Thèse de doctorat
de l'Université Sorbonne Paris Cité
Préparée à l'Université Paris Diderot
Ecole doctorale BioSPC ED157
Unité de recherche: Biologie et Pathogénicité Fongiques

Molecular mechanisms and signaling pathways involved in genome stability in the human fungal pathogen, *Candida albicans*

Par Adeline FERI

Thèse de doctorat d'Infectiologie - Microbiologie

Dirigée par Marie-Elisabeth Bougnoux et Mélanie Legrand

Présentée et soutenue publiquement à Paris, le lundi 26 septembre 2016

Président du jury : Philippe SILAR, Professeur des Universités, Université Paris Diderot
Rapporteur : Cécile FAIRHEAD, Professeur des Universités, Université Paris Sud
Rapporteur : Timothy HUMPHREY, Chef de laboratoire, Université d'Oxford, Royaume Uni
Examinateur : Alix COSTE, Chargée de recherche, Institut de Microbiologie, Suisse
Directrice de thèse : Marie-Elisabeth BOUGNOUX, MCU-PH, Université Paris Descartes
Membre invité : Mélanie LEGRAND, Chargée de recherche, Institut Pasteur

Acknowledgements

I would like to start this section by thanking Cécile Fairhead, Timothy Humphrey, Alix Coste and Philippe Silar for agreeing to be part of this thesis jury. This is the final step of this thesis and your presence means a lot to me. In particular, I thank Cécile and Timothy who helped in the improvement of this manuscript.

I would like to thank Christophe for giving me the chance to be a student in his lab, accompanying me in this adventure all along these years, sharing his knowledge and allocating a lot of his time everytime I needed it. I thank Mélanie for her advices and for giving the freedom that I needed to grow both personally and professionally and Raphaël, for teaching me how to work in the lab, the different techniques required to succeed, and to be there for me when I was still a baby in this lab. I would like also to thank Marie-Elisabeth for giving me her trust by agreeing to be my thesis director.

I have spent almost 4 years in this lab, and I really enjoyed it. However, this would never be as great as it is if I did not have the chance to share happiness and laughter with my present and former labmates and especially, Raphaël and Vitor with whom I had the chance to share hilarious moments; Sophie for her constant availability and above all our numerous moments of uncontrollable laughter during conferences (shhhh this is a secret!!) and other moments; Murielle with whom I could express my devilish side and be always helped and supported (in her own way =)) and with whom I could share unforgettable moments (in spite of my bad memory); Emilie, Natacha and Anne who encouraged me and with whom I had pleasure to have long conversations, Fred for her smile and delicious chocolate cakes; Arturo, Virginia and Sadri for the fun moments we had and notably during meetings; Corinne for her help with my best enemy: bioinformatics and finally, short term students as Aurélie, Hélène, Jérôme, Léa, Alexandra and Valentine, I was happy to meet and share great moments with them. Outside the lab, I was lucky to get to know Haser that I would like to thank for his friendship and Laetitia PJ, my favorite southwestern girl, with whom I had great time and great laughter.
Overall, I think I can say that beyond the work, this experience offered me new friends.

I do also want to thank Guilhem and Guy-Franck for their help and support all over my thesis, Pierre-Henri for his kindness as well as Lucia and Monique for helping me with the large amount of media that I needed.

Outside the doors of Pasteur I have wonderful friends. Alba, my other half, Delphine, my former “coloc”, my little Hind, Tharsha (also known as Thor) and Aurélien that I have the chance to have in my life for 6 years and my friends from Marseille, Laetitia F and Pauline, I thank you all for being you and being there for me.

J’aimerais à présent remercier, Denis, celui qui partage ma vie et qui a su me supporter pendant tout ce temps, malgré mes périodes de doutes et de stress ! Je suis heureuse que tu m’accompagnes dans chaque étape de ma vie.
Enfin, je souhaite remercier ma famille et surtout mes parents. Ils m’ont toujours soutenu et m’ont tout donné pour être heureuse et réussir. Merci, c’est grâce à vous tous que je suis ici.

You all played a major role in this thesis, and I sincerely thank you for that!!

