1004344

Mohr G., M. Del Campo, S. Mohr, Q. Yang, H. Jia, *et al.*, 2008 Function of the C-terminal Domain of the DEAD-Box Protein Mss116p Analyzed In Vivo and In Vitro. J. Mol. Biol. 375: 1344–1364. https://doi.org/10.1016/j.jmb.2007.11.041

Montague T. G., J. M. Cruz, J. A. Gagnon, G. M. Church, and E. Valen, 2014 CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. 42: W401–W407. https://doi.org/10.1093/nar/gku410

Morrical S. W., 2015 DNA-Pairing and Annealing Processes in Homologous Recombination and Homology-Directed Repair. Cold Spring Harb. Perspect. Biol. 7. https://doi.org/10.1101/cshperspect.a016444

Morschhäuser J., P. Staib, and G. Köhler, 2005 Targeted Gene Deletion in *Candida albicans* Wild-Type Strains by MPAR Flipping BT - Antifungal Agents: Methods and Protocols, pp. 35–44 in edited by Ernst E. J., Rogers P. D. Humana Press, Totowa, NJ.

Mosbach V., L. Poggi, D. Viterbo, M. Charpentier, and G.-F. Richard, 2018 TALEN-Induced Double-Strand Break Repair of CTG Trinucleotide Repeats. Cell Rep. 22: 2146–2159. https://doi.org/10.1016/j.celrep.2018.01.083

Mu H., N. E. Geacintov, S. Broyde, J.-E. Yeo, and O. D. Schärer, 2018 Molecular basis for damage recognition and verification by *XPC-RAD23B* and TFIIH in nucleotide excision repair. DNA Repair (Amst). 71: 33–42. https://doi.org/https://doi.org/10.1016/j.dnarep.2018.08.005

Muzzey D., K. Schwartz, J. S. Weissman, and G. Sherlock, 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., G. Sherlock, and J. S. Weissman, 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

Myung K., and R. D. Kolodner, 2003 Induction of genome instability by DNA damage in *Saccharomyces cerevisiae*. DNA Repair (Amst). 2: 243–258. https://doi.org/10.1016/s1568-7864(02)00216-1

Nackley A. G., S. A. Shabalina, I. E. Tchivileva, K. Satterfield, O. Korchynskyi, *et al.*, 2006 Human Catechol-*O*-Methyltransferase Haplotypes Modulate Protein Expression by Altering mRNA Secondary Structure. Science (80-.). 314: 1930 LP – 1933.

Navarro-Garcia F., R. M. Perez-Diaz, B. B. Magee, J. Pla, C. Nombela, *et al.*, 1995 Chromosome reorganization in *Candida albicans* 1001 strain. J. Med. Vet. Mycol. 33: 361–366.

Nguyen N., M. M. F. Quail, and A. D. Hernday, 2017 An Efficient, Rapid, and Recyclable System for CRISPR-Mediated Genome Editing in *Candida albicans*. mSphere 2: e00149-17. https://doi.org/10.1128/mSphereDirect.00149-17

Noble S. M., and A. D. Johnson, 2005 Strains and Strategies for Large-Scale Gene Deletion Studies of the Diploid Human Fungal Pathogen *Candida albicans*. Eukaryot. Cell 4: 298–309. https://doi.org/10.1128/EC.4.2.298-309.2005

Nugent C. I., and V. Lundblad, 1998 The telomerase reverse transcriptase: components and regulation. Genes Dev. 12: 1073–1085.

Odds F. C., M.-E. Bougnoux, D. J. Shaw, J. M. Bain, A. D. Davidson, *et al.*, 2007 Molecular Phylogenetics of *Candida albicans*. Eukaryot. Cell 6: 1041 LP – 1052. https://doi.org/10.1128/EC.00041-07

Otto S. P., and J. Whitton, 2000 Polyploid Incidence and Evolution. Annu. Rev. Genet. 34: 401–437. https://doi.org/10.1146/annurev.genet.34.1.401

Otto S. P., 2007 The evolutionary consequences of polyploidy. Cell 131: 452–462. https://doi.org/10.1016/j.cell.2007.10.022

Pamidi A., R. Cardoso, A. Hakem, E. Matysiak-Zablocki, A. Poonepalli, *et al.*, 2007 Functional interplay of p53 and Mus81 in DNA damage responses and cancer. Cancer Res. 67: 8527–8535. https://doi.org/10.1158/0008-5472.CAN-07-1161

Pande K., C. Chen, and S. M. Noble, 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

Papon N., V. Courdavault, M. Clastre, and R. J. Bennett, 2013 Emerging and Emerged Pathogenic Candida Species: Beyond the *Candida albicans* Paradigm. PLOS Pathog. 9: e1003550.

Papon N., M.-E. Bougnoux, and C. D'Enfert, 2020 Tracing the Origin of Invasive Fungal Infections. Trends Microbiol. https://doi.org/https://doi.org/10.1016/j.tim.2020.01.007

Pappas P. G., M. S. Lionakis, M. C. Arendrup, L. Ostrosky-Zeichner, and B. J. Kullberg, 2018 Invasive candidiasis. Nat. Rev. Dis. Prim. 4: 18026. https://doi.org/10.1038/nrdp.2018.26

Pardo B., M. Moriel-Carretero, T. Vicat, A. Aguilera, and P. Pasero, 2020 Homologous recombination and Mus81 promote replication completion in response to replication fork blockage. EMBO Rep. n/a: e49367. https://doi.org/10.15252/embr.201949367

Parfrey L. W., D. J. G. Lahr, and L. A. Katz, 2008 The dynamic nature of eukaryotic genomes. Mol. Biol. Evol. 25: 787–794. https://doi.org/10.1093/molbev/msn032

Park Y.-N., and J. Morschhäuser, 2005 Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. Eukaryot. Cell 4: 1328–1342. https://doi.org/10.1128/EC.4.8.1328-1342.2005

Pedersen B. S., and S. De, 2013 Loss of heterozygosity preferentially occurs in early replicating regions in cancer genomes. Nucleic Acids Res. 41: 7615–7624. https://doi.org/10.1093/nar/gkt552

Peltomäki P., 2003 Role of DNA mismatch repair defects in the pathogenesis of human cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 21: 1174–1179. https://doi.org/10.1200/JCO.2003.04.060

Peter J., M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, *et al.*, 2018 Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature 556: 339–344. https://doi.org/10.1038/s41586-018-0030-5

Petukhova G., S. Van Komen, S. Vergano, H. Klein, and P. Sung, 1999 Yeast Rad54 Promotes Rad51-dependent Homologous DNA Pairing via ATP Hydrolysis-driven Change in DNA Double Helix Conformation. J. Biol. Chem. 274: 29453–29462.

Pfaller M. A., and D. J. Diekema, 2010 Epidemiology of Invasive Mycoses in North America. Crit. Rev. Microbiol. 36: 1–53. https://doi.org/10.3109/10408410903241444

Pfander B., and J. Matos, 2017 Control of Mus81 nuclease during the cell cycle. FEBS Lett. 591: 2048–2056. https://doi.org/10.1002/1873-3468.12727

Piazza A., S. S. Shah, W. D. Wright, S. K. Gore, R. Koszul, *et al.*, 2019 Dynamic Processing of Displacement Loops during Recombinational DNA Repair. Mol. Cell 73: 1255-1266.e4. https://doi.org/10.1016/j.molcel.2019.01.005

Pla J., R. M. Pérez-Diaz, F. Navarro-Garcia, M. Sánchez, and C. Nombela, 1995 Cloning of the *Candida albicans HIS1* gene by direct complementation of a *C. albicans* histidine auxotroph using an improved double-ARS shuttle vector. Gene 165: 115–120. https://doi.org/https://doi.org/10.1016/0378-1119(95)00492-O

Plessis A., A. Perrin, J. E. Haber, and B. Dujon, 1992 Site-specific recombination determined by I-*Sce*I, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. Genetics 130: 451–460.

Prado F., and A. Aguilera, 1995 Role of reciprocal exchange, one-ended invasion crossover and single-strand annealing on inverted and direct repeat recombination in yeast: different requirements for the *RAD1, RAD10*, and *RAD52* genes. Genetics 139: 109–123.

Prakash R., D. Satory, E. Dray, A. Papusha, J. Scheller, *et al.*, 2009 Yeast Mph1 helicase dissociates Rad51-made D-loops: implications for crossover control in mitotic recombination. Genes Dev. 23: 67–79.

Prieto D., I. Correia, J. Pla, and E. Román, 2016 Adaptation of *Candida albicans* to commensalism in the gut. Future Microbiol. 11: 567–583. https://doi.org/10.2217/fmb.16.1

Pujol C., S. Joly, B. Nolan, T. Srikantha, and D. R. Soll, 1999 Microevolutionary changes in *Candida albicans i*dentified by the complex Ca3 fingerprinting probe involve insertions and deletions of the full-length repetitive sequence RPS at specific genomic sites. Microbiology 145: 2635–2646.

Pujol C., M. A. Pfaller, and D. R. Soll, 2004 Flucytosine resistance is restricted to a single genetic clade of *Candida albicans*. Antimicrob. Agents Chemother. 48: 262–266. https://doi.org/10.1128/aac.48.1.262-266.2004

Ramakrishna S., A.-B. Kwaku Dad, J. Beloor, R. Gopalappa, S.-K. Lee, *et al.*, 2014 Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res. 24: 1020–1027. https://doi.org/10.1101/gr.171264.113

Ramakrishnan S., Z. Kockler, R. Evans, B. D. Downing, and A. Malkova, 2018 Single-strand annealing between inverted DNA repeats: Pathway choice, participating proteins, and genome destabilizing consequences. PLOS Genet. 14: e1007543.

Reuss O., A. Vik, R. Kolter, and J. Morschhäuser, 2004 The *SAT1* flipper, an optimized tool for gene disruption in *Candida albicans*. Gene 341: 119–127. https://doi.org/10.1016/j.gene.2004.06.021

Rodgers K., and M. McVey, 2016 Error-Prone Repair of DNA Double-Strand Breaks. J. Cell. Physiol. 231: 15–24. https://doi.org/10.1002/jcp.25053

Ropars J., C. Maufrais, D. Diogo, M. Marcet-Houben, A. Perin, *et al.*, 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

Rosenberg A., I. V Ene, M. Bibi, S. Zakin, E. S. Segal, *et al.*, 2018 Antifungal tolerance is a subpopulation effect distinct from resistance and is associated with persistent candidemia. Nat. Commun. 9: 2470. https://doi.org/10.1038/s41467-018-04926-x

Roy B., and K. Sanyal, 2011 Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. Eukaryot. Cell 10: 1384–1395. https://doi.org/10.1128/EC.05165-11

Rustchenko E. P., T. M. Curran, and F. Sherman, 1993 Variations in the number of ribosomal DNA units in morphological mutants and normal strains of *Candida albicans* and in normal strains of Saccharomyces cerevisiae. J. Bacteriol. 175: 7189 LP – 7199. https://doi.org/10.1128/jb.175.22.7189-7199.1993

Saini N., S. Ramakrishnan, R. Elango, S. Ayyar, Y. Zhang, *et al.*, 2013 Migrating bubble during break-induced replication drives conservative DNA synthesis. Nature 502: 389–392. https://doi.org/10.1038/nature12584

Sakofsky C. J., S. Ayyar, and A. Malkova, 2012 Break-induced replication and genome stability. Biomolecules 2: 483–504. https://doi.org/10.3390/biom2040483

Sakofsky C. J., S. A. Roberts, E. Malc, P. A. Mieczkowski, M. A. Resnick, *et al.*, 2014 Break-induced replication is a source of mutation clusters underlying kataegis. Cell Rep. 7: 1640–1648. https://doi.org/10.1016/j.celrep.2014.04.053

Sakofsky C. J., and A. Malkova, 2017 Break induced replication in eukaryotes: mechanisms, functions, and consequences. Crit. Rev. Biochem. Mol. Biol. 52: 395–413. https://doi.org/10.1080/10409238.2017.1314444

Sakofsky C. J., N. Saini, L. J. Klimczak, K. Chan, E. P. Malc, *et al.*, 2019 Repair of multiple simultaneous double-strand breaks causes bursts of genome-wide clustered hypermutation. PLoS Biol. 17: e3000464. https://doi.org/10.1371/journal.pbio.3000464

San Filippo J., P. Sung, and H. Klein, 2008 Mechanism of Eukaryotic Homologous Recombination. Annu. Rev. Biochem. 77: 229–257. https://doi.org/10.1146/annurev.biochem.77.061306.125255

Sanglard D., B. Hube, M. Monod, F. C. Odds, and N. A. R. Gow, 1997 A triple deletion of the secreted aspartyl proteinase genes *SAP4, SAP5,* and *SAP6* of *Candida albicans* causes attenuated virulence. Infect. Immun. 65: 3539–3546. https://doi.org/10.1128/iai.65.9.3539-3546.1997

Sanyal K., and J. Carbon, 2002 The CENP-A homolog CaCse4p in the pathogenic yeast *Candida albicans* is a centromere protein essential for chromosome transmission. Proc. Natl. Acad. Sci. 99: 12969 LP – 12974. https://doi.org/10.1073/pnas.162488299

Sanyal K., M. Baum, and J. Carbon, 2004 Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. Proc. Natl. Acad. Sci. U. S. A. 101: 11374 LP – 11379. https://doi.org/10.1073/pnas.0404318101

Schaub Y., A. Dünkler, A. Walther, and J. Wendland, 2006 New pFA-cassettes for PCR-based gene manipulation in *Candida albicans*. J. Basic Microbiol. 46: 416–429. https://doi.org/10.1002/jobm.200510133

Schoeters F., C. A. Munro, C. D'Enfert, and P. Van Dijck, 2018 A High-Throughput *Candida albicans* Two-Hybrid System, (A. P. Mitchell, Ed.). mSphere 3: e00391-18. https://doi.org/10.1128/mSphere.00391-18

Schoustra S. E., A. J. M. Debets, M. Slakhorst, and R. F. Hoekstra, 2007 Mitotic Recombination Accelerates Adaptation in the Fungus *Aspergillus nidulans*. PLOS Genet. 3: e68.

Segal E. S., V. Gritsenko, A. Levitan, B. Yadav, N. Dror, *et al.*, 2018 Gene Essentiality Analyzed by *In Vivo* Transposon Mutagenesis and Machine Learning in a Stable Haploid Isolate of *Candida albicans*, (A. S. Di Pietro Nathan Hube, Bernhard, Ed.). MBio 9: e02048-18. https://doi.org/10.1128/mBio.02048-18

Selmecki A., S. Bergmann, and J. Berman, 2005 Comparative genome hybridization reveals widespread aneuploidy in *Candida albicans* laboratory strains. Mol. Microbiol. 55: 1553–1565. https://doi.org/10.1111/j.1365-2958.2005.04492.x

Selmecki A., A. Forche, and J. Berman, 2006a Aneuploidy and Isochromosome Formation in Drug-Resistant *Candida albicans*. Science 313: 367–370. https://doi.org/10.1126/science.1128242

Selmecki A., A. Forche, and J. Berman, 2006b Aneuploidy and Isochromosome Formation in Drug-Resistant *Candida albicans*. Science (80-.). 313: 367 LP – 370. https://doi.org/10.1126/science.1128242

Selmecki A., M. Gerami-Nejad, C. Paulson, A. Forche, and J. Berman, 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., K. Dulmage, L. E. Cowen, J. B. Anderson, and J. Berman, 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

Selmecki A. M., Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shoresh, *et al.*, 2015 Polyploidy can drive rapid adaptation in yeast. Nature 519: 349–352. https://doi.org/10.1038/nature14187

Seoighe C., and K. H. Wolfe, 1999 Yeast genome evolution in the post-genome era. Curr. Opin. Microbiol. 2: 548–554. https://doi.org/10.1016/s1369-5274(99)00015-6

Shampay J., J. W. Szostak, and E. H. Blackburn, 1984 DNA sequences of telomeres maintained in yeast. Nature 310: 154–157. https://doi.org/10.1038/310154a0

Shcherbakova P. V, and T. A. Kunkel, 1999 Mutator phenotypes conferred by *MLH1* overexpression and by heterozygosity for *mlh1* mutations. Mol. Cell. Biol. 19: 3177–3183. https://doi.org/10.1128/mcb.19.4.3177

Shell S. S., C. D. Putnam, and R. D. Kolodner, 2007 The N Terminus of *Saccharomyces cerevisiae* Msh6 Is an Unstructured Tether to PCNA. Mol. Cell 26: 565–578. https://doi.org/https://doi.org/10.1016/j.molcel.2007.04.024

Shi Q.-M., Y.-M. Wang, X.-D. Zheng, R. T. H. Lee, and Y. Wang, 2007 Critical role of DNA checkpoints in mediating genotoxic-stress-induced filamentous growth in *Candida albicans*. Mol. Biol. Cell 18: 815–826. https://doi.org/10.1091/mbc.e06-05-0442

Shin J. H., M.-E. Bougnoux, C. D'Enfert, S. H. Kim, C.-J. Moon, *et al.*, 2011 Genetic diversity among Korean *Candida albicans* bloodstream isolates: assessment by multilocus sequence typing and restriction endonuclease analysis of genomic DNA by use of *BssHII*. J. Clin. Microbiol. 49: 2572–2577. https://doi.org/10.1128/JCM.02153-10

Sirbu B. M., and D. Cortez, 2013 DNA damage response: three levels of DNA repair regulation. Cold Spring Harb. Perspect. Biol. 5: a012724–a012724. https://doi.org/10.1101/cshperspect.a012724

Sitterle E., C. Maufrais, N. Sertour, M. Palayret, C. d'Enfert, *et al.*, 2019 Within-Host Genomic Diversity of *Candida albicans* in Healthy Carriers. Sci. Rep. 9: 2563. https://doi.org/10.1038/s41598-019-38768-4

Sitterlé E., A. T. Coste, T. Obadia, C. Maufrais, M. Chauvel, *et al.*, 2020 Large-scale genome mining allows identification of neutral polymorphisms and novel resistance mutations in genes involved in *Candida albicans* resistance to azoles and echinocandins. J. Antimicrob. Chemother. 75: 835–848. https://doi.org/10.1093/jac/dkz537

Skrzypek M. S., J. Binkley, G. Binkley, S. R. Miyasato, M. Simison, *et al.*, 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

Sloan R., S. N. Huang, Y. Pommier, and S. Jinks-Robertson, 2017 Effects of camptothecin or *TOP1* overexpression on genetic stability in *Saccharomyces cerevisiae*. DNA Repair (Amst). 59: 69–75. https://doi.org/https://doi.org/10.1016/j.dnarep.2017.09.004

Slutsky B., J. Buffo, and D. R. Soll, 1985 High-frequency switching of colony morphology in *Candida albicans*. Science 230: 666–669. https://doi.org/10.1126/science.3901258

Smerdon M. J., and M. W. Lieberman, 1978 Nucleosome rearrangement in human chromatin during UV-induced DNA- reapir synthesis. Proc. Natl. Acad. Sci. 75: 4238 LP – 4241. https://doi.org/10.1073/pnas.75.9.4238

Smith G. R., M. N. Boddy, P. Shanahan, and P. Russell, 2003 Fission yeast Mus81.Eme1 Holliday junction resolvase is required for meiotic crossing over but not for gene conversion. Genetics 165: 2289–2293.

Smith C. E., B. Llorente, and L. S. Symington, 2007 Template switching during break-induced replication. Nature 447: 102–105. https://doi.org/10.1038/nature05723

Smith S. B., D. L. Kiss, E. Turk, A. M. Tartakoff, and E. D. Andrulis, 2011 Pronounced and Extensive Microtubule Defects in a *Saccharomyces cerevisiae DIS3* Mutant. Yeast 28: 755–769. https://doi.org/10.1002/yea.1899

Smukowski Heil C. S., C. G. DeSevo, D. A. Pai, C. M. Tucker, M. L. Hoang, *et al.*, 2017 Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. Mol. Biol. Evol. 34: 1596–1612. https://doi.org/10.1093/molbev/msx098

Stafa A., R. A. Donnianni, L. A. Timashev, A. F. Lam, and L. S. Symington, 2014 Template switching during break-induced replication is promoted by the Mph1 helicase in *Saccharomyces cerevisiae*. Genetics 196: 1017–1028. https://doi.org/10.1534/genetics.114.162297

Staib P., M. Kretschmar, T. Nichterlein, H. Hof, and J. Morschhauser, 2002 Host versus *in vitro* signals and intrastrain allelic differences in the expression of a *Candida albicans* virulence gene. Mol. Microbiol. 44: 1351–1366. https://doi.org/10.1046/j.1365-2958.2002.02967.x

Staib P., and J. Morschhäuser, 2007 Chlamydospore formation in *Candida albicans* and *Candida dubliniensis*– an enigmatic developmental programme. Mycoses 50: 1–12. https://doi.org/10.1111/j.1439-0507.2006.01308.x

Sugawara N., and J. E. Haber, 2012 Monitoring DNA recombination initiated by HO endonuclease. Methods Mol. Biol. 920: 349–370. https://doi.org/10.1007/978-1-61779-998-3_25

Sun S., and J. Heitman, 2011 Is sex necessary? BMC Biol. 9: 56. https://doi.org/10.1186/1741-7007-9-56

Sun Y., C. Cao, W. Jia, L. Tao, G. Guan, *et al.*, 2015 pH Regulates White-Opaque Switching and Sexual Mating in *Candida albicans.* Eukaryot. Cell 14: 1127 LP – 1134. https://doi.org/10.1128/EC.00123-15

Swanson R. L., N. J. Morey, P. W. Doetsch, and S. Jinks-Robertson, 1999 Overlapping Specificities of Base Excision Repair, Nucleotide Excision Repair, Recombination, and Translesion Synthesis Pathways for DNA Base Damage in *Saccharomyces cerevisiae* Mol. Cell. Biol. 19: 2929 LP – 2935. https://doi.org/10.1128/MCB.19.4.2929

Symington L. S., 2014 End Resection at Double-Strand Breaks: Mechanism and Regulation. Cold Spring Harb. Perspect. Biol. 6.