Contents

Figures

Tables and supplemental materials

Lists of abbreviations

5-FOA	5-fluoro orotic acid
2-DG	2-deoxygalactose
Alt-NHEJ	Alternative-NHEJ
BER	Base Excision Repair
BFP	Blue Fluorescent Protein
BIR	Break-induced replication
CDE	Centromeric DNA element
CEN	Centromere
CENP	Centromeric protein
CGH	Comparative Genomic Hybridization
CHEF	Contour-clamped Homogenous Electric Field DSB
Chr	Chromosome
C-NHEJ	Classical or canonical-NHEJ
CNV	Copy Number Variation
CO	Crossover
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dHJ	double Holliday junction
dNTP	desoxyribosenucleotide triphosphate
DSB	Double-strand break
FACS	Fluorescence-activated cell sorting
GC	Gene conversion
GFP	Green Fluorescent Protein
GG-NER	Global genome-NER
HDR	Homology-directed repair
HR	Homologous recombination
HU	Hydroxyurea
LOH	Loss-of-heterozygosity

LN-BER	Long-patch BER
LTR	Long-terminal repeats
MAT	Mating Type
MCO	Mitotic crossover
MMBIR	Microhomology-mediated BIR
MMEJ	Microhomology mediated end-joining
MMR	Mismatch Repair
MMS	methyl methanesulfonate
MRN	Mre11-Rad50-Nbs1
MRS	Major repeat sequence
MRX	Mre11-Rad50-Xrs2
MTL	Mating Type-Like
NER	Nucleotide Excision Repair
NHEJ	Non-homologous end-joining
PARP1	Poly(ADP)-ribose polymerase 1
PCNA	Proliferating Cell Nuclear Antigen
PFGE	Pulse-field gel electrophoresis
rDNA	ribosomal DNA
RER	Ribonucleotide Excision Repair
RF-C	Replication Factor C
rNMP	ribonucleoside monophosphate
RPA	Replication Protein A
RPS	Repeated sequences
SC	Synthetic complete
SCL	Segmental Chromosome Loss
SDL	Synthetic Dosage Lethality
SDSA	Synthesis-dependent strand-annealing
SGA	Synthetic Genetic Array
SN-BER	Short-patch BER
SSA	Single strand annealing
SNP	Single Nucleotide Polymorphism
SNP-RFLP	Single Nucleotide Polymorphism - Restriction Fragment Length Polymorphism
TALEN	Transcription Activator Like Effector Nuclease
TC-NER	Transcription-coupled NER
TER	RNA template
TERT	Reverse transcriptase
TET	Tetracycline
WCL	Whole Chromosome Loss
WGS	Whole Genome Sequencing
YPD	Yeast extract Peptone Dextrose

List of publications

Feri A, Loll-Krippleber R, Commere PH, Maufrais C, Sertour N, Schwartz K, Sherlock G, Bougnoux ME, d'Enfert C and Legrand M. *(accepted for publication)*. Analysis of repair mechanisms following an induced double strand break uncovers recessive deleterious alleles in the *Candida albicans* diploid genome. mBio

d'Enfert C, Bougnoux ME, Feri A, Legrand M, Loll-Krippleber R, Marton T, Maufrais C, Ropars J, Sertour N and Sitterlé E. (2016). Chapter genome diversity and dynamics in *Candida albicans*. *Candida albicans*: cellular and molecular biology. Springer. Prasad R.

Loll-Krippleber R, Feri A, Nguyen M, Maufrais C, Yansouni J, d'Enfert C and Legrand M. (2015). A FACS-Optimized Screen Identifies Regulators of Genome Stability in *Candida albicans*. Eukaryot Cell **14:**311-322

Loll-Krippleber R, d'Enfert C, Feri A, Diogo D, Perin A, Marcet-Houben M, Bougnoux ME and Legrand M. (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.

Loll-Krippleber R, Feri A, d'Enfert C and Legrand M. (2014). Chapter genome integrity: mechanisms and contribution to antifungal resistance. Antifungals: from genomics to resistance and the development of novel agents. Caister Academic Press. Coste A and Vandeputte P.

Foreword

One of the most fascinating facts about living organisms is how everything is made to favor survival. The most representative example is the preservation of the DNA molecule integrity. This molecule carries the genetic sources necessary to the living and has to be conserved to ensure life. Indeed, DNA replication is highly accurate but copying over millions of bases in a little timeframe is at the origin of mistakes, generating either mutations (with 1 mutation every 10^7 bases) or DNA double-strand breaks (DSBs). DNA DSBs are the most dangerous form of DNA damage and must be repaired to ensure cell viability. Cells have evolved many different mechanisms to guarantee genome stability and repair DNA DSBs. The first line of defense of a cell is the DNA damage checkpoint. Checkpoint activation following a DNA damage allows a pause during G1-S or G2-M cell cycle phases to leave time to the cells for repair. The repair mechanisms required for genome integrity are various and can be distinguished by several characteristics: (i) the specific proteins involved, (ii) the cell cycle timing of repair and (iii) the accuracy of repair. The molecular mechanisms are classified in two categories: the homology-directed repair (HDR) mechanisms, in particular, homologous recombination (HR) that uses the sister chromatid or the homologous chromosome as a template to repair DNA breaks; and classical non-homologous end joining (C-NHEJ) ligating the two broken ends together. Alternatively, cells can repair a DNA DSB by single-strand annealing (SSA) or alternative-NHEJ (alt-NHEJ). Failure to repair DSBs can have dramatic effects to the cell and be at the origin of gross chromosomal rearrangements or cell death.

Some living organisms are more tolerant than others to genome rearrangements: indeed, taking into account the fact that the human cells environment is relatively constant, one can expect that microorganisms can be more tolerant to genomic rearrangement as they are more frequently exposed to abrupt environmental changes. Many examples either in bacteria (1, 2) or in fungi (3, 4) show that genome rearrangements appear upon exposure to changing environments, leading to genetic diversity and to better adaptation. *Candida albicans* is found as part of the normal microflora of the human mucosa but is also known as the most common fungal pathogen causing minor or disseminated infections in immunocompromised hosts (5). This fungus can be exposed to various stresses within the host such as hypoxia, oxidative stress by contact with macrophages, fever and antifungals

among others. *C. albicans* is a prominent representative of the *Saccharomycotina* subphylum among the *Ascomycota* phylum (6-8) and belongs to the CTG clade, decoding leucine instead of serine (9) (Figure 1). An estimated twenty million of years ago, *C. tropicalis* on one hand, and *C. albicans* and *C. dubliniensis* on the other hand, derived from their common ancestor.

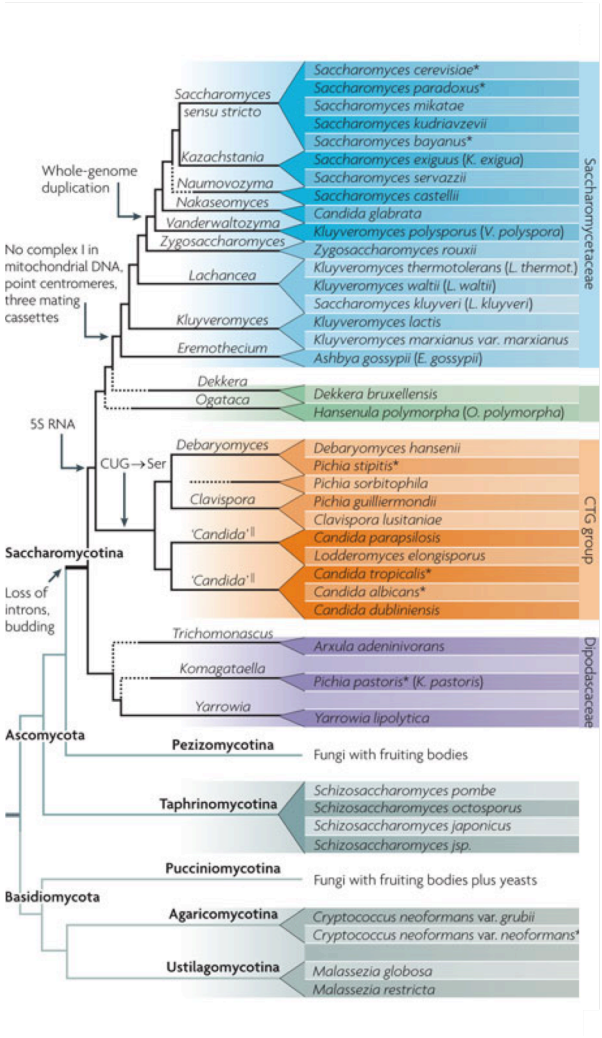


Figure 1: Current view on the phylogeny of *Candida albicans*
C. albicans is a fungus and a major representative of the *Saccharomycotina* subphylum among the *Ascomycota* phylum. It belongs to the CTG clade, whose codon usage allows the decoding of serine into leucine. Figure from Dujon, B (7) .