Tay Y. D., J. M. Sidebotham, and L. Wu, 2010 Mph1 requires mismatch repair-independent and -dependent functions of MutSalpha to regulate crossover formation during homologous recombination repair. Nucleic Acids Res. 38: 1889–1901. https://doi.org/10.1093/nar/gkp1199

Thakur J., and K. Sanyal, 2012 A coordinated interdependent protein circuitry stabilizes the kinetochore ensemble to protect CENP-A in the human pathogenic yeast *Candida albicans*. PLoS Genet. 8: e1002661. https://doi.org/10.1371/journal.pgen.1002661

Thakur J., and K. Sanyal, 2013 Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in *Candida albicans*. Genome Res. 23: 638–652. https://doi.org/10.1101/gr.141614.112

Tietz H. J., A. Küssner, M. Thanos, M. P. De Andrade, W. Presber, *et al.*, 1995 Phenotypic and genotypic characterization of unusual vaginal isolates of *Candida albicans* from Africa. J. Clin. Microbiol. 33: 2462 LP – 2465.

Todd R. T., A. Forche, and A. Selmecki, 2017 Ploidy Variation in Fungi – Polyploidy, Aneuploidy, and Genome Evolution. Microbiol. Spectr. 5: 10.1128/microbiolspec.FUNK-0051–2016. https://doi.org/10.1128/microbiolspec.FUNK-0051-2016

Todd R. T., T. D. Wikoff, A. Forche, and A. Selmecki, 2019 Genome plasticity in *Candida albicans* is driven by long repeat sequences, (K. J. Verstrepen, D. Weigel, and M. Anderson, Eds.). Elife 8: e45954. https://doi.org/10.7554/eLife.45954

Tomaska L., J. Nosek, J. Kramara, and J. D. Griffith, 2009 Telomeric circles: universal players in telomere maintenance? Nat. Struct. Mol. Biol. 16: 1010–1015. https://doi.org/10.1038/nsmb.1660

Tsai S. Q., Z. Zheng, N. T. Nguyen, M. Liebers, V. V Topkar, *et al.*, 2015 GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33: 187–197. https://doi.org/10.1038/nbt.3117

Tso G. H. W., J. A. Reales-Calderon, A. S. M. Tan, X. Sem, G. T. T. Le, *et al.*, 2018 Experimental evolution of a fungal pathogen into a gut symbiont. Science (80-.). 362: 589 LP – 595. https://doi.org/10.1126/science.aat0537

Tutaj H., E. Pogoda, K. Tomala, and R. Korona, 2019 Gene overexpression screen for chromosome instability in yeast primarily identifies cell cycle progression genes. Curr. Genet. 65: 483–492. https://doi.org/10.1007/s00294-018-0885-x

Tzung K. W., R. M. Williams, S. Scherer, N. Federspiel, T. Jones, *et al.*, 2001 Genomic evidence for a complete sexual cycle in *Candida albicans*. Proc. Natl. Acad. Sci. U. S. A. 98: 3249–3253. https://doi.org/10.1073/pnas.061628798

Uhl M. A., M. Biery, N. Craig, and A. D. Johnson, 2003 Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen *C. albicans*. EMBO J. 22: 2668–2678. https://doi.org/10.1093/emboj/cdg256

Vyas V. K., M. I. Barrasa, and G. R. Fink, 2015 A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. Sci. Adv. 1: e1500248. https://doi.org/10.1126/sciadv.1500248

Vyas V. K., G. G. Bushkin, D. A. Bernstein, M. A. Getz, M. Sewastianik, *et al.*, 2018 New CRISPR Mutagenesis Strategies Reveal Variation in Repair Mechanisms among Fungi. mSphere 3. https://doi.org/10.1128/mSphere.00154-18

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

Wang J. M., R. J. Bennett, and M. Z. Anderson, 2018 The Genome of the Human Pathogen *Candida albicans* Is Shaped by Mutation and Cryptic Sexual Recombination. MBio 9: e01205-18. https://doi.org/10.1128/mBio.01205-18

Wensing L., J. Sharma, D. Uthayakumar, Y. Proteau, A. Chavez, *et al.*, 2019 A CRISPR Interference Platform for Efficient Genetic Repression in *Candida albicans.* mSphere 4. https://doi.org/10.1128/mSphere.00002-19

Westermann S., D. G. Drubin, and G. Barnes, 2007 Structures and Functions of Yeast Kinetochore Complexes. Annu. Rev. Biochem. 76: 563–591. https://doi.org/10.1146/annurev.biochem.76.052705.160607

Wickes B., J. Staudinger, B. B. Magee, K. J. Kwon-Chung, P. T. Magee, *et al.*, 1991 Physical and genetic mapping of *Candida albicans*: several genes previously assigned to chromosome 1 map to chromosome R, the rDNA-containing linkage group. Infect. Immun. 59: 2480–2484.

Wiele T. Van de, P. Van den Abbeele, W. Ossieur, S. Possemiers, and M. Marzorati, 2015 The Simulator of the Human Intestinal Microbial Ecosystem (SHIME(®))., pp. 305–317 in edited by Verhoeckx K., Cotter P., López-Expósito I., Kleiveland C., Lea T., *et al.* Cham (CH).

Wood A. J., T.-W. Lo, B. Zeitler, C. S. Pickle, E. J. Ralston, *et al.*, 2011 Targeted Genome Editing Across Species Using ZFNs and TALENs. Science (80-.). 333: 307 LP – 307. https://doi.org/10.1126/science.1207773

Wright W. D., S. S. Shah, and W.-D. Heyer, 2018 Homologous recombination and the repair of DNA double-strand breaks. J. Biol. Chem. 293: 10524–10535. https://doi.org/10.1074/jbc.TM118.000372

Yadav V., L. Sreekumar, K. Guin, and K. Sanyal, 2018 Five pillars of centromeric chromatin in fungal pathogens. PLOS Pathog. 14: e1007150.

Yim E., K. E. O'Connell, J. St Charles, and T. D. Petes, 2014 High-resolution mapping of two types of spontaneous mitotic gene conversion events in *Saccharomyces cerevisiae*. Genetics 198: 181–192. https://doi.org/10.1534/genetics.114.167395

Yin Y., and T. D. Petes, 2015 Recombination between homologous chromosomes induced by unrepaired UV-generated DNA damage requires Mus81p and is suppressed by Mms2p. PLoS Genet. 11: e1005026. https://doi.org/10.1371/journal.pgen.1005026

Yona A. H., Y. S. Manor, R. H. Herbst, G. H. Romano, A. Mitchell, *et al.*, 2012 Chromosomal duplication is a transient evolutionary solution to stress. Proc. Natl. Acad. Sci. 109: 21010 LP – 21015. https://doi.org/10.1073/pnas.1211150109

Yu E. Y., 2012 Telomeres and telomerase in *Candida albicans*. Mycoses 55: e48–e59. https://doi.org/10.1111/j.1439-0507.2011.02123.x

Zhai B., M. Ola, T. Rolling, N. L. Tosini, S. Joshowitz, *et al.*, 2020 High-resolution mycobiota analysis reveals dynamic intestinal translocation preceding invasive candidiasis. Nat. Med. 26: 59–64. https://doi.org/10.1038/s41591-019-0709-7

Zhang Y., M. L. Hefferin, L. Chen, E. Y. Shim, H.-M. Tseng, *et al.*, 2007 Role of Dnl4–Lif1 in nonhomologous end-joining repair complex assembly and suppression of homologous recombination. Nat. Struct. Mol. Biol. 14: 639–646. https://doi.org/10.1038/nsmb1261

Zhang N., B. B. Magee, P. T. Magee, B. R. Holland, E. Rodrigues, *et al.*, 2015 Selective Advantages of a Parasexual Cycle for the Yeast *Candida albicans.* Genetics 200: 1117–1132. https://doi.org/10.1534/genetics.115.177170

Zhang J.-M., T. Yadav, J. Ouyang, L. Lan, and L. Zou, 2019a Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. Cell Rep. 26: 955-968.e3. https://doi.org/https://doi.org/10.1016/j.celrep.2018.12.102

Zhang K., D.-Q. Zheng, Y. Sui, L. Qi, and T. D. Petes, 2019b Genome-wide analysis of genomic alterations induced by oxidative DNA damage in yeast. Nucleic Acids Res. 47: 3521–3535. https://doi.org/10.1093/nar/gkz027

Zhu Z., W.-H. Chung, E. Y. Shim, S. E. Lee, and G. Ira, 2008 Sgs1 helicase and two nucleases Dna2 and Exo1 resect DNA double-strand break ends. Cell 134: 981–994. https://doi.org/10.1016/j.cell.2008.08.037

Zordan R. E., D. J. Galgoczy, and A. D. Johnson, 2006 Epigenetic properties of white–opaque switching in *Candida albicans* are based on a self-sustaining transcriptional feedback loop. Proc. Natl. Acad. Sci. 103: 12807 LP – 12812. https://doi.org/10.1073/pnas.0605138103

Appendix

Appendix 1:

Identification of Recessive Lethal Alleles in the Diploid Genome of a *Candida albicans* Laboratory Strain Unveils a Potential Role of Repetitive Sequences in Buffering Their Deleterious Impact

Timea MARTON, Adeline FERI, Pierre-Henri COMMERE, Corinne MAUFRAIS, Christophe D'ENFERT, Melanie LEGRAND

mSphere Feb 2019, 4 (1) e00709-18; DOI: 10.1128/mSphere.00709-18

AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE
Host-Microbe Biology

Identification of Recessive Lethal Alleles in the Diploid Genome of a *Candida albicans* Laboratory Strain Unveils a Potential Role of Repetitive Sequences in Buffering Their Deleterious Impact

Timea Marton,[a,b] Adeline Feri,[a,b*] Pierre-Henri Commere,[c] Corinne Maufrais,[a,d] Christophe d'Enfert,[a] Melanie Legrand[a]

[a]Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, France
[b]Université Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, Paris, France
[c]Institut Pasteur, Unité de Technologie et de Service Cytométrie et Biomarqueurs, Plate-Forme Cytométrie, Paris, France
[d]Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI), USR 3756 IP CNRS, Institut Pasteur, Paris, France

ABSTRACT The heterozygous diploid genome of *Candida albicans* is highly plastic, with frequent loss of heterozygosity (LOH) events. In the SC5314 laboratory strain, while LOH events are ubiquitous, a chromosome homozygosis bias is observed for certain chromosomes, whereby only one of the two homologs can occur in the homozygous state. This suggests the occurrence of recessive lethal allele(s) (RLA) preventing large-scale LOH events on these chromosomes from being stably maintained. To verify the presence of an RLA on chromosome 7 (Chr7), we utilized a system that allows (i) DNA double-strand break (DSB) induction on Chr7 by the I-*Sce*I endonuclease and (ii) detection of the resulting long-range homozygosis. I-*Sce*I successfully induced a DNA DSB on both Chr7 homologs, generally repaired by gene conversion. Notably, cells homozygous for the right arm of Chr7B were not recovered, confirming the presence of RLA(s) in this region. Genome data mining for RLA candidates identified a premature nonsense-generating single nucleotide polymorphism (SNP) within the HapB allele of C7_03400c whose *Saccharomyces cerevisiae* ortholog encodes the essential Mtr4 RNA helicase. Complementation with a wild-type copy of *MTR4* rescued cells homozygous for the right arm of Chr7B, demonstrating that the *mtr4*^{K880*} RLA is responsible for the Chr7 homozygosis bias in strain SC5314. Furthermore, we observed that the major repeat sequences (MRS) on Chr7 acted as hot spots for interhomolog recombination. Such recombination events provide *C. albicans* with increased opportunities to survive DNA DSBs whose repair can lead to homozygosis of recessive lethal or deleterious alleles. This might explain the maintenance of MRS in this species.

IMPORTANCE *Candida albicans* is a major fungal pathogen, whose mode of reproduction is mainly clonal. Its genome is highly tolerant to rearrangements, in particular loss of heterozygosity events, known to unmask recessive lethal and deleterious alleles in heterozygous diploid organisms such as *C. albicans*. By combining a site-specific DSB-inducing system and mining genome sequencing data of 182 *C. albicans* isolates, we were able to ascribe the chromosome 7 homozygosis bias of the *C. albicans* laboratory strain SC5314 to an heterozygous SNP introducing a premature STOP codon in the *MTR4* gene. We have also proposed genome-wide candidates for new recessive lethal alleles. We additionally observed that the major repeat sequences (MRS) on chromosome 7 acted as hot spots for interhomolog recombination. Maintaining MRS in *C. albicans* could favor haplotype exchange, of vital importance to LOH events, leading to homozygosis of recessive lethal or deleterious alleles that inevitably accumulate upon clonality.

Citation Marton T, Feri A, Commere P-H, Maufrais C, d'Enfert C, Legrand M. 2019. Identification of recessive lethal alleles in the diploid genome of a *Candida albicans* laboratory strain unveils a potential role of repetitive sequences in buffering their deleterious impact. mSphere 4:e00709-18. https://doi.org/10.1128/mSphere.00709-18.

Editor Aaron P. Mitchell, Carnegie Mellon University

Address correspondence to Melanie Legrand, melanie.legrand@pasteur.fr.

* Present address: Adeline Feri, Pathoquest, BioPark, Paris, France.

Received 25 December 2018
Accepted 24 January 2019
Published 13 February 2019

KEYWORDS *Candida albicans*, homologous recombination, loss of heterozygosity, major repeat sequences, recessive lethal alleles

In diploid genomes, new mutations are heterozygous, and their effect is generally masked by the presence of the ancestral allele. As claimed in Haldane's sieve, only mutations that confer a fitness advantage as heterozygotes can invade the population. Although true, it does not make specific prediction about the fitness of the mutant homozygotes. Recent studies of *Saccharomyces cerevisiae* have observed maintenance of genetic variation due to heterozygote advantage, as a result of overdominance of mutated alleles (1). In addition, Gerstein and colleagues used the model organism *S. cerevisiae* to show that recessive beneficial mutations can avoid Haldane's sieve in clonal organisms, through rapid loss of heterozygosity (LOH), and thus contribute to rapid evolutionary adaptation (1). Similarly, in *Candida albicans*, mutations followed by genomic rearrangements such as LOH events and isochromosome formation have been associated with the acquisition of antifungal resistance (2, 3), bringing forth the idea that mechanisms favoring genome plasticity could contribute to *C. albicans* fitness within the host and upon exposure to antifungal agents. *C. albicans* is a frequent human commensal yeast responsible for both mucosal fungal infections and the majority of life-threatening nosocomial fungal infections (4). Its diploid genome displays a high degree of plasticity that includes, in particular, LOH events. Despite frequent LOH, overall heterozygosity is maintained in the *C. albicans* population, as illustrated by various studies that highlighted the elevated levels of natural heterozygosity, with a heterozygous position every ~200 to 250 bp on average (5–7). Furthermore, several studies revealed that genome heterozygosity showed a significant correlation with higher growth rates (6, 8, 9).

Essentially, mutations can be categorized as beneficial, harmful, or neutral and can be differently assigned depending on the organism's environment. Because the mode of reproduction of *C. albicans* is mainly clonal, and therefore mimics inbreeding in higher eukaryotes, an increased number of recessive lethal alleles (RLA) in the *C. albicans* genome is expected compared to other eukaryotes that undergo true sexual reproduction. Various types of mutations can impact the functionality of alleles and render them inactive; however, mutations introducing premature STOP codons would convey the most obvious effect. Within the *C. albicans* laboratory strain SC5314, Muzzey et al. (10) reported almost 200 genes for which one of the alleles contains a single nucleotide polymorphism (SNP) that introduces a premature STOP codon. Functional differences have already been reported for the two alleles of a heterozygous gene, and in all instances, the effect of the recessive mutation was visible only upon homozygosis toward the mutated allele (*HIS4* [11], *MBP1* [12], *GPI16/MRF2* [13]). Moreover, SNPs in promoter regions have been shown to alter expression regulation of two alleles (14). Of interest, LOH is pervasive in *C. albicans* isolates, as homozygous regions can be found in all sequenced isolates and affect all of the chromosomes. These LOH events vary in size: they can be limited to a single chromosomal region, affect an entire chromosomal arm, or even cover the entire chromosome (6, 7, 9).

Recently, a combination of molecular tools has been developed to study genome dynamics in *C. albicans*. First, an LOH reporter system takes advantage of fluorescent markers at an artificial heterozygous locus containing the BFP and GFP genes (15). Consequently, the appearance of spontaneous LOH events for the given locus can be monitored by the fluorescent status of cells using flow cytometry (15). Second, LOH events are often a result of DNA double-strand breaks (DSB) (16) resolved by means of various DNA repair mechanisms which can either be independent or dependent of homologous recombination. Feri et al. (13) developed an inducible, locus-specific DNA DSB system that uses the I-*Sce*I meganuclease from *S. cerevisiae*. When coupled to the BFP/GFP LOH reporter system, this system can be used to study the consequence of a targeted DNA DSB on the appearance of LOH events. Indeed, Feri et al. (13) have shown that I-*Sce*I-induced DNA breaks are predominantly repaired by gene conversion result-

ing in limited LOH. Nevertheless, various patterns of long-range LOH can also be obtained. Of note, the engineered system, alongside sequence resources, helped identify a RLA on Chr4B of *C. albicans* strain SC5314 (13). This RLA is the consequence of a nonsense mutation in the *GPI16* gene that encodes an essential component of the glycosylphosphatidylinositol (GPI) anchor biosynthetic machinery and explains why Chr4B is never observed in the homozygous state in *C. albicans* strains SC5314. Notably, this RLA appeared unique to strain SC5314 (13).

Although LOH can be observed on all eight chromosome pairs, prior studies conducting haplotype characterization of (i) progeny from the parasexual life cycle (17), (ii) homozygous diploid isolates derived from *RAD*52 double knockout mutants (18), and (iii) haploid strains of *C. albicans* (8) showed a chromosome homozygosis bias in the *C. albicans* laboratory strain SC5314. This suggests that mutations, potentially RLAs, could apply constraints on the directionality of LOH events. Indeed, the homozygosis state of some chromosomes was observed only for a given homolog while recurrently absent for the other homolog. This is the case for chromosomes 1, 4, 6, as well as chromosome 7 (Chr7) for which homozygosis of haplotype B (HapB) is never observed, while haplotype A homozygosis is, suggesting the presence of RLAs on Chr7 HapB (Chr7B).

In this study, we aimed to identify the RLA(s) on Chr7B using an approach similar to that developed by Feri et al. (13) when searching for RLAs on Chr4B. This approach also allowed addressing the role that repetitive sequences, such as the major repeat sequences (MRS), might play on the overall genome dynamics of *C. albicans*. MRS are unique to *C. albicans* and *Candida dubliniensis*, and are found throughout their genomes. MRS are composed of three subregions: RB2 which contains the *FRG8* gene, the RPS region which varies in repeat numbers (and thus, in size), and the HOK region. *C. albicans* possesses eight MRS, one on each chromosome with the exceptions of Chr7 where the presence of one MRS on each arm is observed and of Chr3 where an incomplete MRS is located (19). MRS expansion and contraction have previously been shown to be involved in chromosome loss where the chromosome copy containing the shorter MRS region is spontaneously lost (20). Furthermore, MRS have also been shown to be involved in chromosome translocation (19), when two different chromosomes exchange large regions of an arm. Results presented below confirm that homozygosis of Chr7B is not recovered and that this is the consequence of a premature STOP codon in the Chr7B-borne allele of the RNA helicase-encoding gene *MTR4*. Furthermore, we highlight that repeat regions such as MRS are hot spots for interhomolog recombination upon DNA repair and play a role in LOH dynamics in *C. albicans*.