One particularity of *C. albicans* is that this organism has been proven to be a quasi-obligate diploid: although no meiosis has been described, rare haploid cells resulting from concerted chromosome loss have been found (10).

At a genomic level, the *C. albicans* 28Mb diploid genome is organized in 8 pairs of chromosomes with a high level of heterozygosity (11-13). Furthermore, genome studies have shown that the *C. albicans* genome displays a high level of plasticity: indeed, aneuploidies and loss-of-heterozygosity (LOH) events are observed in clinical isolates (14-17) and in response to various *in vitro* stresses (18-20) and have been associated with the acquisition of resistance to antifungals (for review: (21)). Apart from chromosomal missegregations, LOH events can arise from DSBs occurring throughout the cell cycle and being repaired by

homologous recombination. If not repaired, the damaged chromosome is either truncated or lost. In *C. albicans*, according to the stress the cells have been exposed to, the nature and/or frequency of the resulting LOH varies (18). Relatively little is known on *C. albicans* genome biology while this aspect in the model organism *S. cerevisiae* has been extensively studied. However, these species have diverged some 900 million of years ago. *C. albicans* had undergone rewiring of multiple biological pathways (22, 23). Thus, the commonalities and specificities of the biology of this opportunistic pathogen cannot be understood without detailed investigation.

First, I propose to review the recent advances in the genome structure and dynamics of *C. albicans*, the molecular mechanisms controlling the DNA repair in yeasts and finally, the tools that are already or will be used to study genome stability. Then, I will focus on my research work that was built on two aims: (i) the development of a new tool to study DNA repair mechanisms and recessive alleles in the *C. albicans* genome and (ii) the elucidation of signaling pathways involved in genome dynamics.

I. Introduction

A. Organization of the *Candida albicans* genome

1. Chromosomes

The *C. albicans* genome is about 14 Mb (haploid) that is shared by 8 chromosomes. The chromosomes are named according to their size, ranging from 3.2 Mb for Chr1 to 0.9 Mb for Chr7. The last chromosome is called ChrR and can vary in size according to the variability in the number of ribosomal DNA repeats. Each chromosome bears a centromere and telomeres that constitute protecting ends at both extremities.

i. Centromeres

During mitosis and/or meiosis, eukaryotic cells divide to give rise to daughter cells that would carry as many chromosomes as the parental cells. However, the whole division process can fail if the chromosomes do not reach their destination and this important step relies on centromeres. For review see: (24).

In *S. cerevisiae*, centromeres are regions of condensed chromatin providing the basis for kinetochore assembly, a prerequisite for sister chromatid attachment. The kinetochore is a multiprotein complex not only needed to attach the chromosomes to the mitotic spindles (microtubules) but also ensure chromosome bi-orientation (attachment of sister kinetochores to microtubules extending from opposite spindle poles) and provide a catalytic site for the synchronization of chromosomal segregation along with the cell cycle. The kinetochore is recruited to the centromere along with signaling proteins such as cyclin-dependent kinases that allow setting a signal to the mitotic spindles. The *S. cerevisiae* centromeres have a 116 to 120 bp-conserved sequence among its 16 chromosomes (25). This region contains three centromeric DNA elements (CDE), called CDEI, CDEII and CDEIII (25) and found to be sufficient to generate stable artificial chromosomes in yeast (26) (Figure 2). Indeed, the 8 bp long CDEI is necessary for high fidelity chromosomal segregation, CDEII is an AT-rich region of 78-86 bp and is required for chromosome segregation, while CDEIII contains 7 conserved bases among its 26 bp sequence and mutations in one of these invariant nucleotides would result in the abolition of centromere function (27). Because centromeric regions of *S. cerevisiae* are small and as a unique microtubule attaches to each centromere, the centromeres are called point centromeres.