RESULTS

Strain engineering to promote and detect long-range LOH on Chr7. Genome analysis revealed that the left arm of Chr7 carries only 9 heterozygous SNPs in 3 ORFs, while the right arm of Chr7 carries 784 heterozygous SNPs in 105 ORFs. Because RLAs are more likely to be found in heterozygous regions, we focused on the right arm of Chr7 to understand Chr7 homozygosis bias. To efficiently screen for the presence of RLAs on Chr7 right arm, we engineered strains carrying an artificial heterozygous BFP/GFP LOH reporter system (15) associated with an I-*Sce*I DNA DSB-inducing system (13) (Fig. 1A). Because we chose to insert the I-*Sce*I target site (TS) associated with the auxotrophic marker *URA3* conferring uridine prototrophy in the most *mrs-7b* proximal, gene-free region found on the right arm of Chr7, this setup allows rendering a maximum number of alleles homozygous on this arm while avoiding *mrs-7b*. Integration of the *URA3*-I-*Sce*I TS cassette can occur on either Chr7 homologs (Chr7A or Chr7B); thus, transformants were screened by SNP-RFLP to identify which Chr7 haplotype was targeted (see Fig. S1 in the supplemental material). Using the heterozygous SNP at position 727,328, we showed that 28/51 *C. albicans* transformants had integrated the I-*Sce*I TS on Chr7A (55%) and 23/51 on Chr7B (45%), demonstrating the absence of integration bias for this locus. Two independent transformants that had integrated the

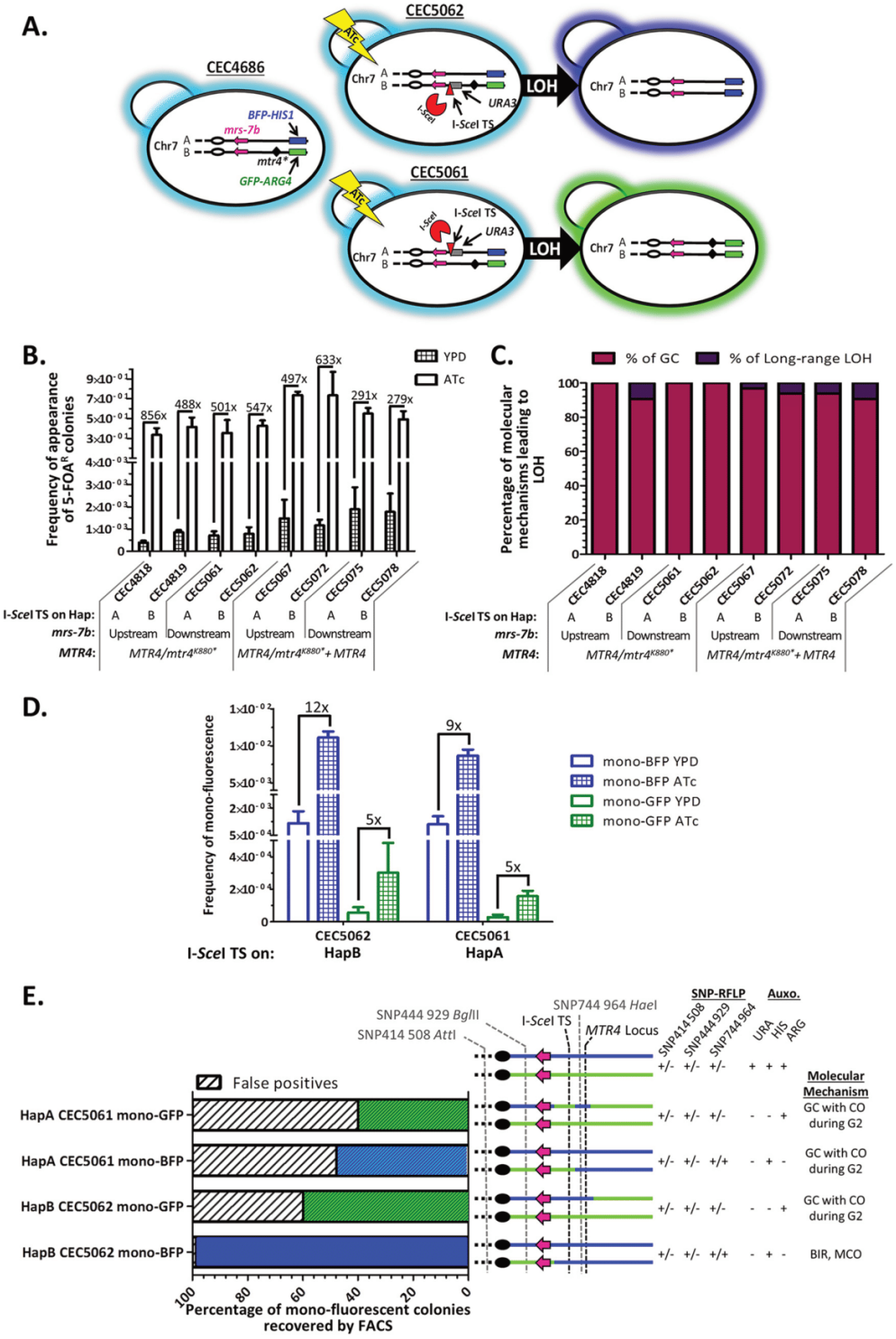

FIG 1 Coupling of a DNA double-strand-break-inducing system and a FACS-optimized LOH reporter system on Chr7. (A) Illustration of strains bearing the BFP/GFP LOH reporter system on the right arm of Chr7 associated with an I-*Sce*I-inducible DNA DSB system on Chr7B (CEC5062) or Chr7A (CEC5061). Upon addition of ATc, the mega-endonuclease I-*Sce*I is expressed and will generate a DNA DSB

(Continued on next page)

I-*Sce*I TS on Chr7 HapA (CEC5061) or HapB (CEC5062), were selected and used in subsequent analysis (Fig. 1A).

Strains CEC5061 and CEC5062 underwent preliminary characterization regarding their fluorescence status, as well as their growth rate to ensure that the successive transformation steps did not significantly alter their fitness. The fluorescence status of the intermediate and final strains was verified by flow cytometry. The flow cytometry outputs clearly displayed (i) the absence of fluorescence signals in the parental strain CEC4591, (ii) a shift toward the mono-GFP gate upon integration of the P_{TDH3}-*GFP-ARG4* cassette in CEC4679, (iii) a shift toward the double-fluorescent BFP/GFP gate upon integration of the P_{TDH3}-*BFP-HIS1* cassette in CEC4685, and (iv) the double-fluorescence status of the population upon integration of the *URA3*-I-*Sce*I-TS cassette in strains CEC5061 and CEC5062 (Fig. S2A).

Growth curves performed demonstrated that only the insertion of the *URA3*-I-*Sce*I TS cassette in strains CEC5061 and CEC5062 resulted in a higher growth rate, almost certainly due to the uridine prototrophy in these strains, as *URA3* deletion has been shown to result in significant decreases in *C. albicans* growth rate even when *ura3*Δ strains are grown in rich medium (21) (Fig. S2B).

Validation of the I-*Sce*I DNA DSB induction system on Chr7 by 5-FOA counter-selection. As the I-*Sce*I TS is associated with the genetic auxotrophic marker *URA3*, we could assess the frequency of cells that have survived an I-*Sce*I-induced DNA DSB at the TS by monitoring the frequency of appearance of 5-fluoroorotic acid-resistant (5-FOA^R) colonies upon I-*Sce*I induction. Indeed, 5-FOA^R colonies should have lost the *URA3* genetic marker (uridine auxotrophy) and are likely to represent cells that have sustained an I-*Sce*I-dependent DNA DSB through a LOH event, even though point mutations in the *URA3* gene cannot be excluded. We obtained 501 times more 5-FOA^R colonies after I-*Sce*I induction when the I-*Sce*I TS was located on Chr7A and 547 times more 5-FOA^R colonies when the I-*Sce*I TS was located on Chr7B (Fig. 1B). These data confirmed the efficiency of the I-*Sce*I-dependent DNA DSB induction system on Chr7. The majority of I-*Sce*I DNA DSB-induced 5-FOA^R colonies (90 to 100%) resulted from DNA DSB repair by gene conversion, as suggested by fluorescence and auxotrophy profiles of 32 Ura$^-$ colonies (Fig. 1C). Similar to what has been observed for Chr4 (13), DNA DSBs by I-*Sce*I on Chr7 are predominantly repaired by gene conversion repair mechanisms, resulting in short-range LOH events.

An I-SceI-induced DNA DSB on Chr7B leads to viable cells homozygous for the right arm of Chr7A. Although the 5-FOA assays yield information on the overall occurrence of LOH events encompassing the *URA3* gene, it does not allow us to efficiently study the underrepresented long-range LOH events. Thus, we also investigated LOH frequency upon I-*Sce*I expression using flow cytometry, an assay that specifically detects long-range LOH events. As expected, upon induction of I-*Sce*I in strain CEC5062, possessing the I-*Sce*I TS on the GFP-bearing Chr7B, we observed a

FIG 1 Legend (Continued)
at its target sequence (I-*Sce*I TS linked to the *URA3* marker) on Chr7. Theoretically, when the DNA DSB is repaired by break-induced replication or mitotic crossover, doubly fluorescent cells harboring the I-*Sce*I-TS on BFP-bearing Chr7A will become mono-GFP, while doubly fluorescent cells harboring the I-*Sce*I-TS on GFP-bearing Chr7B will become mono-BFP. The centromere (oval), MRS (pink arrow), and candidate RLA (diamond) on Chr7 are shown. (B) Average frequencies of 5-FOA^R colonies obtained after 8 h of I-*Sce*I overexpression and recovery from three independent experiments ($n = 3$) (plus standard deviations [S.D.] [error bars]) alongside fold changes observed between YPD and ATc conditions. (C) Percentage of molecular mechanisms leading to LOH among 5-FOA^R colonies in induced condition ($n = 32$). (D) Histogram representing average frequency ($n = 6$) (plus S.D.) of appearance of monofluorescent cells in the presence (ATc) and absence (YPD) of expression of I-*Sce*I and hence induced DNA DSBs on Chr7B (CEC5062) or Chr7A (CEC5061). Fold changes between uninduced and induced conditions for each monofluorescent population are indicated. (E) Characterization of monofluorescence-sorted individuals ($n = 32$) from induced CEC5062 (I-*Sce*I TS on HapB) and CEC5061 (I-*Sce*I TS on HapA) strains. A subset of sorted individuals was characterized for fluorescence and auxotrophy status in addition to SNP-RFLP for haplotype appointment in order to profile homozygosis of the right arm of Chr7 and determine the molecular mechanism used for I-*Sce*I-induced DNA DSB repair. In the stacked histogram, white and black hatched portions represent missorted cells (false-positive results), while green and blue portions represent the proportions of properly FACS-sorted monofluorescent individuals. The presence of stripes on the green or blue bars indicate that those individuals are not fully homozygous for one given haplotype from the I-*Sce*I TS to the BFP/GFP LOH reporter system on the right arm of Chr7. Abbreviations of molecular mechanisms are as follows; gene conversion (GC), break-induced replication (BIR), mitotic crossover (MCO), gene conversion with crossover (GC with CO).

12-fold increase in the appearance of mono-BFP cells (Fig. 1D). We also observed a fivefold increase in the appearance of mono-GFP cells (Fig. 1D). A subset of each population was recovered by fluorescence-activated cell sorting (FACS) and further characterized. We observed that, while the majority of the mono-BFP population included true mono-BFP cells displaying complete Chr7A homozygosis distal to the I-*Sce*I TS, 100% of the rare true mono-GFP cells displayed only partial homozygosis of Chr7B. From a mechanistic point of view, the mono-BFP cells resulted most likely from the repair of the DNA DSBs by mechanisms of break-induced replication or mitotic crossover, while the mono-GFP cells could be one of the possible outcomes of DNA DSB repair by gene conversion with crossover during the G_2 phase of the cell cycle (Fig. 1E).

Absence of recovery of cells being fully homozygous for the right arm of Chr7B. Unlike targeting the right arm of Chr7B, a DNA DSB on the right arm of Chr7A in strain CEC5061 should lead to a higher increase in frequency of the mono-GFP cells compared to the mono-BFP cells (Fig. 1D). Although an augmentation in frequency of both mono-BFP and mono-GFP cells was obtained, the mono-BFP cells still appeared at a higher frequency, 8×10^{-3} ($\pm 8 \times 10^{-4}$), compared to 1×10^{-4} ($\pm 3 \times 10^{-5}$) for the mono-GFP cells in the induced condition (Fig. 1D). Characterization of a subset of FACS-sorted mono-BFP cells confirmed that the true mono-BFP cells had arisen from a DNA DSB repaired by gene conversion with crossover during G_2. Further characterization of FACS-sorted mono-GFP cells (corresponding to the expected fluorescence) revealed that the I-*Sce*I-induced homozygosis of Chr7B was only partial in the targeted region. Thus, rather than having experienced break-induced replication or mitotic crossover events extending from the I-*Sce*I site to the BFP/GFP locus, the rare mono-GFP individuals were likely to have resulted from DNA DSB repair by gene conversion with crossover during G_2 (Fig. 1E). In conclusion, we were unable to recover a single individual that had undergone complete homozygosis of the right arm of Chr7B distal to the I-*Sce*I TS. This suggests that complete homozygosis of the right arm of Chr7B is associated with lethality in the SC5314 genetic background, thus confirming the chromosome homozygosis bias previously observed and validating the hypothesis that the presence of RLA(s) in this region could be a cause.

A data mining strategy identifies a heterozygous mutation in the *MTR4* gene as a possible cause of the homozygosis bias of Chr7. Genome sequence data obtained from a collection of 182 *C. albicans* isolates (7), including the reference strain SC5314, was used to compile all heterozygous SNPs within ORFs of Chr7 and search for SNPs (i) generating a premature STOP codon, (ii) showing a heterozygous genotype in SC5314, (iii) never observed in the homozygous state in the collection of 182 genomes, and (iv) located in a coding region never found to be dispensable in *C. albicans*. Only one such SNP was identified, located at position 746,359 on Chr7B. In *C. albicans* strain SC5314, this SNP causes a change from AAA (lysine) on Chr7A to TAA (STOP) on Chr7B in the C7_03400C gene. This gene is the orthologue of *S. cerevisiae MTR4* that encodes an essential ATP-dependent RNA helicase involved in RNA processing in *S. cerevisiae* (22–24). In *C. albicans* SC5314, the HapA allele of *MTR4* encodes a full-length Mtr4 protein, while the HapB allele carrying the STOP-introducing SNP encodes a truncated $Mtr4^{K880*}$ protein that misses a C-terminal DSHCT domain, common to DEAD box helicases (Fig. 2A).

The *mtr4* ^{K880*} allele is responsible for the Chr7 homozygosis bias. To validate that the identified RLA candidate $mtr4^{K880*}$ is truly responsible for the Chr7 homozygosis bias, the full-length *MTR4* ORF was placed ectopically under the control of the P_{TDH3} constitutive promoter in strain CEC5061, giving rise to strain CEC5075 (Fig. 2B). Results shown in Fig. 1B and C confirmed that the DNA DSB-inducing system was functional in this strain and that the I-*Sce*I-induced DNA DSB was predominantly repaired by gene conversion as observed in other instances. Strikingly, flow cytometry analysis revealed a significant elevation in the fold increase of mono-GFP cells obtained upon induction of I-*Sce*I in strain CEC5075 by comparison to CEC5061 (Fig. 2C). Upon cell sorting and characterization of these mono-GFP cells, the majority (93%) appeared

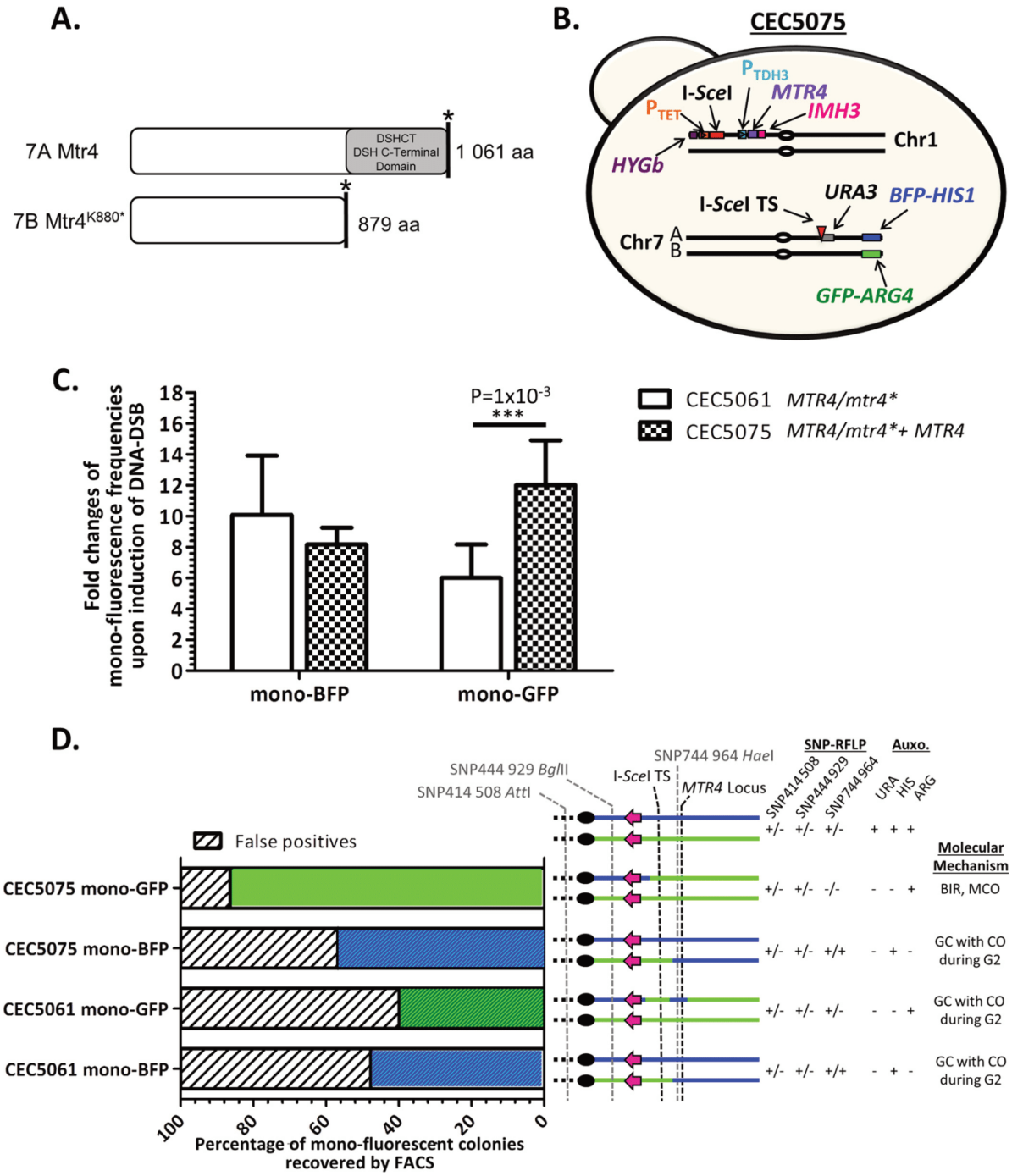

FIG 2 Complementation with a functional *MTR4* allele restores viability of Chr7B homozygous cells. (A) Schematic alignment of the wild-type *MTR4* and *mtr4*^{K880*}-encoded proteins. *MTR4* encodes an ATP-dependent RNA-helicase, and the nonsense mutation in the *mtr4*^{K880*} allele leads to loss of the DSH C-terminal domain shared by DEAD box helicases. aa, amino acids. (B) Illustration of the CEC5075 complemented strain possessing the functional *MTR4* allele at the *RPS1* locus on Chr1. Centromeres are indicated by ovals. (C) Comparison of the average fold increase of the monofluorescent populations upon DNA DSB induction (ATc/YPD) in both parental (CEC5061) and complemented (CEC5075) strains ($n = 6$) (plus S.D.), significance was determined using a bilateral *t* test (*P* value). (D) Characterization of monofluorescence-sorted individuals ($n = 32$) from induced CEC5061 (I-*Sce*I TS on HapA) and *MTR4*-complemented CEC5075 strains. See the legend to Fig. 1E for explanation.

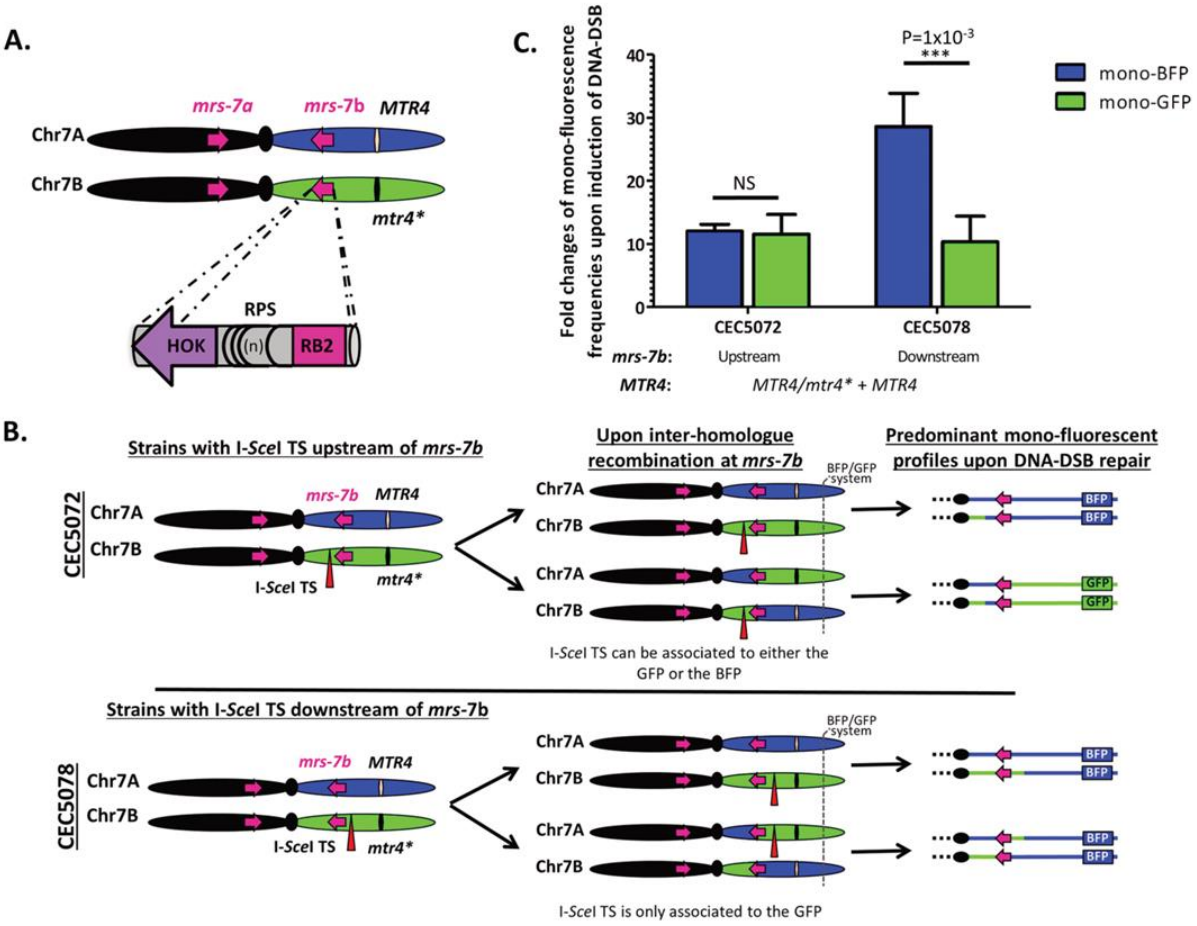

FIG 3 Upon DNA DSB, major repeat sequences are a source of interhomolog recombination. (A) Three subunits compose the MRS, which are hot spots for mitotic crossover. (B) Engineering of strains where the I-*Sce*I TS is located on HapB upstream (CEC5072) or downstream (CEC5078) of *mrs-7b*. Illustration of the right arm of Chr7 upon interhomolog recombination at *mrs-7b* followed by the predominant monofluorescence profiles upon DNA DSB repair leading to long-range LOH events. (C) Histogram showing average fold changes ($n = 6$) (plus S.D.) of monofluorescent populations upon I-*Sce*I induction (ATc versus YPD). While no significant difference in augmentation of both populations is detected when the I-*Sce*I TS is placed upstream of *mrs-7b* (CEC5072), a significant difference is observed when the I-*Sce*I TS is placed downstream of *mrs-7b* (CEC5078). Significance was determined using a bilateral *t* test (*P* value).

as true mono-GFP individuals that had become fully homozygous for Chr7B from the I-*Sce*I TS to the BFP/GFP LOH reporter system. These cells were likely to have resulted from the repair of the induced DNA DSB by break-induced replication or mitotic crossover, indicating that complete homozygosis of the right arm of Chr7B is compatible with viability upon addition of a functional *MTR4* allele (Fig. 2D). Similar to what was observed with the CEC5061 parental strain, I-*Sce*I induction in the CEC5075 *MTR4*-complemented strain resulted in an augmentation of mono-BFP cells (Fig. 2C). Eighty-eight percent of these mono-BFP cells corresponded to cells where the DNA DSB had been repaired by gene conversion with crossover during G_2 (Fig. 2D).