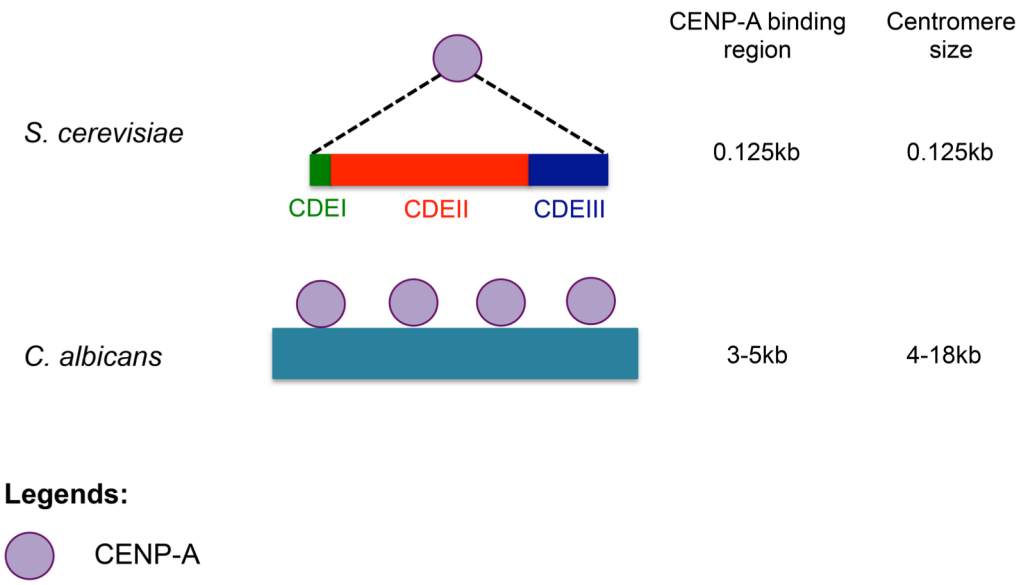


Figure 2: Centromeres of *S. cerevisiae* and *C. albicans*
In the two species, DNA is organized differently at centromeres. In *S. cerevisiae*, the DNA elements are organized in a point centromere, while in *C. albicans* the centromeres are small and regional. Figure adapted from Roy and Sanyal (27).

Several studies in *S. cerevisiae* showed evidence of the occurrence of epigenetics marks as being the determinant that makes the difference between the surrounding chromatin and the chromatin at the centromeric regions (28). However, the exact origin of centromere regulation is not well understood, yet. In *S. cerevisiae,* two well conserved proteins, namely Cse4 and Mif2, were found to be localized at the centromere throughout the cell cycle and are thought to be essential for centromere functionality (29).

By contrast, in *C. albicans*, centromeres are called regional centromeres (Figure 2) and no conserved motif was found between the 8 chromosomes (30). *C. albicans* centromeres were first characterized by the work of Sanyal and collaborators (30) and were identified through the binding of the orthologues of *S. cerevisiae* Cse4 and Mif2 to a 3 kb region, rich in AT (65%). This 3 kb region is uniquely found within a 4 to 18 kb gene-free region flanked by euchromatin. In addition, all chromosomes except Chr7 carry unique short repeats surrounding centromeres: inverted repeats for ChrR, 1, 4 and 5 and long terminal repeats for Chr2, 3 and 6 (31). The construction of an artificial chromosome carrying a 3 kb centromeric region, resulted in mitotically unstable chromosome suggesting that centromere inheritance is propagated by a mechanism that does not rely on the DNA sequence (32). Indeed, the introduction of naked DNA carrying a centromeric region does not allow the spontaneous activation of the centromere, i.e. epigenetic marks (either histone modifications or RNAi) are necessary to have a functional centromere (32). This was also supported by the work of

Mishra *et al.* (31) who showed that the sizes (3 to 4.5 kb) and positions of the 8 centromeres were conserved among several *C. albicans* strains estimated to have diverged from each other by 1 to 3 million of years.