Upon DNA DSB, major repeat sequences are a source of interhomolog recombination. Our initial strategy for unveiling RLAs was to induce a DNA DSB downstream of *mrs-7b* (the MRS located on the right arm of Chr7 [Fig. 3A]) in order to ensure that the repeat sequences would not interfere with the DNA repair mechanisms. To seek validation of this initial assumption, the I-*Sce*I TS was moved upstream of *mrs-7b*, between the centromere and *mrs-7b*, on either HapA or HapB and in the presence of an ectopic copy of *MTR4* (Fig. 3B). By plating on 5-FOA-containing medium, we showed

TABLE 1 Summary of chromosome homozygosis bias observed in the literature

Chr	Parasexuality[a]	*RAD52* mutants[b]	Obligate diploid[c]	Bias summary
R	No BB	No BB		None
1	No AA, no BB	No BB	No BB	BB bias
2	No BB			None
3	No AA	No AA	No BB	None
4	No AA, no BB	No BB	No BB	BB bias
5		No AA		None
6	No BB	No BB	No BB	BB bias
7	No BB	No BB	No BB	BB bias

[a]Data from reference 17.
[b]Data from reference 18.
[c]Data from reference 8.

that the inducible I-*Sce*I DNA DSB system was functional in the new location and that gene conversion was the preferred mechanism of repair, independently of the targeted haplotype (Fig. 1B and C). Notably, when the I-*Sce*I TS was localized on HapB upstream of *mrs-7b* in an *MTR4*-complemented strain (CEC5072), induction of I-*Sce*I expression resulted in almost equal increases in the mono-GFP and mono-BFP cell populations relative to noninduced conditions (Fig. 3C). This contrasted to what was observed when the I-*Sce*I TS was localized on HapB downstream of *mrs-7b* in an *MTR4*-complemented strain (CEC5078), whereby a greater increase in the mono-BFP cell population than mono-GFP cell population was observed upon I-*Sce*I induction (Fig. 3C). This observation could be explained by interhomolog recombination at *mrs-7b*, which would link the I-*Sce*I TS to either the BFP or GFP markers when the TS is located upstream of *mrs-7b* (Fig. 3B). This would result in equal proportions of mono-BFP and mono-GFP cells upon repair of the I-*Sce*I-induced DNA DSB, keeping in mind that break-induced replication and mitotic crossover are the predominant molecular mechanisms leading to long-range LOH (Fig. 3C). In contrast, when the I-*Sce*I TS is located downstream of *mrs-7b*, it remains linked to the GFP marker, regardless of interhomolog recombination at *mrs-7b*, thus predominantly yielding mono-BFP cells upon I-*Sce*I-induced DNA DSB repair (Fig. 3B and C). Overall, our data suggest that the MRS is a hot spot for interhomolog recombination upon DNA repair on the right arm of Chr7.

Of interest, *C. albicans* Chr7 is characterized by the occurrence of a second MRS on the left arm, namely, *mrs-7a* (Fig. 3A). Strains possessing the I-*Sce*I TS between the centromere and *mrs-7a* and the BFP/GFP LOH reporter system on the left arm of Chr7 also exhibited an equal rate in mono-BFP and mono-GFP cells upon I-*Sce*I induction, whatever the location of the I-*Sce*I TS on HapA or HapB (data not shown), suggesting an increase in the number of cells that have undergone crossover at *mrs-7a*, linking the I-*Sce*I TS to either the BFP or GFP marker. Thus, *mrs-7a* also appears to be a hot spot for interhomolog recombination on Chr7.

DISCUSSION

Previous studies have shown that *C. albicans* strains harbor recessive lethal alleles (RLAs) that are responsible for a bias upon homozygosis of certain chromosomes, whereby only one of the two homologs can be retained in the homozygous state (Table 1). Yet, the nature of these RLAs is generally unknown. In this report, we have ascribed the Chr7 homozygosis bias of the *C. albicans* laboratory strain SC5314 to a heterozygous SNP introducing a premature STOP codon in the *MTR4* gene. Furthermore, we have unveiled the contribution of the major repeat sequence (MRS) to interhomolog recombination and hence, chromosome dynamics in *C. albicans*.

Our work focused on the homozygous bias observed for Chr7, suggesting the presence of at least one RLA on Chr7B (Table 1). Using a fluorescence-based LOH reporter system and an I-*Sce*I-dependent DNA DSB-inducing system, we have shown that while long-range homozygosis of Chr7A does not affect cell viability, long-range homozygosis of Chr7B is nonviable in *C. albicans* strain SC5314. A library of SNPs compiled from 182 clinical *C. albicans* genomes (7) was searched for SNPs that would

result in the most drastic outcome, i.e., a premature STOP codon. In this respect, we surely underestimated the presence of recessive lethal alleles, as mutations other than premature STOP-introducing SNPs are likely to result in lethal phenotypes. Indeed, nonsynonymous SNPs in coding regions have been shown to negatively impact protein function or regulation (25, 26). Our approach pinpointed an SNP in the *C. albicans MTR4* gene that complementation experiments confirmed to be the RLA responsible for the lethality of individuals homozygous for Chr7B in strain SC5314. The identified SNP results in a truncated form of Mtr4 that lacks the last 182 C-terminal amino acids, encompassing a DEAD box family DSHCT domain ($Mtr4^{K880*}$ [Fig. 2A]). In *S. cerevisiae*, *MTR4* encodes an ATP-dependent RNA helicase whose deletion results in nuclear accumulation of unprocessed RNAs (27), and reduction of function results in increased sensitivity to both benomyl and nocodazole (28). Importantly, it has been shown that the C-terminal domain of other fungal RNA helicases is critical for their proper RNA-unwinding function (29), consistent with $mtr4^{K880*}$ being a loss-of-function allele and homozygosity of this allele being lethal.

Our study and that of Feri et al. (13) indicate that the mining of the genomes of a large panel of *C. albicans* isolates for premature STOP-introducing SNPs is a suitable approach to identify RLAs responsible for chromosome homozygosis bias in *C. albicans* strains. As a reference, we provide a list of 70 genes that harbor a STOP-introducing SNP in one of the two alleles in the laboratory strain SC5314 in Table S1 in the supplemental material, including 12 alleles (Table 2) that were never found in the homozygous state in a collection of 182 genome-sequenced isolates representative of the *C. albicans* population (7). The locations of these 12 candidate RLAs across the *C. albicans* genome are shown in Fig. 4. Subsequent Sanger sequencing will be necessary to confirm these alleles. Our reduced number of genes with SNPs introducing a premature STOP codon in strain SC5314 compared to Muzzey et al. (10) could be explained by the stringency of our analysis. Indeed, only positions with high SNP quality and coverage depth of >20 for all 182 strains of our collection were considered for further analysis.

A promising candidate to explain the homozygosis bias observed for Chr1 is the *gcd6** allele (C1_08600C_B) that carries a SNP introducing a premature STOP codon, truncating the protein from the 17 C-terminal amino acids. In *S. cerevisiae*, *GCD6* encodes the catalytic epsilon subunit of the translation initiation factor eIF2B. The truncated domain is found at the C terminus of several translation initiation factors and is important for mediating protein-protein interactions. *GCD6* is essential in *S. cerevisiae*, as null mutants are not viable. The same is likely to be true in *C. albicans*, since heterozygous mutants are viable, but null mutants have not been reported in the literature. Therefore, it can be hypothesized that homozygosis of *gcd6**, located on Chr1B, could result in cell death and be responsible for the homozygosis bias observed on Chr1 (Table 1).

Interestingly, a recent transposon mutagenesis screen in haploid *C. albicans* also argues that Ca*GCD6* is essential (30). Among the 12 candidate RLAs presented in Table 2, the same screen argues for essentiality of Ca*GPI16* (previously identified by Feri et al. [13]) and Ca*MTR4* (identified in this study). Although the 10 remaining genes in Table 2 were defined as nonessential genes by Segal et al. (30), it is puzzling to see that the mutated alleles with a premature STOP codon were never observed in the homozygous state in the natural population of 182 *C. albicans* isolates. This observation highlights the complementarity of the study by Segal et al. (30) and our work, given that gene essentiality can vary depending on growth conditions and *in vitro* assessment of gene essentiality does not necessarily correlate with *in vivo* essentiality.

From a mechanistic point of view, our work revealed that most of the cells that have undergone an I-*Sce*I-mediated DNA DSB on Chr7 use gene conversion as a repair mechanism, thus limiting LOH extent and the loss of genetic information. Break-induced replication, mitotic crossover, gene conversion with crossover and chromosome truncation that lead to long-range LOH events in decreasing order are less frequently utilized. Our results are consistent with those of Feri et al. (13) on Chr4, arguing that our observation is not locus specific but can be applied genome-wide in

TABLE 2 Genome-wide RLA candidates in the *C. albicans* strain SC5314[a]

Chr	Position	ORF name	Gene name	Description	Haplotype of mutated allele	Essentiality in:	
						C. albicans	*S. cerevisiae*
Ca22chr1A	1131425	C1_05380C_A		Ortholog(s) have telomeric DNA binding activity, role in DNA double-strand break processing, DNA replication initiation, chromatin silencing at silent mating-type cassette, telomere capping, and shelterin complex localization	B	No info	No
Ca22chr1A	1760072	C1_08070W_A	CDR4	Putative ABC transporter superfamily; fluconazole, Sfu1, Hog1, core stress response induced; caspofungin repressed; fluconazole resistance not affected by mutation or correlated with expression; rat catheter and flow model biofilm induced	B	No	No
Ca22chr1A	1882812	C1_08600C_A	GCD6	Ortholog of *S. cerevisiae* Gcd6; catalytic epsilon subunit of the translation initiation factor eIF2B; genes encoding translation factors are repressed by phagocytosis by murine macrophages	B	Het ok	Yes
Ca22chr1A	1986384	C1_09100W_A	ZCF28	Zn(II)2Cys6 transcription factor; required for yeast cell adherence to silicone substrate; Spider biofilm induced	A	No	No
Ca22chr1B	634346	C1_03040W_B		Protein of unknown function; induced by nitric oxide; rat catheter and Spider biofilm induced	A	No info	No
Ca22chr1B	834720	C1_03960C_B		Ortholog(s) have role in TOR signaling, reentry into mitotic cell cycle after pheromone arrest and endoplasmic reticulum membrane, endoplasmic reticulum-Golgi intermediate compartment localization	A	No	No
Ca22chr4A	659155	C4_03130W_A		Putative GPI transamidase component; possibly an essential gene, disruptants not obtained by UAU1 method	B	Het ok (no UAU mutant)	Yes
Ca22chr4A	796679	C4_03750C_A		Ortholog(s) have translation release factor activity, role in mitochondrial translational termination and mitochondrial inner membrane localization	B	No info	No
Ca22chr4B	531082	C4_02560C_B	UME7	Putative transcription factor with zinc cluster DNA binding motif; similar to *S. cerevisiae* Ume6p, which is a transcription factor involved in the regulation of meiotic genes	No allelic variation in feature	No	No
Ca22chr7A	746359	C7_03400C_A		Ortholog of *S. cerevisiae* Mtr4, an ATP-dependent 3′–5′ RNA helicase of the DEAD box family; Hap43-induced gene; Spider biofilm induced	B	Het ok	Yes
Ca22chrRA	1603863	CR_07380C_A	KSR1	3-Ketosphinganine reductase, catalyzes the second step in phytosphingosine synthesis	B	No	Yes
Ca22chrRA	1782353	CR_08200C_A		Putative MRP/CFTR-subfamily ABC transporter; member of multidrug resistance-associated protein (MRP) subfamily of ABC family; similar to *S. cerevisiae* Bpt1p	B	Het ok	No

[a]Confirmed RLAs are indicated by medium gray shading (reference 13 and this study), and the recessive deleterious allele is indicated by light gray shading (13).

C. albicans. However, we cannot exclude the possibility that the relative usage of these repair mechanisms is specific to I-*Sce*I-induced DNA DSB, i.e., DNA DSBs induced by the CRISPR-*Cas9* RNA-guided endonuclease could be preferentially handled by another mechanism of DNA repair.

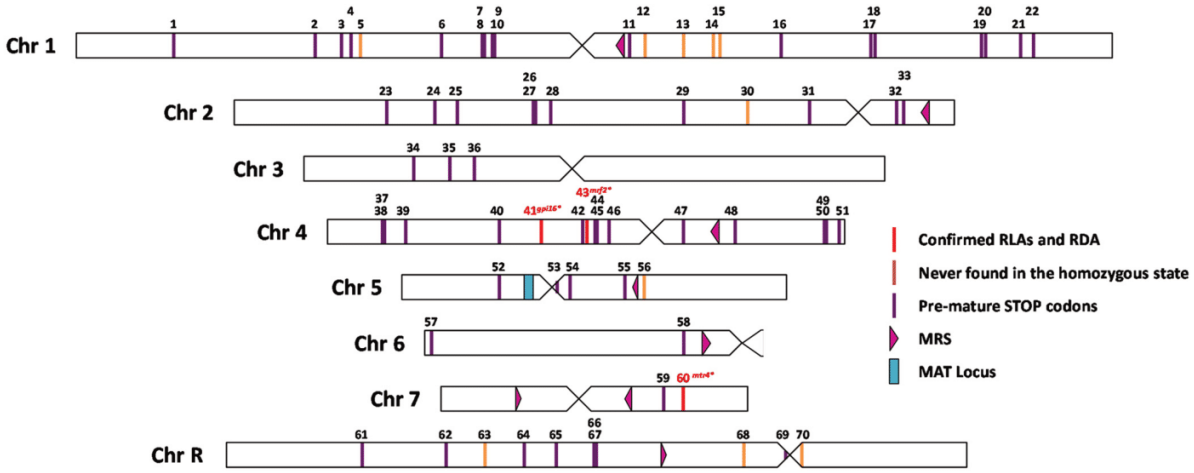

FIG 4 Schematic representation of the localization of premature STOP-inducing SNPs in the *C. albicans* SC5314 genome. Representation of 70 heterozygous SNPs (vertical bars) inducing a premature STOP codon distributed on all eight chromosomes of the reference strain SC5314. SNPs inducing premature STOP codons that are never found in the homozygous state in the library of 182 genomes of clinical *C. albicans* isolates are represented in orange, while confirmed recessive lethal and deleterious alleles are identified in red. Details regarding each SNP can be found in Table S1 using the assigned numbers in this figure.

In the course of this study, we also addressed the role that repetitive sequences, such as MRS, might play on the overall genome dynamics of *C. albicans*. This was achieved by studying how MRS position affects the outcome of I-*Sce*I-induced LOH. Even though most *C. albicans* chromosomes possess a unique MRS region, the presence of two MRS regions, one on each arm, is unique to Chr7 (19). We observed that the presence of *mrs-7b* or *mrs-7a* between the I-*Sce*I TS and the telomere-proximal LOH reporter system on Chr7 results in equal augmentation of the mono-BFP and mono-GFP populations upon I-*Sce*I-dependent DNA DSB (Fig. 3). This contrasts with what is observed when the I-*Sce*I TS is located downstream of the MRS, whereby the mono-fluorescent population arising by break-induced replication or mitotic crossover is increased relative to the monofluorescent population arising by gene conversion with crossover in the G_2 phase of the cell cycle (a scarce molecular mechanism), unless a RLA is present. This suggests that the MRS could be a hot spot for interhomolog mitotic crossover on Chr7. Indeed, upon recombination events at the MRS, DSBs repaired by break-induced replication or mitotic crossover would result in a relatively equal appearance of both monofluorescent populations when I-*Sce*I-TS is located between the centromere and the MRS (Fig. 3B and C).

Mitotic crossovers at MRS could be an intrinsic feature of these repeat regions or be triggered by either (i) stress resulting from I-*Sce*I overexpression or (ii) the physical I-*Sce*I-induced DNA DSB. The former would imply that increased mitotic crossovers at MRS should be expected on all chromosomes remaining to be tested. The latter would imply that increased mitotic crossovers at Chr7 MRS are observed only upon repair of a DNA DSB upstream of these repeated regions. These latter hypotheses are consistent with a role for stress in the enhancement of the recombination frequency near the MRS or in general, as already suggested by Lephart et al. (20). Importantly, our results confirm the original proposal of Pujol et al. (31) that MRS on Chr7 are hot spots for recombination. Concretely, MRS would allow switching *C. albicans* haplotypes by generating a new combination of alleles. Information regarding the biological importance of MRS remains scarce despite the positive selection on MRS leading to the maintenance of these large and unique repeats in the *C. albicans* genome. Such recombination events provide *C. albicans* with increased opportunities to survive DNA DSBs whose repair can lead to homozygosis of recessive lethal or deleterious alleles. This might explain the maintenance of MRS in this species.

TABLE 3 *C. albicans* strains used in this study

Strain	Genotype	Characteristic	Auxotrophies	Parental strain	Reference
CEC4591	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1*Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436	pNIMx-pTet-NLS-ISceI	arg− his− uri− HygR NsnR	CEC1448	13
CEC4679	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::PTDH3-GFP-CdARG4	CEC4591 + Ch7Right_pTDH3-ARG4-GFP	ARG+ his− uri− NsnR HygR	CEC4591	This study
CEC4685	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1	CEC4679 + Chr7Right_pTDH3-HIS1-BFP	ARG+ HIS+ uri− NsnR HygR	CEC4679	This study
CEC4818	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347:: PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:446552 to 446704::PTDH3-CdURA3-I-SceI-TS	CEC4685 + relocated I-SceI TS on HapA	ARG+ HIS+ URI+ NsnR HygR	CEC4685	This study
CEC4819	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1*Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347:: PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:446552 to 446704::PTDH3-CdURA3-I-SceI-TS	CEC4685 + relocated I-SceI TS on HapB	ARG+ HIS+ URI+ NsnR HygR	CEC4685	This study
CEC5061	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1*Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::PTDH3 -GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:725867 to 726122::PTDH3-CdURA3-I-SceI-TS	CEC4685 + I-SceI TS on HapA	ARG+ HIS+ URI+ NsnR HygR	CEC4685	This study
CEC5062	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1*Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347:: PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:725867 to 726122::PTDH3-CdURA3-I-SceI-TS	CEC4685 + I-SceI TS on HapB	ARG+ HIS+ URI+ NsnR HygR	CEC4685	This study
CEC5067	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1*Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347:: PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:446552 to 446704:: PTDH3-CdURA3-I-SceI-TS RPS1:: Clp10/RPS1::PTDH3-MTR4-IMH3r	CEC4818 + pTDH3-IMH3-MTR4	ARG+ HIS+ URI+ NsnR HygR MpaR	CEC4818	This study

(Continued on next page)

TABLE 3 (Continued)

Strain	Genotype	Characteristic	Auxotrophies	Parental strain	Reference
CEC5072	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1*Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347:: PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:446552 to 446704:: PTDH3-CdURA3-I-SceI-TS RPS1::Clp10/RPS1::PTDH3-MTR4-IMH3r	CEC4819 + pTDH3-IMH3-MTR4	ARG+ HIS+ URI+ NsnR HygR MpaR	CEC4819	This study
CEC5075	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1*Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347::PTDH 3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:725867 to 726122:: PTDH3-CdURA3-I-SceI-TS RPS1::Clp10/RPS1::PTDH3-MTR4-IMH3r	CEC5061 + pTDH3-IMH3-MTR4	ARG+ HIS+ URI+ NsnR HygR MpaR	CEC5061	This study
CEC5078	*arg4Δ/arg4Δ his1Δ/his1Δ ura3Δ::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 ADH1/adh1*::PTDH3-*carTA*::*SAT1* Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::ClpXH-HYGb-PTET-ATG-NLS-I-SceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca22chr7_C_albicans_SC5314:911053 to 911347:: PTDH3-GFP-CdARG4/PTDH3-BFP-CdHIS1 Ca22chr7_C_albicans_SC5314:725867 to 726122:: PTDH3-CdURA3-I-SceI-TS RPS1::Clp10/RPS1::PTDH3-MTR4-IMH3r	CEC5062 + pTDH3-IMH3-MTR4	ARG+ HIS+ URI+ NsnR HygR MpaR	CEC5062	This study

MATERIALS AND METHODS

Strains and media. *C. albicans* strains used in this work are derived from the reference strain SC5314. The cloning experiments were carried out using One Shot TOP10 chemically competent *Escherichia coli* cells (ThermoFisher Scientific). All *C. albicans* strains and *E. coli* plasmids generated and used throughout this investigation are listed in Table 3 and Table S2 in the supplemental material, respectively.