The study conducted by Ketel and coworkers (33) revealed new properties of *C. albicans* centromeres and demonstrated that the loss of a functional centromere results in the formation of a stable neocentromere near the initial centromere. Indeed, by replacing the centromere of Chr5 (CEN5) by *URA3* or the nourseothricin resistance marker, *NAT1*, the authors observed that neocentromere formation is highly efficient when the native centromere is not active. Additionally, the position at which neocentromeres are formed is not unique. Indeed, the new centromere could be formed at different regions on Chr5 within long intergenic regions in the vicinity of repeated DNA sequences and were found to move over a few kb to up to 100 kb (33). Surprisingly, while the centromeres were thought to be recombination-free loci, several studies demonstrated that gene conversions (GC, see section I.B.3.ii p.39) as well as mitotic recombinations at centromeres were occurring at high frequencies in other organisms (34, 35). *In C. albicans*, GC events were also detected when one of the two CEN7 was replaced by a *URA3* cassette (*URA3/CEN7*) (36). Indeed, the formation of a neocentromere at another locus on the chromosome through the inactivation of one of the two native centromeres led to either the propagation of the neocentromere on both chromosome homologs or the reintroduction via GC of the native centromere on the homolog that was deleted for its initial centromere. As GC is known to be a mutagenic process in other organisms, mutations can be introduced in the centromeric sequence (36).

Furthermore, recent work in *C. albicans* made by Mitra and colleagues (37) revealed a strong link between DNA replication, DNA repair mechanisms and chromosomal segregation. They observed that the position of centromeres is not only connected to the proximal presence of an active origin of replication but also governs early and constitutive replication at these proximal origins. Indeed, centromeres in *C. albicans* are replicated earlier than any other locus in the genome (38). This early replication takes place within the first seconds of the S phase and leads to the recruitment of Cse4 at the centromere locus. Again, the epigenetic marks are required for centromere positioning and replication timing (38). Tsai *et al.* (39) detected all putative replication origins within the *C. albicans* genome and more precisely the replication origins close to centromeres, without prior information. They identified two types of replication origins in the *C. albicans* genome: either sequence-dependent origins located along the chromosome arms or epigenetic-dependent centromeric origins (39).

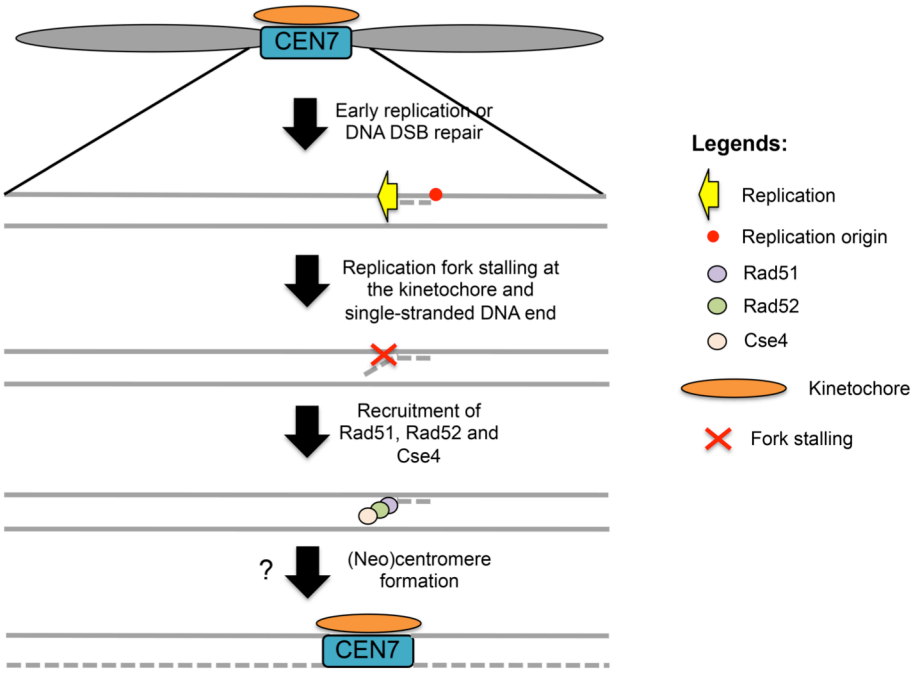


Figure 3: Interpretative model for (neo)centromere formation
Upon fork stalling during the early replication of centromeres, the presence of single stranded DNA would trigger the recruitment of Rad51, Rad52 and Cse4, favoring the formation of a neocentromere. Interpreted from the work of Mitra *et al.* (37).