C. albicans cells were maintained on rich yeast extract-peptone-dextrose (YPD) medium (1% yeast extract, 2% peptone, 2% dextrose). Synthetic complete (SC) (0.67% yeast nitrogen base without amino acids, 2% dextrose, 0.08% drop out mix) and synthetic defined (SD) (0.67% yeast nitrogen base without amino acids, 2% dextrose) media, both with drop out amino acids, depending on auxotrophy of the strains, were used as selective media. Solid media were obtained by adding 2% agar. 5-Fluoroorotic acid (5-FOA)-containing agar medium (0.67% yeast nitrogen base without amino acid, 0.0625% 5-fluoroorotic acid [Toronto Research Chemicals], 0.01% uridine, 2% glucose, 2% agar supplemented with arginine and histidine) was used to detect cells that have an nonfunctional *URA3* gene. Mycophenolic acid (MPA) medium is composed of SD medium with 5 μg/ml MPA (Sigma-Aldrich). Meanwhile, *E. coli* strains were cultured on/in LB or 2YT medium, with appropriate antibiotics for selection purposes (kanamycin [50 μg/ml], ticarcillin [50 μg/ml], and gentamicin [10 μg/ml]).

***In silico* identification of potential recessive lethal alleles.** A collection of 182 *C. albicans* genomes (7) was utilized with the intention to search for potential RLAs. SNPs were compiled and filtered with the following criteria: (i) an SNP that introduces a premature STOP codon in a coding region (open reading frame [ORF]), (ii) this SNP was never observed in the homozygous state, (iii) the SNP lies in a coding region never found to be dispensable in *C. albicans*, and (iv) the SNP was found in the heterozygous state in the reference strain SC5314.

Plasmid constructions. Methods for the construction of plasmids used for I-*Sce*I target sequence (TS) insertion (pFA-*URA3*-I*Sce*I-TS-*HSP90/CUP9*) and for complementation with the wild-type *MTR4* allele (Clp10-P_{TDH3}-*MTR4-IMH3*) are described in Text S1 in the supplemental material.

***C. albicans* strain constructions.** Construction of strain CEC4685 is described in Text S1. Briefly, this strain is derived from the common laboratory strain SN76 and carries the I-*Sce*I-encoding gene under the control of the inducible P_{TET} promoter (32), pNIMX coding for the transactivator necessary for activation of the P_{TET} promoter in the presence of tetracycline derivatives (33), and the BFP/GFP LOH reporter system (15) on the right arm of Chr7. Integration of the I-*Sce*I TS was achieved by transformation of strain CEC4685 with a PCR-amplified cassette bearing the I-*Sce*I TS and Chr7 homology regions, the latter borne from primers 9 and 10 (Table S3) used in amplification of pFA-*URA3*-I*Sce*I-TS-*CDR3*/TG(GCC)2 (13). This led to prototroph strains named CEC5061 (I-*Sce*I TS on HapA) and CEC5062 (I-*Sce*I TS on HapB) (Table 3).

Relocalization of I-*Sce*I TS upstream of the MRS was conducted by transforming strain CEC4685 with *Hpa*I+*Sca*II-cut plasmid of pFA-*URA3*-I*Sce*I-TS-*HSP90/CUP9*, which carries the *URA3*-I-*Sce*I TS construction for integration on Chr7. The now prototroph transformants were selected on SD medium. The final strains were named CEC4818 (I-*Sce*I TS on HapA) and CEC4819 (I-*Sce*I TS on HapB) (Table 3).

Strains possessing the BFP/GFP LOH reporter system on the left arm of Chr7 were generated by positioning the heterozygous BFP/GFP locus close to the left telomere (position 108,458 to 108,838) and the I-*Sce*I TS at position 367,060 to 367,145 between *mrs-7a* (position 228,342 to 242,083) and the centromere (position 425,808 to 428,708).

Complementation of strains CEC5061, CEC5062, CEC4818, and CEC4819 with the wild-type *MTR4* allele at the *RPS1* locus was achieved using the *Stu*I-linearized Clp10-P_{TDH3}-*MTR4-IMH3* plasmid described in Text S1. Transformants were selected on SD medium containing 5 μg/ml MPA (Sigma-Aldrich), and the final strains were named CEC5067, CEC5072, CEC5075, and CEC5078, respectively (Table 3).

Following each transformation step, junction PCRs were conducted to ensure proper integrations. Additional controls were conducted such as fluorescence check (flow cytometry of 20,000 cells) and auxotrophy testing of strains (by spotting on SD supplemented with either Arg, His, Ura, or MPA). To ensure that the strains generated in the course of this study did not display growth defects resulting from transformation events, growth curves were generated using a TECAN Sunrise with OD readings (620 nm) every 10 min for a period of 48 h. Doubling times were outputted from the generated OD values and analyzed using GraphPad Prism 5.

Induction of the *Tet*-On system. In order to activate the *Tet*-On promoter and achieve I-*Sce*I protein production, single colonies were precultured in liquid SC-His-Arg medium at 30°C. After overnight growth, induction was conducted in YPD plus anhydrotetracycline (ATc) (3 μg/ml) (Thermofisher ACROS Organics) for 8 h at 30°C, followed by an overnight recovery in YPD.

5-Fluoroorotic acid selection. Following the I-*Sce*I induction protocol, as seen above, three different cell dilutions of cultures (20,000 cells, 2,000 cells, and 200 cells) grown in the presence (induced) or absence (noninduced; control) of ATc were plated on 5-fluoroorotic acid (5-FOA)-containing plates in triplicates. Dilutions were verified by plating a volume corresponding to 100 cells on YPD plates. The plates were incubated at 30°C for 3 days before analysis.

Cell preparation for flow cytometry and analysis. All flow cytometry analyses were conducted on the MACSQuant analyzer (Miltenyi Biotec) where BFP is detected with a 405-nm laser and 425- to 475-nm filters and GFP is detected with a 488-nm laser and 500- to 550-nm filters. Data for a maximum of 10^6 cells were analyzed using the FlowJo V10.1 software. The gates to determine the LOH frequencies were arbitrarily selected but conserved throughout sample analysis.

Cell sorting. Induced and noninduced cultures were filtered using BD Falcon Cell strainers in order to remove large debris and filamentous cells that could obstruct the tubing system of the cytometer. The MoFlo Astrios flow cytometer was used to analyze and sort the cells of interest. For each sorted gate, 1,000 cells were recovered in 400 μl of liquid YPD medium, plated immediately after cell sorting on four YPD petri plates, and incubated at 30°C for 48 h before collection of results.

SNP-RFLP. *In silico* identification of heterozygous SNPs affecting a restriction site on only one haplotype of Chr7 was used for haplotype characterization of strains. First, a nucleotide multiple sequence alignment by MUSCLE was conducted using the Chr7A and Chr7B sequences from the reference strain SC5314 (22). Heterozygous SNPs were selected using the following criteria. (i) It interrupts a commonly known restriction site on one haplotype. (ii) The selected enzyme does not cut again within a range of 1 kb. (iii) The heterozygous SNP is present in most strains of the collection of 182 *C. albicans* clinical strains (7). Second, primer pairs were designed to result in PCR products with different digestion profiles and used to verify the presence of the heterozygous SNP in our strain of interest.

These SNPs were used for two distinct purposes: (i) to assign the Chr7 homolog targeted by I-*Sce*I TS integration and (ii) to monitor the heterozygous status of the left and right arms of Chr7 upon DNA DSB repair. Regions of roughly 2.5 kb surrounding the heterozygous SNPs were amplified by PCR and digested with the appropriate restriction enzyme overnight. *C. albicans* gDNA extractions, PCRs, and amplicon verifications were conducted following the methods of Feri et al. (13), while a list of the primers used can be found in Table S3. The SNPs at positions 444,929 (*Bgl*II cutting HapA) and 727,328 (*Hpa*I cutting HapA) were used to identify the I-*Sce*I-targeted haplotype of the right arm of Chr7, when the I-*Sce*I TS is located upstream and downstream of the *mrs-7b*, respectively. The 2.2-kb and 2.16-kb regions around the heterozygous SNPs located at positions 414,508 (left arm) and 744,964 (right arm) utilizing *Att*I (cutting HapB) and *Hae*II (cutting HapB) enzymes, respectively, were used to assess the heterozygous status of both Chr7 arms. All SNP-RFLP sites and the enzyme-sensitive haplotypes are summarized in Fig. S1.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSphere.00709-18.

TEXT S1, DOCX file, 0.05 MB.
FIG S1, TIF file, 0.1 MB.
FIG S2, TIF file, 0.5 MB.
TABLE S1, DOCX file, 0.04 MB.
TABLE S2, DOCX file, 0.03 MB.
TABLE S3, DOCX file, 0.03 MB.

ACKNOWLEDGMENTS

T.M. is the recipient of a Ph.D. fellowship from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID). A.F. was the recipient of a Ph.D. fellowship from INRA and Institut Pasteur. We acknowledge support from the French Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases [ANR-10-LABX-62-IBEID]).

REFERENCES

1. Gerstein AC, Kuzmin A, Otto SP. 2014. Loss-of-heterozygosity facilitates passage through Haldane's sieve for *Saccharomyces cerevisiae* undergoing adaptation. Nat Commun 5:3819. https://doi.org/10.1038/ncomms4819.
2. Coste A, Selmecki A, Forche A, Diogo D, Bougnoux M-E, d'Enfert C, Berman J, Sanglard D. 2007. Genotypic evolution of azole resistance mechanisms in sequential *Candida albicans* isolates. Eukaryot Cell 6:1889–1904. https://doi.org/10.1128/EC.00151-07.
3. 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.
4. Prieto D, Correia I, Pla J, Román E. 2016. Adaptation of *Candida albicans* to commensalism in the gut. Future Microbiol 11:567–583. https://doi.org/10.2217/fmb.16.1.
5. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, 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.
6. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, Zeng Q, Zisson E, Wang JM, Greenberg JM, Berman J, Bennett RJ, Cuomo CA. 2015. Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res 25:413–425. https://doi.org/10.1101/gr.174623.114.
7. Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, Sertour N, Mosca K, Permal E, Laval G, Bouchier C, Ma L, Schwartz K, Voelz K, May RC, Poulain J, Battail C, Wincker P, Borman AM, Chowdhary A, Fan S, Kim SH, Le Pape P, Romeo O, Shin JH, Gabaldon T, Sherlock G, Bougnoux M-E, d'Enfert C. 2018. Gene flow contributes to diversification of the major fungal pathogen Candida albicans. Nat Commun 9:2253. https://doi.org/10.1038/s41467-018-04787-4.
8. Hickman MA, Zeng G, Forche A, Hirakawa MP, Abbey D, Harrison BD, Wang Y-M, Su C, Bennett RJ, Wang Y, Berman J. 2013. The 'obligate diploid' *Candida albicans* forms mating-competent haploids. Nature 494:55–59. https://doi.org/10.1038/nature11865.
9. Abbey DA, Funt J, Lurie-Weinberger MN, Thompson DA, Regev A, Myers CL, Berman J. 2014. YMAP: a pipeline for visualization of copy number variation and loss of heterozygosity in eukaryotic pathogens. Genome Med 6:1–16. https://doi.org/10.1186/s13073-014-0100-8.
10. Muzzey D, Schwartz K, Weissman JS, 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.
11. Gómez-Raja J, Andaluz E, Magee B, Calderone R, Larriba G. 2008. A single SNP, G929T (Gly310Val), determines the presence of a functional and a non-functional allele of *HIS4* in *Candida albicans* SC5314: detection of the non-functional allele in laboratory strains. Fungal Genet Biol 45:527–541. https://doi.org/10.1016/j.fgb.2007.08.008.
12. Ciudad T, Hickman M, Bellido A, Berman J, Larriba G. 2016. Phenotypic consequences of a spontaneous loss of heterozygosity in a common laboratory strain of *Candida albicans*. Genetics 203:1161–1176. https://doi.org/10.1534/genetics.116.189274.
13. Feri A, Loll-Krippleber R, Commere P-H, Maufrais C, Sertour N, Schwartz K, Sherlock G, Bougnoux M-E, D'Enfert C, Legrand M. 2016. Analysis of repair mechanisms following an induced double strand break uncovers recessive deleterious alleles in the *Candida albicans* diploid genome. mBio 7:e01109-16. https://doi.org/10.1128/mBio.01109-16.
14. Muzzey D, Sherlock G, Weissman JS. 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.
15. Loll-Krippleber R, Feri A, Nguyen M, Maufrais C, Yansouni J, d'Enfert C, Legrand M. 2015. A FACS-optimized screen identifies regulators of genome stability in *Candida albicans*. Eukaryot Cell 14:311–322. https://doi.org/10.1128/EC.00286-14.
16. Malkova A, Klein F, Leung W-Y, Haber JE. 2000. HO endonuclease-induced recombination in yeast meiosis resembles Spo11-induced events. Proc Natl Acad Sci U S A 97:14500–14505. https://doi.org/10.1073/pnas.97.26.14500.
17. Forche A, Alby K, Schaefer D, Johnson AD, Berman J, Bennett RJ. 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.
18. Andaluz E, Bellido A, Gómez-Raja J, Selmecki A, Bouchonville K, Calderone R, Berman J, Larriba G. 2011. Rad52 function prevents chromosome loss and truncation in *Candida albicans*. Mol Microbiol 79:1462–1482. https://doi.org/10.1111/j.1365-2958.2011.07532.x.
19. Chibana H, Magee PT. 2009. The enigma of the major repeat sequence of *Candida albicans*. Future Microbiol 4:171–179. https://doi.org/10.2217/17460913.4.2.171.
20. Lephart PR, Chibana H, Magee PT. 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.
21. Kirsch DR, Whitney RR. 1991. Pathogenicity of *Candida albicans* auxotrophic mutants in experimental infections. Infect Immun 59:3297–3300.
22. Skrzypek MS, Binkley J, Binkley G, Miyasato SR, 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.
23. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian K-D, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Güldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kötter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 418:387. https://doi.org/10.1038/nature00935.
24. Bernstein KA, Granneman S, Lee AV, Manickam S, Baserga SJ. 2006. Comprehensive mutational analysis of yeast DEXD/H box RNA helicases involved in large ribosomal subunit biogenesis. Mol Cell Biol 26:1195–1208. https://doi.org/10.1128/MCB.26.4.1195-1208.2006.
25. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ, Vernimmen D, Yoshinaga Y, de Jong P, Cheng J-F, Rubin EM, Wood WG, Bowden D, Higgs DR. 2006. A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science 312:1215–1217. https://doi.org/10.1126/science.1126431.
26. Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS, Maixner W, Diatchenko L. 2006. Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. Science 314:1930–1933. https://doi.org/10.1126/science.1131262.
27. Liang S, Hitomi M, Hu YH, Liu Y, Tartakoff AM. 1996. A DEAD-box-family protein is required for nucleocytoplasmic transport of yeast mRNA. Mol Cell Biol 16:5139–5146. https://doi.org/10.1128/MCB.16.9.5139.
28. Smith SB, Kiss DL, Turk E, Tartakoff AM, Andrulis ED. 2011. Pronounced and extensive microtubule defects in a *Saccharomyces cerevisiae* DIS3 mutant. Yeast 28:755–769. https://doi.org/10.1002/yea.1899.
29. Mohr G, Del Campo M, Mohr S, Yang Q, Jia H, Jankowsky E, Lambowitz AM. 2008. Function of the C-terminal domain of the DEAD-box protein

Mss116p analyzed *in vivo* and *in vitro*. J Mol Biol 375:1344–1364. https://doi.org/10.1016/j.jmb.2007.11.041.

30. Segal ES, Gritsenko V, Levitan A, Yadav B, Dror N, Steenwyk JL, Silberberg Y, Mielich K, Rokas A, Gow NAR, Kunze R, Sharan R, Berman J. 2018. Gene essentiality analyzed by *in vivo* transposon mutagenesis and machine learning in a stable haploid isolate of *Candida albicans*. mBio 9:e02048-18. https://doi.org/10.1128/mBio.02048-18.
31. Pujol C, Joly S, Nolan B, Srikantha T, Soll DR. 1999. Microevolutionary changes in *Candida albicans* identified by the complex Ca3 fingerprinting probe involve insertions and deletions of the full-length repetitive sequence RPS at specific genomic sites. Microbiology 145:2635–2646. https://doi.org/10.1099/00221287-145-10-2635.
32. Park Y-N, Morschhäuser J. 2005. Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. Eukaryot Cell 4:1328–1342. https://doi.org/10.1128/EC.4.8.1328-1342.2005.
33. Chauvel M, Nesseir A, Cabral V, Znaidi S, Goyard S, Bachellier-Bassi S, Firon A, Legrand M, Diogo D, Naulleau C, Rossignol T, d'Enfert C. 2012. A versatile overexpression strategy in the pathogenic yeast *Candida albicans*: identification of regulators of morphogenesis and fitness. PLoS One 7:e45912. https://doi.org/10.1371/journal.pone.0045912.

Appendix 2:

Use of CRISPR-Cas9 To Target Homologous Recombination Limits Transformation-Induced Genomic Changes in *Candida albicans*

Timea MARTON, Corinne MAUFRAIS, Christophe D'ENFERT, Melanie LEGRAND

mSphere Sep 2020, 5 (5) e00620-20; DOI: 10.1128/mSphere.00620-20

AMERICAN SOCIETY FOR MICROBIOLOGY | mSphere®

RESEARCH ARTICLE
Molecular Biology and Physiology

Use of CRISPR-Cas9 To Target Homologous Recombination Limits Transformation-Induced Genomic Changes in *Candida albicans*

Timea Marton,[a,b] Corinne Maufrais,[a,c] Christophe d'Enfert,[a] Melanie Legrand[a]

[a]Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, France
[b]Université Paris Diderot, Sorbonne Paris Cité, Paris, France
[c]Hub de Bioinformatique et Biostatistique, Département de Biologie Computationnelle, USR 3756 IP CNRS, Institut Pasteur, Paris, France

ABSTRACT Most of our knowledge relating to molecular mechanisms of human fungal pathogenesis in *Candida albicans* relies on reverse genetics approaches, requiring strain engineering. DNA-mediated transformation of *C. albicans* has been described as highly mutagenic, potentially accentuated by the organism's genome plasticity, including the acquisition of genomic rearrangements, notably upon exposure to stress. The advent of CRISPR-Cas9 has vastly accelerated the process of genetically modifying strains, especially in diploid (such as *C. albicans*) and polyploid organisms. The effects of unleashing this nuclease within the genome of *C. albicans* are unknown, although several studies in other organisms report Cas9-associated toxicity and off-target DNA breaks. Upon the construction of a *C. albicans* strain collection, we took the opportunity to compare strains which were constructed using CRISPR-Cas9-free and CRISPR-Cas9-dependent transformation strategies, by quantifying and describing transformation-induced loss-of-heterozygosity and hyperploidy events. Our analysis of 57 strains highlights the mutagenic effects of transformation in *C. albicans*, regardless of the transformation protocol, but also underscores interesting differences in terms of genomic changes between strains obtained using different transformation protocols. Indeed, although strains constructed using the CRISPR-Cas9-free transformation method display numerous concomitant genomic changes randomly distributed throughout their genomes, the use of CRISPR-Cas9 leads to a reduced overall number of genome changes, particularly hyperploidies. Overall, in addition to facilitating strain construction by reducing the number of transformation steps, the CRISPR-Cas9-dependent transformation strategy in *C. albicans* appears to limit transformation-associated genome changes.

IMPORTANCE Genome editing is essential to nearly all research studies aimed at gaining insight into the molecular mechanisms underlying various biological processes, including those in the opportunistic pathogen *Candida albicans*. The adaptation of the CRISPR-Cas9 system greatly facilitates genome engineering in many organisms. However, our understanding of the effects of CRISPR-Cas9 technology on the biology of *C. albicans* is limited. In this study, we sought to compare the extents of transformation-induced genomic changes within strains engineered using CRISPR-Cas9-free and CRISPR-Cas9-dependent transformation methods. CRISPR-Cas9-dependent transformation allows one to simultaneously target both homologs and, importantly, appears less mutagenic in *C. albicans*, since strains engineered using CRISPR-Cas9 display an overall decrease in concomitant genomic changes.

KEYWORDS CRISPR-Cas9, *Candida albicans*, genome rearrangements

Citation Marton T, Maufrais C, d'Enfert C, Legrand M. 2020. Use of CRISPR-Cas9 to target homologous recombination limits transformation-induced genomic changes in *Candida albicans*. mSphere 5:e00620-20. https://doi.org/10.1128/mSphere.00620-20.

Editor Aaron P. Mitchell, University of Georgia

Address correspondence to Melanie Legrand, mlegrand@pasteur.fr.

Received 22 June 2020
Accepted 9 August 2020
Published 2 September 2020

Candida albicans, often found as a commensal yeast of the human gastrointestinal tract, is also a notorious clinical fungal pathogen. It ranks as the fourth leading cause of nosocomial infections, principally due to the increasing number of immunocompromised individuals (1). Nowadays, *C. albicans* is arguably the pathogenic yeast that is used to the largest extent in order to investigate fungal pathogenesis at the molecular level. As a consequence, many molecular resources have been generated to facilitate the study of *C. albicans* biology, including but not restricted to gene knockout and gene overexpression strain collections (2–4), all requiring *C. albicans* genome editing. Generation of gene knockouts or gene fusions with reporter genes has relied principally on homologous recombination (HR) at target loci using PCR-generated cassettes with 70 to 100 nucleotides of homology to the targeted genomic locus and a lithium acetate-heat shock protocol (5) or an electroporation protocol (6). These strategies allow single-allele disruption, which can quickly become cumbersome in a diploid organism such as *C. albicans*, since multiple transformation rounds are required for the construction of null mutants. Another inconvenience that has emerged over time is the occurrence of nonspecific genome rearrangements in the course of transformation; several studies have reported transformation-associated genome changes, which sometimes impact the phenotypic traits of engineered strains (7–11). The occurrence of loss-of-heterozygosity (LOH) events and aneuploidies has been described previously and can occur on chromosomes other than those targeted by genetic manipulations (12–15). Interestingly, while a short exposure of *C. albicans* to heat during the transformation heat shock process tends to increase aneuploidy events, longer exposure to a milder heat favors the appearance of LOH events (14). Previous work has also shown that aneuploid strains are particularly susceptible to undergoing genomic changes upon transformation (14). As a safeguard, researchers have been advised to test the ploidy of their mutants at various genome locations using tools such as the quantitative PCR (qPCR) ploidy screen (13).

The discovery and adaptation of the CRISPR (clustered regularly interspaced short palindromic repeat)-Cas9 system has revolutionized genome editing in a multitude of organisms. Initially described in *Streptococcus pyogenes*, the CRISPR and its endonuclease Cas9 act as a bacterial defense system against non-self DNA. A single guide RNA (sgRNA) directs the cleaving activity of Cas9 to a specific genomic site. This sgRNA is composed of a recognition motif of 20 bp directly followed by a protospacer-adjacent motif sequence (PAM), composed of the nucleotides NGG (16). Bacterial protection is achieved by recognition and induction of DNA double-strand breaks (DSB) in the non-self DNA (17). The CRISPR-Cas9 technology is the latest of several customizable DNA-binding nucleases that have been engineered to facilitate the introduction of DNA fragments at a specific genomic location thanks to HR upon repair of the induced DNA DSB (18). Its ability to target DSB during genetic manipulations stimulates HR at a specific locus, thus enhancing transformation efficiency over that with classical transformation approaches.

The CRISPR-Cas9 technology has been successfully implemented in mammalian cells, plants, bacteria, and fungi, including *C. albicans*, and continues to be adapted to an increasing number of organisms. In diploid organisms, such as *C. albicans*, this technique has facilitated the construction of null mutants, where both alleles may be modified at once (19). CRISPR-Cas9 has also been successfully used to generate multiple simultaneous knockouts using a unique sgRNA to target multiple genes in *C. albicans* (S. Bachellier-Bassi, personal communication). However, due to its novelty, the repercussions of the active CRISPR-Cas9 system on the biology of *C. albicans* are not yet fully understood. In *Saccharomyces cerevisiae*, the integration and constitutive expression of Cas9 have been associated with cell toxicity and lowered fitness of strains, as illustrated by a lower growth rate. Additionally, the toxicity of constitutive Cas9 expression has been associated with a lower transformant yield (20). To limit these effects, a transient system has been developed in *C. albicans* circumventing constitutive Cas9 and sgRNA expression (21). Another issue is raised by reports of off-target DSB by Cas9 in mammalian cells (22–24). Overall, our understanding of the effects of CRISPR-Cas9

technology in *C. albicans* is limited, especially in terms of its impact on the integrity of the genome.

We had initially set out to build a collection of isogenic strains possessing an LOH reporter system allowing detection of LOH events throughout the genome of *C. albicans*. This system involves the insertion of two fluorescent marker genes in a neutral genomic region on both homologs of a given chromosome, allowing spontaneous LOH events to be detected by monitoring the loss of one of the fluorescent markers using flow cytometry. While we were validating our collection of transformed strains, whole-genome sequencing revealed the presence of genome changes. Here, we present a descriptive study of the genome changes in our collection of transformed *C. albicans* strains, comparing strains constructed using CRISPR-Cas9-free and CRISPR-Cas9-dependent methods. With a total of 57 sequenced strains, our results illustrate and highlight the important mutagenic properties of each transformation protocol both in terms of quantification and in terms of the nature of those genome changes.

Overall, the democratization of whole-genome sequencing allowed us to address the extent of "unwanted" genome changes during strain construction in *C. albicans*. We show that the CRISPR-Cas9-free transformation method is a high inducer of genome changes, often with multiple genome change events within a single constructed strain, in contrast to the CRISPR-Cas9-dependent transformation method, which results in less-frequent concomitant genome changes.

RESULTS

Strain construction and phenotyping. We previously established an LOH reporter system for *C. albicans* whereby the *GFP* gene (encoding green fluorescent protein) is integrated at a given locus on one chromosome and the *BFP* gene (encoding blue fluorescent protein) is integrated on the homologous chromosome at the same locus (25). LOH at this locus leads to the loss of either the *BFP* or the *GFP* gene, and the frequency at which LOH events arise can be quantified using fluorescence-activated cell sorter (FACS) analysis. In an attempt to explore variations in LOH frequency on the eight *C. albicans* chromosomes, we undertook to construct a collection of isogenic strains, each possessing the BFP/GFP LOH reporter system at a distinct genomic locus (see Table S1 in the supplemental material). In addition to the LOH reporter system, a unique barcode sequence was also integrated into each strain in order to permit pool experiments. All strains derive from SN148, a *C. albicans* laboratory reference strain displaying arginine, histidine, uridine, and leucine auxotrophies. This strain was sequentially transformed in order to integrate (i) the BFP/GFP LOH reporter system, (ii) a unique barcode sequence, and (iii) the leucine marker rendering the strains prototrophic. CRISPR-Cas9-free and CRISPR-Cas9-dependent transformation methods were used to generate a total of 57 strains (30 and 27 strains, respectively) (Fig. 1). Both methods are based on homology-directed recombination with exogenous DNA, prepared using PCR and directed to the target genomic locus by the two flanking homology regions (100 bp). Additionally, both the CRISPR-Cas9-free and CRISPR-Cas9-dependent transformation protocols use lithium acetate/polyethylene glycol (PEG) and heat shock treatments to transform the yeast cells. Transformation by electroporation was not considered, since it has been shown to be associated with frequent ectopic integrations of free DNA (6, 26).

All engineered strains underwent basic phenotyping to assess if the transformation process drastically impacted their behavior. The functionality of auxotrophy markers associated with integration cassettes was tested by spot assay, where all strains demonstrated a capacity to grow on media with the dropout amino acid histidine, arginine, uridine, or leucine. Moreover, the functionality of the fluorescent proteins BFP and GFP was validated by fluorescence microscopy and flow cytometry. No strain exhibited any difference in colony morphology from the parental strain SN148 on yeast extract-peptone-dextrose (YPD) or synthetic defined (SD) medium. Additionally, the doubling times of all constructed strains were measured in YPD medium at 30°C (Fig. 2). We observed that the parental SN148 strain possesses a longer doubling time, which

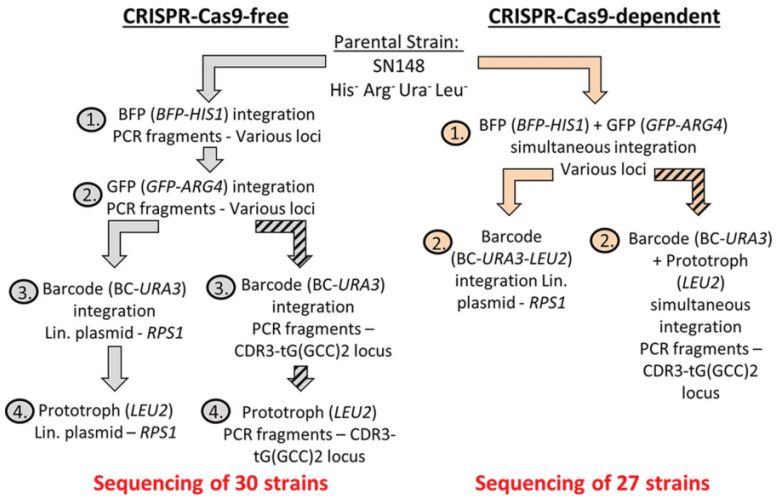

FIG 1 CRISPR-Cas9-free and -dependent transformation strategies used for strain construction. Derived from the *C. albicans* reference strain SC5314, the parental SN148 strain was sequentially transformed using CRISPR-Cas9-free (gray) or CRISPR-Cas9-dependent (orange) strategy in order to integrate (i) the BFP/GFP LOH reporter system, (ii) a unique barcode sequence, and (iii) the leucine marker rendering the strains prototrophic. Strains constructed using the CRISPR-Cas9-free strategy underwent four transformation steps, as opposed to two using the CRISPR-Cas9-dependent strategy. The barcode sequence (BC) and *LEU2* auxotrophic marker are introduced at the *RPS1* locus on Chr1. The BC and *LEU2* marker are introduced at another locus for the strains carrying the LOH reporter system on Chr1 [in the intergenic sequence between the *CDR3* and tG(GCC)2 loci], as indicated by the striped arrows.

is most likely due to uridine auxotrophy (27). Among the strains displaying genomic changes (Fig. 2, yellow symbols), some have doubling times longer than the average doubling time obtained for all 57 strains analyzed in this study.

Genome sequencing. To ensure that the strains did not acquire major genome changes during transformation, we took the opportunity to perform whole-genome sequencing. In view of time and resource constraints, we took advantage of a pipeline available for rapid genome sequencing of our 57 strains that was calibrated to give 50×

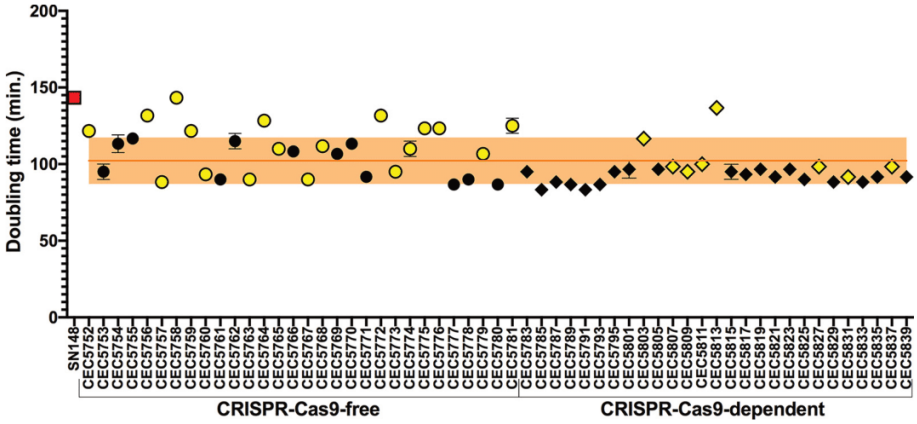

FIG 2 Doubling times in YPD medium at 30°C. Each data point is representative of the average doubling time (in minutes) ($n = 8$) of the parental (SN148) strain (red square) or a constructed strain (circles, CRISPR-Cas9-free method; diamonds, CRISPR-Cas9-dependent method), with error bars indicating standard deviations. The average doubling time of all constructed strains (57 strains) is represented by the orange horizontal line (with orange shading indicating standard deviations). Strains displaying at least one genomic change are shown in yellow, while black indicates strains that are free of transformation-acquired genomic changes.

coverage for a bacterial genome (4.6 Mb). Based on the *C. albicans* genome size, 15× coverage was expected, which would still be sufficient to allow detection of LOH and aneuploidies. An average sequencing depth of 25.5 ± 17.2 (ranging from 7.44× to 73.12×) was obtained upon sequencing of the 57 *C. albicans* strains (Table S4). Postsequencing cleanup and single nucleotide polymorphism (SNP) calling were conducted, and allele ratios at heterozygous sites (ABHet) were evaluated in order to allow identification of genome changes, in particular aneuploidy and LOH. Despite the low sequencing depth, ABHet plots were successfully generated by plotting the allele balance at heterozygous positions across the genome (Fig. S1) and were used to visually identify large-scale LOH events as represented by an absence (or weak presence) of ABHet values. Although less striking in ABHet plots (Fig. S1) due to limited sequencing depth, hyperploidy events (aneuploidy with a chromosome number that is more than the diploid number) were identified by plotting the distribution of ABHet values and calculating the mean ABHet value per chromosome, since hyperploid chromosomes would shift away from the 0.5 ABHet value, which represents a 0.5 allelic ratio (a 50/50 heterozygous nucleotide ratio at the given position). Thus, this method allowed us to differentiate ABHet values corresponding to trisomy (0.33 or 0.66) or tetrasomy (0.25 or 0.75) (Fig. 3A; Table S5). The genomic changes identified throughout the 57 strains are graphically summarized in Fig. 3B.

As illustrated by elevated standard deviations (Table S4), an uneven sequencing depth was obtained across our genomes, probably resulting from library preparations being optimized for bacterial genomes. This prevented us from using sequencing depth data to identify or characterize genome changes, and therefore, we could not distinguish monosomies from diploid whole-chromosome LOH events. Since we relied solely on ABHet analysis to study transformation-induced genome changes, we do acknowledge that we may be underestimating the occurrence of aneuploidy events, notably events resulting in a balance of both haplotypes, such as balanced tetraploidy (2×HapA 2×HapB events), which would also be represented as a 0.5 ratio in the ABHet graphs. Nevertheless, this methodology allowed us to efficiently identify numerous genome changes generated during the transformation process.

Both CRISPR-Cas9-free and CRISPR-Cas9-dependent transformations trigger unwanted genomic changes, but to different extents. Among the 30 strains constructed using the CRISPR-Cas9-free method, a total of 65 genomic changes were identified (Fig. 3B). Only 40% of the strains (12/30) displayed no obvious genomic changes, while the other 60% (18/30) had at least one genomic change, of which half (9/18 strains) possessed three or more genomic changes (Fig. 4A and B). Of interest, 10/18 strains displayed at least one rearrangement on the targeted chromosome, sometimes accompanied by additional changes on other chromosomes. The remaining 8/18 strains displayed changes only on a chromosome(s) not targeted by the transformation process (Fig. 4D). At the chromosome scale, one-quarter of the sequenced chromosomes (60/240) displayed at least one genomic change, including five chromosomes with two identifiable genomic changes (Fig. 4A; Table S6). Although our analysis did not reveal any obvious enrichment of genomic changes on a given chromosome, we did notice slightly higher and lower numbers of changes on chromosome 4 (Chr4) and Chr7, respectively (Fig. 4C). Overall, strains constructed using the CRISPR-Cas9-free method displayed an average of 0.54 genomic change per strain per transformation (Fig. 5A).

Because of the extensive mutagenic effect of CRISPR-Cas9-free transformation, the CRISPR-Cas9-dependent method, via transient Cas9 expression, was favored to generate the second part of our collection, since it reduces the number of sequential transformations (Fig. 1) by targeting DNA DSB on both homologs of the loci of interest. A total of 10 genomic changes, distributed among 9 strains (33.3%), were identified within the 27 strains engineered by this method (Fig. 4A and B). While one-third of the strains (3/9) displayed at least one change on the targeted chromosome, at times accompanied by additional changes on other chromosomes, the others displayed changes only on nontargeted chromosomes (Fig. 4D). At the chromosome scale, only

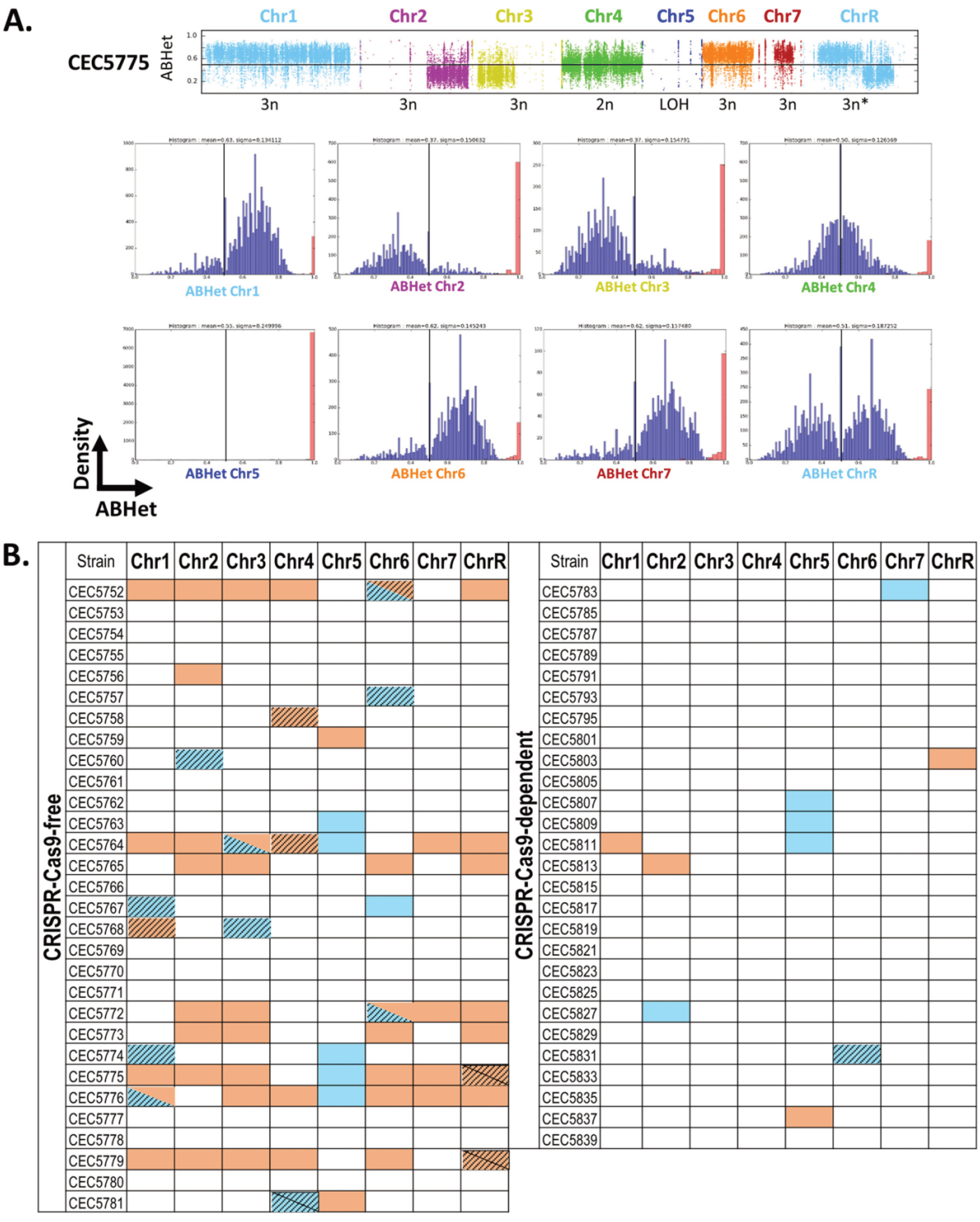

FIG 3 Identification of transformation-induced genome changes within the 57 sequenced *C. albicans* strains. (A) Determining the average ABHet values per chromosome (Chr) for strain CEC5775. (Upper panel) Plots of allele balances for the eight chromosomes of strain CEC5775. (Lower panels) Histograms illustrate the distribution of ABHet values across a given chromosome, where the black vertical bar represents a 0.5 ABHet value (heterozygous diploid). ABHet and

(Continued on next page)

4.6% of the sequenced chromosomes demonstrated genomic changes (10/216), with no chromosomes displaying multiple changes (Fig. 4A; Table S6). These genomic changes were distributed among most chromosomes, apart from Chr3 and Chr4, which, in our hands, were not affected by changes even when targeted. In contrast, Chr5 seemed to exhibit the highest number of genomic changes (Fig. 4C). We estimated that strains constructed using the transient CRISPR-Cas9 method displayed an average of 0.19 genomic change per strain per transformation (Fig. 5A).

Aneuploidies are overrepresented in strains engineered using the CRISPR-Cas9-free method. Genome sequence analysis allowed us to investigate in more detail the nature of the genomic changes observed in the strains engineered using either the CRISPR-Cas9-free or the CRISPR-Cas9-dependent method. The 75 large-scale genomic changes identified above were categorized into two major types: LOH and hyperploidy. Additionally, we report on the sizes of the genome changes, spanning a partial or entire chromosome. Hyperploidies were defined as regions possessing heterozygous SNPs with an allele balance value greater or less than 0.5. For comparison purposes, we took into account the fact that the strain did not go through the same number of transformations in both methods, and we present the data in terms of number of changes per strain per transformation (Fig. 5A).

With the CRISPR-Cas9-free transformation strategy, we observed on average 2.86 times more aneuploidy events (0.40 event) than LOH events (0.14 event) per strain per transformation (*P*, 0.0279 by *t* test). Additionally, both types of genome changes were identified, in decreasing order of abundance, as follows; full chromosome hyperploidies (0.35 event/strain and transformation), partial chromosomal LOH (0.09 event), partial chromosomal hyperploidies (0.05 event), and full-chromosome LOH (0.05 event) (Fig. 5A).

In contrast, strains transformed with the CRISPR-Cas9-dependent transformation strategy displayed comparable average numbers of hyperploidy (0.07 event) and LOH (0.11 event) events per strain per transformation (*P*, 0.493 by *t* test). In decreasing order of abundance, we detected full-chromosome LOH (0.09 event), full-chromosome hyperploidies (0.06 event), partial chromosomal LOH (0.02 event), and partial chromosomal hyperploidy (0.02 event) (Fig. 5A). Our detailed analysis did not reveal any obvious link between the nature of the genomic changes and a specific chromosome, although Chr5 and Chr6 seemed to be implicated in LOH events more frequently than the other chromosomes (Fig. 5B).

DISCUSSION

The construction of a *C. albicans* strain collection, aimed at studying genome-wide LOH dynamics, allowed us to retrospectively compare the mutagenic landscapes of strains engineered using two transformation strategies: a CRISPR-Cas9-free and a CRISPR-Cas9-dependent strategy. Whole-genome sequencing of 57 engineered strains of *C. albicans* permitted the identification and comparison of transformation-acquired genomic changes. Although this study was not designed for this purpose, we observed during the retrospective analysis that *C. albicans* strains engineered using the CRISPR-Cas9-dependent transformation method displayed significantly fewer concomitant genomic changes, notably in terms of hyperploidy events, than strains constructed using the CRISPR-Cas9-free transformation strategy.

As with transformation protocols of other yeast species, *C. albicans* competent cells are prepared and suspended in a lithium acetate solution with single-stranded carrier

FIG 3 Legend (Continued)
ABHom values are shown in blue and red, respectively. Data interpretation for each chromosome is as follows: Chr1, -6, and -7, trisomy (1×HapA, 2×HapB); Chr2 and -3, trisomy (2×HapA, 1×HapB); Chr4, disomy (1×HapA, 1×HapB); Chr5, LOH; ChrR, recombination event localized in proximity of the centromere plus trisomy (left arm, 1×HapA, 2×HapB; right arm, 2×HapA, 1×HapB). Additional LOH events have been described previously in parental strain SN148 (LOH on Chr2) (25) and in SC5314 (LOH on Chr3 and Chr7) (12). (B) Summary of the genomic changes identified across the eight chromosomes for all 57 sequenced *C. albicans* strains, using the strategy presented in panel A. The plots showing the allele balance at heterozygous positions and the mean ABHet values per chromosome for each strain can be found in Fig. S1 and Table S5, respectively. LOH events are indicated in blue, aneuploidies in orange. Genomic changes impacting whole chromosomes are identified by solid colors, while those partially impacting chromosomes are identified by a striped pattern.

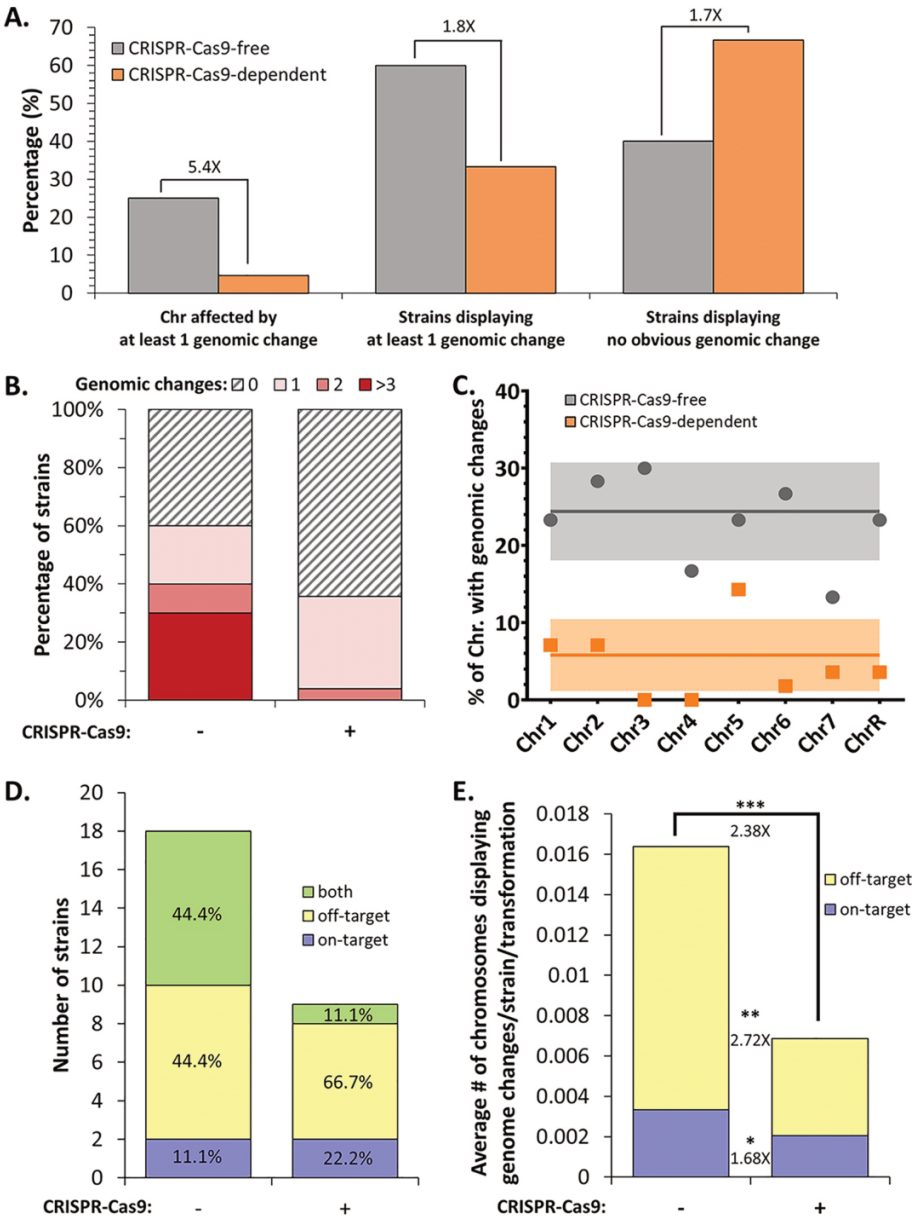

FIG 4 Quantification of genome changes identified within *C. albicans* strains engineered using one of two transformation methods. (A) Percentages of chromosomes and strains impacted by transformation-associated genomic changes, as well as percentages of genomic-change-free strains. (B) Studying concurrent genomic changes within strains. (C) Percentage of each chromosome affected by at least one genomic change. (D) Representations of the numbers of strains displaying genomic change(s) on a targeted chromosome, a nontargeted chromosome, or both types. The fraction of each category of strains within genomic-change-displaying strains is indicated by a percentage. (E) Frequency of transformation-associated genomic changes tabulated as the average number of chromosomes displaying genome changes per strain per transformation. Differences in genome change frequencies between strains constructed using the two transformation strategies are represented as fold changes (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$ by *t* test).

DNA (often single-stranded salmon sperm DNA) and the exogenous DNA, intended to be integrated into the yeast genome. The exogenous DNA usually consists of linearized plasmid DNA with free ends having homology with the yeast genome, or PCR-amplified integration cassettes (5), which are targeted to a specific locus by two homology

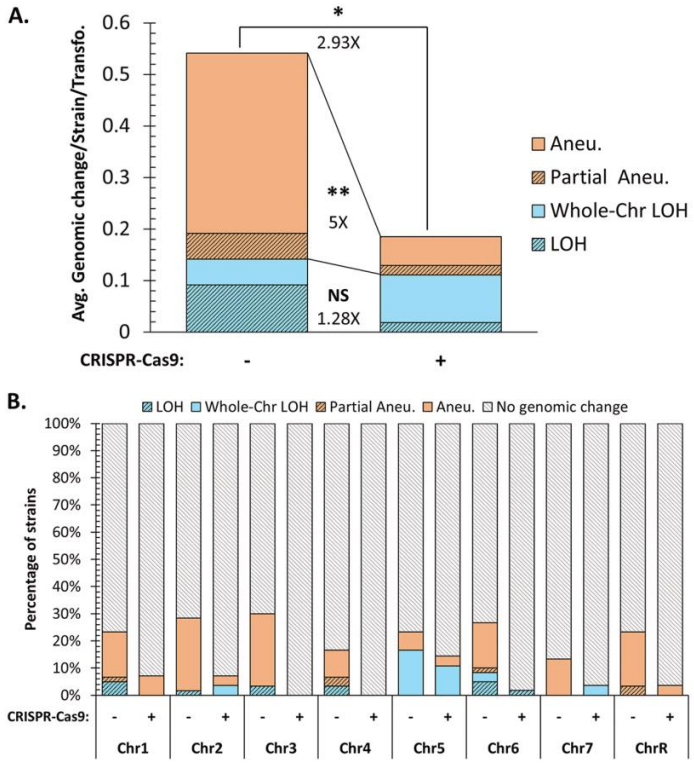

FIG 5 Nature of genomic changes identified within sequenced *C. albicans* strains. (A) Frequency of transformation-associated genomic changes tabulated as the average number of genome changes per strain per transformation. Differences in genomic change frequencies between the two transformation methods are represented as fold changes (P, ≤ 0.05 by t test). (B) Percentage of strains displaying each type of genomic change per chromosome and transformation strategy.

regions (70 bp to 100 bp) flanking the DNA of interest associated with a selection marker. These homology regions on the free DNA allow strand invasion at the genomic homology site, ultimately resulting in the integration of the repair template through recombination events. Studies have shown that DNA DSB in *C. albicans* are repaired predominantly via HR and only rarely by nonhomologous end joining (NHEJ) (28–32). This facilitates genetic manipulations of both laboratory and clinical isolates, since efficient genetic modifications are not restricted to NHEJ-deficient strains (as is the case in numerous fungi [33–35]). However, this technique allows the targeting of a single homolog at one locus, i.e., one integration per transformation process, rendering certain strain constructions cumbersome. The development of molecular tools permitting locus-specific DNA DSB has largely facilitated genome editing in numerous eukaryotic organisms by means of zinc finger nucleases (ZFN) or transcription activator-like effector nucleases (TALEN), which recognize a specific DNA sequence and mediate DNA DSB by the FokI nuclease (18). Nevertheless, the most recent and popular targeted genetic modification system is CRISPR-Cas9, because it is highly versatile and its implementation is less laborious than that of the nucleases mentioned above. A sgRNA guides the Cas9 nuclease to its target sequence, where it induces a DNA break. Promoting DNA DSB at the target locus increases HR, resulting in higher transformation efficiency, and allows the integration of exogenous DNA at the target locus on multiple homologs, facilitating genetic engineering in diploid and polyploid organisms.

We observed that both CRISPR-Cas9-free and CRISPR-Cas9-dependent transformations provoke unwanted genomic changes in *C. albicans*. Our data clearly showed that the transformation process itself is mutagenic, since, by cutting in half the transforma-

tion steps required to engineer our strains, we approximately doubled the number of strains free of unwanted transformation-associated genomic changes (Fig. 4A). Despite differences in the number of transformation steps and constructs (use of one versus two barcode/prototrophy constructs [see Materials and Methods]), both transformation strategies produced genomic changes, though to different extents, since *C. albicans* strains constructed using the CRISPR-Cas9-dependent transformation process displayed obvious reductions in concomitant genomic changes (Fig. 4). Indeed, we observed that 5.4 times fewer chromosomes were affected by undesirable genomic changes among the strains engineered with the CRISPR-Cas9-dependent method. Strains constructed using the CRISPR-Cas9-free strategy displayed multiple coexisting genomic changes; 40% of them carried more than 1 change (up to 8 events), in contrast to 3.7% of strains engineered using the CRISPR-Cas9-dependent method (Fig. 4B). The global mutagenic frequency (average genomic change per strain per transformation) was significantly higher (2.93-fold [$P = 0.0145$]) with the CRISPR-Cas9-free strategy than with the CRISPR-Cas9-dependent method (Fig. 5A). These results suggest that strains engineered by CRISPR-Cas9-free transformations probably derive from cells that endured strong perturbations in genomic stability, thus generating numerous concomitant genomic changes. In contrast, directing DNA DSB with the CRISPR-Cas9-dependent method favors HR-mediated repair and integration of exogenous DNA at the target locus, thus yielding more transformants independently of overall genome perturbation. From a mechanistic perspective, the free linear DNA (repair template) may attempt multiple strand invasions during a homology search, leading to a series of abortive Holliday junctions, perhaps altering overall chromosomal stability, and resulting in large LOH and/or aneuploidies on on- and/or off-target chromosomes. On the other hand, the global mutagenic reduction in strains resulting from CRISPR-Cas9 transformation may potentially be associated with activation of the DNA damage response upon DNA DSB, leading to a global response yielding a cellular environment conducive to repair, including changes in the cell cycle, chromosome mobility, transcription, and nuclear deoxynucleoside triphosphate (dNTP) levels (36).

In other organisms, it has been shown that CRISPR-Cas9 can induce off-target DNA DSB, which may have grave implications (23, 24). To the best of our knowledge, this has not been studied in *C. albicans*. The genomic changes identified within our strain collection are distributed quasi-equally between the eight chromosomes, and LOH and/or hyperploidies are found predominantly on off-target chromosomes in strains derived from both strategies, even though strains engineered using the CRISPR-Cas9-free strategy are always more impacted by genomic changes (Fig. 4E and 5B). Unfortunately, since genomic changes more frequently involved off-target chromosomes (2.38-fold) (Fig. 4E), we cannot rule out the possibility that they resulted from off-target Cas9 DNA DSB in strains that were generated using the CRISPR-Cas9-dependent strategy. Heat shock is a well-known source of DNA breaks, has been associated with the appearance of genomic changes in *C. albicans* (37), and could explain the changes observed in our strain set, which experienced a heat shock during the transformation protocol (15 min, 44°C). Further investigations need to be conducted in order to properly address the potential off-target activity of Cas9 in *C. albicans*.

The genome of *C. albicans* is highly tolerant of genomic rearrangement events, which often arise upon exposure to various stresses, e.g., heat, fluconazole, or oxidative stress (11, 12, 14, 37). Since transformation protocols involve inflicting stress, it comes as no surprise that our study and these previous studies revealed the presence of LOH and aneuploidy events in various laboratory strains. Indeed, microarray experiments and haplotype mapping have highlighted the mutagenic effect of transformation in multiple laboratory backgrounds (RM1000, CAI-4, and BWP17) between strain stocks and their derivatives (12–14, 38). For instance, transformation of SC5314 to generate the uracil auxotrophic CAI-4 strain (39) has been linked to trisomy of Chr1 (10) or Chr2 (11), while *HIS1* disruption in CAI-4 resulted in deletion of telomere-proximal genes of Chr5 in its derivative BWP17 (11, 40). The same was true among our collection of 57 transformed strains, for which we mainly observed two sorts of genomic changes: (i)

LOH and (ii) hyperploidies (defined in our study as a chromosome or a portion of a chromosome present in more than two copies). In our collection, transformation-induced LOH events appeared at comparable rates of 0.14 and 0.11 LOH event/strain/transformation in strains constructed using CRISPR-Cas9-free and CRISPR-Cas9-dependent methods, respectively (1.28-fold change [*P*, 0.5715 by *t* test]). In contrast, strains obtained using the CRISPR-Cas9-dependent strategy displayed a 5-fold decrease (*P*, 0.0086 by *t* test) in hyperploidy events relative to the strains constructed using the CRISPR-Cas9-free method (Fig. 5A). Frequent chromosome trisomies were observed in our strain set (see Tables S5 and S6 in the supplemental material). Chromosome trisomies have often been associated with rapid adaptation to stress conditions, such as that demonstrated by increased fitness in the presence of fluconazole upon Chr5 aneuploidy (41, 42). Additionally, genotypic and phenotypic diversification, namely, Chr5 and Chr6 trisomies, has been described upon exposure of *C. albicans* to the oral niche, leading to a low-virulence phenotype, suggesting an adaptation resulting in a commensal-like phenotype (43, 44).

Taking our findings together, transformation protocols limiting the number of sequential transformation steps should be favored in order to minimize the acquisition of unwanted transformation-associated genomic changes. Additionally, we want to highlight the importance of generating, testing, and analyzing multiple transformants, since, at best, one-third of transformed strains undergo genomic changes (Fig. 4A and B). Though ideal, routine whole-genome sequencing of transformed strains is not necessarily feasible. Alternatives have been proposed by others, such as a qPCR-based assay to monitor copy number variations throughout the genome of *C. albicans* (13). However, this technique is limited to the identification of aneuploidies and does not permit the detection of equally important LOH events. The SNP-restriction fragment length polymorphism (RFLP) technique, allowing one to monitor the allelic status for multiple genomic sites (45), can be a complementary means to detect LOH events. Additionally, recomplementation experiments are key in reverse genetics, allowing one to reinforce the conclusions drawn from phenotypic observations.

To conclude, this study has permitted a detailed investigation of the frequency and nature of genomic changes that occur upon the transformation of *C. albicans* cells, comparing transformation-induced mutagenic landscapes in strains constructed with two transformation methods: a CRISPR-Cas9-free and a CRISPR-Cas9-dependent strategy. By investigating only large genomic rearrangements, we highlight the fact that *C. albicans* transformation is highly mutagenic and recognize that we may be vastly underestimating this mutagenic effect; for example, we did not investigate the presence of indels and SNPs. Nevertheless, the CRISPR-Cas9-dependent strategy seems to reduce transformation-associated concomitant genomic changes, especially with regard to hyperploidy events.

MATERIALS AND METHODS

Strains and culture conditions. The yeast strains described in the study were constructed starting from *C. albicans* strain SN148 (His$^-$ Arg$^-$ Ura$^-$ Leu$^-$) (46). Yeast cells were cultured on/in rich YPD medium (1% yeast extract, 2% peptone, 2% dextrose). synthetic defined (SD) medium (0.67% yeast nitrogen base without amino acids, 2% dextrose), and synthetic complete (SC) medium (0.67% yeast nitrogen base without amino acids, 2% dextrose, 0.08% dropout mix with all the essential amino acids), which were used for selection. Solid media were obtained by adding 2% agar.

Cloning experiments were conducted using One Shot TOP10 chemically competent *Escherichia coli* K-12 cells (Thermo Fisher Scientific). *E. coli* strains were cultured on/in LB (1% Bacto tryptone, 0.5% Bacto yeast extract, 0.5% sodium chloride) or 2YT (1.6% Bacto tryptone, 1% Bacto yeast extract, 0.5% sodium chloride, 0.1% D-glucose) medium, with appropriate antibiotics for selection purposes (50 μg/ml kanamycin, 50 μg/ml ticarcillin). Solid media were obtained by adding 2% agar.

All *C. albicans* strains and *E. coli* plasmids are listed in Tables S1 and S2 in the supplemental material, respectively.

CRISPR-Cas9-free transformation. Strains constructed using the following CRISPR-Cas9-free protocol underwent four sequential heat shock and lithium acetate/PEG rounds of transformations (47) in order to (i and ii) integrate the BFP/GFP LOH reporter system at a distinct genomic locus, (iii) integrate a unique barcode associated with the *URA3* auxotrophic marker at the *RPS1* locus, and finally (iv) integrate the *LEU2* auxotrophic marker at the *RPS1* locus, rendering the strains prototrophic (Fig. 1). The strategy was to integrate the BFP/GFP LOH reporter system into the most telomere-proximal intergenic

region of ≥5 kb on each chromosome arm (Table S1). For this purpose, 120-bp primers were designed, each composed of 20 bp complementary to both the P_{TDH3}-*GFP-ARG4* and P_{TDH3}-*BFP-HIS1* cassettes and 100-bp tails possessing the complementary sequences of the targeted integration locus (Table S3). Each primer pair was used to amplify both the P_{TDH3}-*GFP-ARG4* and P_{TDH3}-*BFP-HIS1* cassettes from plasmids pCRBluntII-P_{TDH3}-*GFP-ARG4* and pCRBluntII-P_{TDH3}-*BFP*-Cd*HIS1*, respectively. Each cassette was amplified in a total PCR volume of 500 μl, precipitated in 100% ethanol, and resuspended in 100 μl of distilled sterile water. For each transformation, competent cells were transformed with approximately 5 μg of the appropriate DNA cassette. The parental SN148 strain was initially transformed with the P_{TDH3}-*GFP-ARG4* cassette and then subjected to a second round of transformation with the P_{TDH3}-*BFP-HIS1* cassette. These two transformation steps allowed the integration of the BFP/GFP LOH reporter system at a given intergenic locus (Table S1). The resulting strains, except those possessing the BFP/GFP LOH reporter system on Chr1, were then retransformed with StuI-linearized Clp10-P_{TET}-BC-*URA3* plasmids, each containing a unique barcode (BC) sequence and targeting the *C. albicans RPS1* locus on Chr1. These plasmids are derived from a private laboratory collection of Cip10-P_{TET}-BC-GTW-*URA3* plasmids (unpublished data), each possessing a unique 25-nucleotide barcode sequence. The Gateway (GTW) cassette was removed by HindIII digestion and self-ligation in order to ensure that its presence did not influence the biology of *C. albicans*. Last, BFP/GFP barcoded strains were rendered prototrophic by a fourth round of transformation involving the integration of the StuI-linearized Clp10-*LEU2* plasmid at the *RPS1* locus. Conversely, in strains bearing the BFP/GFP LOH reporter system on Chr1, the BC-*URA3* and *LEU2* cassettes were integrated on Chr4 at the CDR3-tG(GCC)2 locus to avoid loss of the latter upon LOH. For these strains, the BC-*URA3* and *LEU2* cassettes were generated using the same PCR amplification protocol described above for the P_{TDH3}-*GFP-ARG4* and P_{TDH3}-*BFP-HIS1* cassettes, where long-tailed primers (Table S3) were used for amplification from plasmids Clp10-P_{TET}-BC-*URA3* and Clp10-*LEU2*, respectively. Throughout the strain construction process, selective pressure was always maintained in order to ensure the selection of transformants carrying all integration cassettes. In addition, at each transformation step, junction PCRs were conducted to ensure the proper integration of cassettes using the primers listed in Table S3.

CRISPR-Cas9-dependent transformation. In contrast to the CRISPR-Cas9-free transformation method, both homologs may be simultaneously targeted for cassette integration with the CRISPR-Cas9-dependent transformation protocol. Thus, by directing a DNA DSB with a locus-specific sgRNA, the BFP/GFP LOH reporter system can be introduced with only one exposure to heat shock and lithium acetate/PEG, rather than two treatments. Hence, only two transformation rounds were required for strain construction, where (i) the BFP/GFP LOH reporter system was integrated at distinct genomic loci and (ii) strains were simultaneously barcoded and rendered prototrophic using both *URA3* and *LEU2* auxotrophic markers (Fig. 1). A total of 11 unique 20-bp sgRNAs were designed using CHOPCHOP (48), targeting the same integration loci as those chosen in the CRISPR-Cas9-free protocol for integration of the BFP/GFP LOH reporter system (Table S3). Because the Clp10-derived BC-*URA3-LEU2* plasmids could not be targeted to the *RPS1* locus on Chr1 in the strains bearing the BFP/GFP LOH reporter system on Chr1, an additional sgRNA was designed to target the BC-*URA3* and *LEU2* markers on Chr4 at the CDR3-tG(GCC)2 locus (Table S3) in these strains.

We used a transient CRISPR-Cas9 system (21), which does not necessitate the genomic integration of either Cas9 or sgRNAs. The construction of sgRNAs and the amplification of Cas9 cassettes from the pV1093 plasmid were conducted as described by Min et al. (21), while *BFP-HIS1*-, *GFP-ARG4*-, BC-*URA3*-, and *LEU2*-bearing cassettes were constructed as described above. SN148 cells were cotransformed with 3 μg of the P_{TDH3}-*GFP-ARG4* cassette, 3 μg of the P_{TDH3}-*BFP-HIS1* cassette, 1 μg of the Cas9 cassette, and 1 μg of sgRNA using the lithium acetate/PEG transformation protocol. Transformants were then selected on SC-Arg-His medium, and junction PCRs were performed in order to ensure proper integration of both cassettes at the targeted locus. Strains bearing the BFP/GFP LOH reporter system on Chr1 were then transformed using the transient CRISPR-Cas9 system targeting BC-*URA3* and *LEU2* cassette integration at the CDR3-tG(GCC)2 locus on Chr4. For the remaining strains, both BC-*URA3* and *LEU2* markers were integrated in one transformation step. We did this using the Gateway recombination system, where the *LEU2* gene was transferred from a pDONR-*LEU2* plasmid into Cip10-P_{TET}-BC-GTW-*URA3* plasmids (unpublished data) by an LR reaction. The unique Cip10-P_{TET}-BC-*LEU2-URA3* plasmids were then linearized by StuI and integrated at the *RPS1* locus. These transformants were selected on SD medium, and junction PCRs were performed.

Strain phenotyping. All selected strains underwent basic phenotypic characterization upon validation of cassette integration at targeted loci by junction PCRs. The functionality of the auxotrophic markers was evaluated by drop tests on SC medium with the appropriate dropout amino acid mix, based on the marker tested. Overnight-saturated cultures of selected strains in liquid YPD medium were spotted onto solid YPD, SC-His, SC-Arg, SC-Ura, and SC-Leu media and were placed at 30°C for 24 h to monitor growth. Furthermore, the functionality/intensity of both fluorescent proteins (BFP and GFP) was validated by flow cytometry (MACSQuant analyzer [Miltenyi Biotec]) and fluorescence microscopy (Olympus IX83). The colony morphology of all strains was also assessed on both solid YPD and SD media at 30°C. Finally, doubling times were evaluated in liquid YPD medium at 30°C by measuring the optical density with a Tecan Infinite system over a 24-h period.

DNA extraction and whole-genome sequencing of strains. Prototrophic strains were cultured in 5 ml of liquid SD medium overnight at 30°C, and DNA was extracted by following the manufacturer's protocol using the Qiagen QIAamp DNA minikit. The DNA was eluted in a total volume of 100 μl. The genomes were sequenced at the P2M Platform of Institut Pasteur by using the Illumina sequencing

technology. Libraries were prepared with the Nextera XT sequencing kit, and NextSeq500 platforms were used to generate 151-bp paired-end reads.

Identification of gross chromosomal rearrangements. Sequences and genomic variations were analyzed as described in references 49 and 50. Each set of paired-end reads was mapped against the *C. albicans* reference genome, SC5314 haplotype A and haplotype B (version A22-s07-m01-r57), using Minimap2 (51). SAMtools, version 1.9, and Picard tools, version 2.8.1 (http://broadinstitute.github.io/picard), were then used to filter, sort, and convert SAM files. SNPs were called using the Genome Analysis Toolkit (GATK), version 3.6, according to GATK best practices. SNPs were filtered using the following parameters: VariantFiltration, QD $<$ 2.0, LowQD, ReadPosRankSum $<$ −8.0, LowRankSum, FS $>$ 60.0, HightFS, MQRankSum $<$ −12.5, MQRankSum, MQ $<$ 40.0, LowMQ, HaplotypeScore $>$ 13.0. Sequencing depths were also calculated using the Genome Analysis Toolkit (Table S4). The GATK variant filtration walker (VariantAnnotator) was used to add allele balance information to VCF files. The value of allele balance at heterozygous sites (ABHet) is a number that varies between 0 and 1. ABHet is calculated as the number of reference reads from individuals with heterozygous genotypes divided by the total number of reads from such individuals. Thus, a diploid genome will be defined by an ABHet value of 0.5. In contrast, while a triploid strain may contain either three identical alleles (an allelic frequency of 1) or two identical alleles and one different allele (frequencies of 0.66 and 0.33), a tetraploid strain may have allelic frequencies of either 0.5 (2 $\times$ 2 identical alleles), 1 (4 identical alleles), or 0.25 and 0.75 (3 identical alleles and 1 different allele). In order to obtain an average ABHet value per chromosome, we evaluated the ABHet and allele balance at homozygous positions (ABHom) with the AlleleBalance annotation GATK module (Table S5). Histograms were built based on the number of SNPs with ABHet values with the matplotlib 2D graphics package (52), with blue and red representing ABHet and ABHom values, respectively.

Data availability. Genome sequences of the 57 engineered *C. albicans* isolates described in this study have been deposited in the NCBI Sequence Read Archive under BioProject ID PRJNA659611. All other relevant data are available from the corresponding author upon request.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FIG S1, JPG file, 2.3 MB.

TABLE S1, XLSX file, 0.01 MB.

TABLE S2, XLSX file, 0.01 MB.

TABLE S3, XLSX file, 0.02 MB.

TABLE S4, XLSX file, 0.01 MB.

TABLE S5, XLSX file, 0.01 MB.

TABLE S6, XLSX file, 0.05 MB.

ACKNOWLEDGMENTS

We are grateful to the members of the Institut Pasteur Mutualized Platform for Microbiology (P2M) for Illumina sequencing.

T.M., M.L., and C.D. designed experiments. T.M. performed experiments. T.M., C.M., M.L., and C.D. analyzed data. T.M., M.L., and C.D. wrote the manuscript. All authors read and approved the final manuscript.

T.M. is the recipient of a Ph.D. fellowship from the Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (ANR-10-LABX-62-IBEID). We acknowledge support from the French Government's Investissement d'Avenir program (Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases [ANR10-LABX-62-IBEID]).

REFERENCES

1. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. 2012. Hidden killers: human fungal infections. Sci Transl Med 4:165rv13. https://doi.org/10.1126/scitranslmed.3004404.
2. Homann OR, Dea J, Noble SM, Johnson AD. 2009. A phenotypic profile of the Candida albicans regulatory network. PLoS Genet 5:e1000783. https://doi.org/10.1371/journal.pgen.1000783.
3. Schoeters F, Munro CA, D'Enfert C, Van Dijck P. 2018. A high-throughput Candida albicans two-hybrid system. mSphere 3:e00391-18. https://doi.org/10.1128/mSphere.00391-18.
4. Legrand M, Bachellier-Bassi S, Lee KK, Chaudhari Y, Tournu H, Arbogast L, Boyer H, Chauvel M, Cabral V, Maufrais C, Nesseir A, Maslanka I, Permal E, Rossignol T, Walker LA, Zeidler U, Znaidi S, Schoeters F, Majgier C, Julien RA, Ma L, Tichit M, Bouchier C, Dijck PV, Munro CA, D'Enfert C. 2018. Generating genomic platforms to study Candida albicans pathogenesis. Nucleic Acids Res 46:6935–6949. https://doi.org/10.1093/nar/gky594.
5. Gola S, Martin R, Walther A, Dünkler A, Wendland J. 2003. New modules for PCR-based gene targeting in Candida albicans: rapid and efficient gene targeting using 100 bp of flanking homology region. Yeast 20:1339–1347. https://doi.org/10.1002/yea.1044.
6. De Backer MD, Maes D, Vandoninck S, Logghe M, Contreras R, Luyten WH. 1999. Transformation of Candida albicans by electroporation. Yeast 15:1609–1618. https://doi.org/10.1002/(SICI)1097-0061(199911)15:15<1609::AID-YEA485>3.0.CO;2-Y.
7. Sanglard D, Hube B, Monod M, Odds FC, Gow NAR. 1997. A triple deletion of the secreted aspartyl proteinase genes SAP4, SAP5, and SAP6 of Candida albicans causes attenuated virulence. Infect Immun 65:3539–3546. https://doi.org/10.1128/IAI.65.9.3539-3546.1997.

8. Dunkel N, Morschhäuser J. 2011. Loss of heterozygosity at an unlinked genomic locus is responsible for the phenotype of a Candida albicans sap4Δ sap5Δ sap6Δ mutant. Eukaryot Cell 10:54–62. https://doi.org/10.1128/EC.00281-10.
9. Ciudad T, Hickman M, Bellido A, Berman J, Larriba G. 2016. Phenotypic consequences of a spontaneous loss of heterozygosity in a common laboratory strain of Candida albicans. Genetics 203:1161–1176. https://doi.org/10.1534/genetics.116.189274.
10. Chen X, Magee BB, Dawson D, Magee PT, Kumamoto CA. 2004. Chromosome 1 trisomy compromises the virulence of Candida albicans. Mol Microbiol 51:551–565. https://doi.org/10.1046/j.1365-2958.2003.03852.x.
11. Selmecki A, Bergmann S, Berman J. 2005. Comparative genome hybridization reveals widespread aneuploidy in Candida albicans laboratory strains. Mol Microbiol 55:1553–1565. https://doi.org/10.1111/j.1365-2958.2005.04492.x.
12. 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 (Bethesda) 1:523–530. https://doi.org/10.1534/g3.111.000885.
13. Arbour M, Epp E, Hogues H, Sellam A, Lacroix C, Rauceo J, Mitchell A, Whiteway M, Nantel A. 2009. Widespread occurrence of chromosomal aneuploidy following the routine production of Candida albicans mutants. FEMS Yeast Res 9:1070–1077. https://doi.org/10.1111/j.1567-1364.2009.00563.x.
14. Bouchonville K, Forche A, Tang KES, 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.
15. Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH, Marton MJ. 2000. Widespread aneuploidy revealed by DNA microarray expression profiling. Nat Genet 25:333–337. https://doi.org/10.1038/77116.
16. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–821. https://doi.org/10.1126/science.1225829.
17. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712. https://doi.org/10.1126/science.1138140.
18. Wood AJ, Lo T-W, Zeitler B, Pickle CS, Ralston EJ, Lee AH, Amora R, Miller JC, Leung E, Meng X, Zhang L, Rebar EJ, Gregory PD, Urnov FD, Meyer BJ. 2011. Targeted genome editing across species using ZFNs and TALENs. Science 333:307–307. https://doi.org/10.1126/science.1207773.
19. Vyas VK, Barrasa MI, Fink GR. 2015. A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. Sci Adv 1:e1500248. https://doi.org/10.1126/sciadv.1500248.
20. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. 2013. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343. https://doi.org/10.1093/nar/gkt135.
21. Min K, Ichikawa Y, Woolford CA, Mitchell AP. 2016. Candida albicans gene deletion with a transient CRISPR-Cas9 system. mSphere 1:e00130-16. https://doi.org/10.1128/mSphere.00130-16.
22. Tsai SQ, Zheng Z, Nguyen NT, Liebers M, Topkar VV, Thapar V, Wyvekens N, Khayter C, Iafrate AJ, Le LP, Aryee MJ, Joung JK. 2015. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol 33:187–197. https://doi.org/10.1038/nbt.3117.
23. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotechnol 31:822–826. https://doi.org/10.1038/nbt.2623.
24. Banakar R, Eggenberger AL, Lee K, Wright DA, Murugan K, Zarecor S, Lawrence-Dill CJ, Sashital DG, Wang K. 2019. High-frequency random DNA insertions upon co-delivery of CRISPR-Cas9 ribonucleoprotein and selectable marker plasmid in rice. Sci Rep 9:19902. https://doi.org/10.1038/s41598-019-55681-y.
25. Loll-Krippleber R, Feri A, Nguyen M, Maufrais C, Yansouni J, d'Enfert C, Legrand M. 2015. A FACS-optimized screen identifies regulators of genome stability in Candida albicans. Eukaryot Cell 14:311–322. https://doi.org/10.1128/EC.00286-14.
26. Bachellier-Bassi S, d'Enfert C. 2015. Chemical transformation of Candida albicans, p 81–85. *In* van den Berg MA, Maruthachalam K (ed), Genetic transformation systems in fungi, vol 1. Springer International Publishing, Cham, Switzerland.
27. Kirsch DR, Whitney RR. 1991. Pathogenicity of Candida albicans auxotrophic mutants in experimental infections. Infect Immun 59:3297–3300. https://doi.org/10.1128/IAI.59.9.3297-3300.1991.
28. Feri A, Loll-Krippleber R, Commere P-H, Maufrais C, Sertour N, Schwartz K, Sherlock G, Bougnoux M-E, D'Enfert C, Legrand M. 2016. Analysis of repair mechanisms following an induced double-strand break uncovers recessive deleterious alleles in the Candida albicans diploid genome. mBio 7:e01109-16. https://doi.org/10.1128/mBio.01109-16.
29. Vyas VK, Bushkin GG, Bernstein DA, Getz MA, Sewastianik M, Barrasa MI, Bartel DP, Fink GR. 2018. New CRISPR mutagenesis strategies reveal variation in repair mechanisms among fungi. mSphere 3:e00154-18. https://doi.org/10.1128/mSphere.00154-18.
30. Legrand M, Chan CL, Jauert PA, Kirkpatrick DT. 2007. Role of DNA mismatch repair and double-strand break repair in genome stability and antifungal drug resistance in Candida albicans. Eukaryot Cell 6:2194–2205. https://doi.org/10.1128/EC.00299-07.
31. Andaluz E, Calderone R, Reyes G, Larriba G. 2001. Phenotypic analysis and virulence of Candida albicans LIG4 mutants. Infect Immun 69:137–147. https://doi.org/10.1128/IAI.69.01.137-147.2001.
32. Ciudad T, Andaluz E, Steinberg-Neifach O, Lue NF, Gow NAR, Calderone RA, Larriba G. 2004. Homologous recombination in Candida albicans: role of CaRad52p in DNA repair, integration of linear DNA fragments and telomere length. Mol Microbiol 53:1177–1194. https://doi.org/10.1111/j.1365-2958.2004.04197.x.
33. Choquer M, Robin G, Le Pecheur P, Giraud C, Levis C, Viaud M. 2008. Ku70 or Ku80 deficiencies in the fungus Botrytis cinerea facilitate targeting of genes that are hard to knock out in a wild-type context. FEMS Microbiol Lett 289:225–232. https://doi.org/10.1111/j.1574-6968.2008.01388.x.
34. Goins CL, Gerik KJ, Lodge JK. 2006. Improvements to gene deletion in the fungal pathogen Cryptococcus neoformans: absence of Ku proteins increases homologous recombination, and co-transformation of independent DNA molecules allows rapid complementation of deletion phenotypes. Fungal Genet Biol 43:531–544. https://doi.org/10.1016/j.fgb.2006.02.007.
35. Krappmann S, Sasse C, Braus GH. 2006. Gene targeting in Aspergillus fumigatus by homologous recombination is facilitated in a nonhomologous end-joining-deficient genetic background. Eukaryot Cell 5:212–215. https://doi.org/10.1128/EC.5.1.212-215.2006.
36. Sirbu BM, Cortez D. 2013. DNA damage response: three levels of DNA repair regulation. Cold Spring Harb Perspect Biol 5:a012724. https://doi.org/10.1101/cshperspect.a012724.
37. Forche A, Abbey D, Pisithkul T, Weinzierl MA, Ringstrom T, Bruck D, Petersen K, Berman J. 2011. Stress alters rates and types of loss of heterozygosity in Candida albicans. mBio 2:e00129-11. https://doi.org/10.1128/mBio.00129-11.
38. Ahmad A, Kabir MA, Kravets A, Andaluz E, Larriba G, Rustchenko E. 2008. Chromosome instability and unusual features of some widely used strains of Candida albicans. Yeast 25:433–448. https://doi.org/10.1002/yea.1597.
39. Fonzi WA, Irwin MY. 1993. Isogenic strain construction and gene mapping in Candida albicans. Genetics 134:717–728.
40. Pla J, Pérez-Diaz RM, Navarro-Garcia F, Sánchez M, Nombela C. 1995. Cloning of the Candida albicans HIS1 gene by direct complementation of a C. albicans histidine auxotroph using an improved double-ARS shuttle vector. Gene 165:115–120. https://doi.org/10.1016/0378-1119(95)00492-O.
41. 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.
42. Selmecki AM, Dulmage K, Cowen LE, Anderson JB, 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.
43. Forche A, Solis NV, Swidergall M, Thomas R, Guyer A, Beach A, Cromie GA, Le GT, Lowell E, Pavelka N, Berman J, Dudley AM, Selmecki A, Filler SG. 2019. Selection of Candida albicans trisomy during oropharyngeal infection results in a commensal-like phenotype. PLoS Genet 15:e1008137. https://doi.org/10.1371/journal.pgen.1008137.
44. Forche A, Cromie G, Gerstein AC, Solis NV, Pisithkul T, Srifa W, Jeffery E, Abbey D, Filler SG, Dudley AM, Berman J. 2018. Rapid phenotypic and genotypic diversification after exposure to the oral host niche in Can-

dida albicans. Genetics 209:725–741. https://doi.org/10.1534/genetics.118.301019.
45. Forche A, Steinbach M, Berman J. 2009. Efficient and rapid identification of Candida albicans allelic status using SNP-RFLP. FEMS Yeast Res 9:1061–1069. https://doi.org/10.1111/j.1567-1364.2009.00542.x.
46. Noble SM, Johnson AD. 2005. Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen Candida albicans. Eukaryot Cell 4:298–309. https://doi.org/10.1128/EC.4.2.298-309.2005.
47. Walther A, Wendland J. 2003. An improved transformation protocol for the human fungal pathogen Candida albicans. Curr Genet 42:339–343. https://doi.org/10.1007/s00294-002-0349-0.
48. Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. 2014. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res 42:W401–W407. https://doi.org/10.1093/nar/gku410.
49. Ropars J, Maufrais C, Diogo D, Marcet-Houben M, Perin A, Sertour N, Mosca K, Permal E, Laval G, Bouchier C, Ma L, Schwartz K, Voelz K, May RC, Poulain J, Battail C, Wincker P, Borman AM, Chowdhary A, Fan S, Kim SH, Le Pape P, Romeo O, Shin JH, Gabaldon T, Sherlock G, Bougnoux M-E, D'Enfert C. 2018. Gene flow contributes to diversification of the major fungal pathogen Candida albicans. Nat Commun 9:2253. https://doi.org/10.1038/s41467-018-04787-4.
50. Sitterle E, Maufrais C, Sertour N, Palayret M, d'Enfert C, Bougnoux M-E, Sitterlé E, Maufrais C, Sertour N, Palayret M, d'Enfert C, Bougnoux M-E. 2019. Within-host genomic diversity of Candida albicans in healthy carriers. Sci Rep 9:2563. https://doi.org/10.1038/s41598-019-38768-4.
51. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
52. Hunter JD. 2007. Matplotlib: a 2D graphics environment. Comput Sci Eng 9:90–95. https://doi.org/10.1109/MCSE.2007.55.

Summary

Candida albicans is an opportunistic human pathogen possessing a relatively heterozygous diploid genome that is highly tolerant to particular genomic rearrangements, namely loss-of-heterozygosity (LOH). These two genomic hallmarks are thought to participate in the successful adaptation to its fluctuating host-dependent environment. In this thesis, we **first** made an inventory of LOH events within a collection of clinical *C. albicans* isolates and highlighted that LOH are principally short in size, which permits to generate new allelic combinations while maintaining relatively high overall heterozygosity level. **Secondly**, we identified recessive lethal alleles which constrains the directionality of these LOH events, and demonstrated that such constrains could be alleviated by recombinogenic properties of repeat sequences. A major source of LOH are DNA double-strand breaks (DSB), which have been previously shown to primarily be repaired through homologous recombination-mediated repair pathways in *C. albicans*, resulting in various lengths of LOH. **Thirdly**, we focused on break-induced replication (BIR) in *C. albicans*, which leads to long-tract LOH, spanning numerous kilobases until the telomere. Through characterization of different genes involved in DNA repair, we found that BIR was unexpectedly associated with homozygosis tracts between the break site and the centromere caused by intrinsic features of BIR and the activity of the mismatch repair pathway on natural heterozygous positions. **Fourth**, because strains often display numerous concomitant genomic rearrangements upon use of a classical transformation protocol, we investigated the recently developed CRISPR-Cas9 protocol and showed that this method seems to reduce simultaneous rearrangements, namely aneuploidy events. Although LOH are pervasive and implicated in various aspect of *C. albicans* biology, including acquisition of adaptive phenotypes, little is known regarding genome-wide LOH dynamics. **Fifth**, using a molecular approach, we highlighted elevated levels of heterogeneity in terms of genome stability, LOH frequency, at several levels, inter-strain, intra-strain and inter-chromosomes in basal and stress conditions. The presence of such heterogeneity participates in generating a wide-spectrum of genome instability potential within a population, perhaps facilitating the generation of new allelic combinations, a powerful adaptive strategy in a predominantly clonal organism such as *C. albicans*, which is continuously exposed to an evolving environment.

Key words: *Candida albicans,* Loss-of-heterozygosity, Genome stability, Heterogeneity