In addition, Mitra *et al.* (37) have also shown that, in *C. albicans*, Rad51 and Rad52 – homologous recombination proteins (see Section I.B.2.ii p.37) – are detected at stalled replication forks and upon *RAD51* and *RAD52* deletion, kinetochores were shown to be unstable and not functional. Besides, in other organisms, kinetochores are known to be bi-directional barrier DNA binding proteins and are at the origin of replication fork stalling (40) (Figure 3). Thus, Mitra and coworkers (37) proposed the following model (Figure 3) for neocentromere formation in *C. albicans*: when the replication forks coming from the nearby replication origin stalls at the kinetochore complex, the presence of single stranded DNA requires the recruitment of Rad51 and Rad52 proteins. Cse4 is transiently recruited to the DNA break along with an up-to-now unknown chaperone protein stabilizing Cse4 at the replication fork stalling, yielding a functional centromere (Figure 3). Nevertheless, despite the importance of Rad51 and Rad52 proteins in centromeres stability, their absence is not lethal, suggesting overlapping pathways playing a role in centromeres functionality and maintenance.

In summary, *C. albicans* centromeres are short regional centromeres, on which Cse4 binds a region of 3-5 kb long having the particularities of carrying no repeated regions. The determination of centromere formation and the characteristics needed for it to be functional

remains a mystery. However, studies showed that origins of replication are located close to the centromeres and replication forks pause at centromeres by the functional kinetochore, allowing the recruitment of Cse4, Rad51 and Rad52 proteins that play an important role in centromere biology.

ii. Telomeres

Each eukaryotic chromosome end is characterized by a nucleoprotein structure called telomere ensuring their protection from exonucleases and unwanted telomere fusions, through specific DNA folding and proteins. Telomeres have a variable length. For example, in *S. cerevisiae*, telomeres are 350 bp long while in humans, telomeres are large (several kb), nevertheless their function remains the same. In *S. cerevisiae*, the telomeric sequences are composed of repetitions of $T(G)_{2–3}(TG)_{1–6}$ (41), and exhibits a G-rich 3' tail (about 12nt) (G-strand overhang) (42). Chromosomes that lack the $T(G)_{2–3}(TG)_{1–6}$ repeats tend to be highly unstable and therefore, lost.

In *C. albicans*, telomeres carry long divergent repeats 5'-ACTTCTTGGTGTACGGATGTCTA-3', a single-stranded 3' overhang (43) and a common $(GT)_3$ motif despite sequence differences between the Candida species (44). Several telomere-related proteins are known and were found to play multiple roles in telomere biology. The *S. cerevisiae* proteins involved in telomere replication and maintenance are reviewed in (45).

(i) **Rap1** plays a central role in telomere length determination, prevents telomere-telomere fusions, determines telomere location to the nuclear periphery and protects chromosome ends in *S. cerevisiae* and in *C. albicans.* This protein also affects silencing via interactions via the SIR complex in both *S. cerevisiae* and *C. glabrata* but not *C. albicans* (46).

(ii**) Rif1 and Rif2**, negative regulators of telomere elongation are also involved in telomere silencing in *C. glabrata* and *S. cerevisiae* (47, 48) – the role of the Rif1 protein has not been investigated in *C. albicans*, however, no Rif2 ortholog has been identified yet.

(iii) **Yku70 and Yku80** forming the Ku complex, known to be involved in Non Homologous End Joining (NHEJ, see section I.B.2.i p.34) – a repair mechanism that should be avoided at telomeres to prevent telomere shortening – surprisingly, play an essential role in telomere maintenance in both *C. albicans* and *S. cerevisiae* (49).

(iv) **the CST complex, composed of Cdc13, Stn1 and Ten1**, is a single-stranded DNA binding protein complex that acts as a cap in both *C. albicans* and *S. cerevisiae* impeding telomeres to be recognized as DNA DSBs that could trigger DNA repair and impact telomeric structure.

(v) the *S. cerevisiae* **SIR complex** is responsible for gene silencing at the telomeric and subtelomeric regions (telomere position effect) and has been shown to localize at the nuclear periphery.

(vi) **the telomerase**, a reverse transcriptase (TERT), is found in highly replicative cells to compensate the loss of information linked to the incapacity of the DNA replication machinery to replicate the chromosomal ends. In *S. cerevisiae*, TERT is associated with the regulatory proteins, Est1, Est2 and Est3 and is recruited by Cdc13 to ensure the replication of the 3' single-stranded end from a RNA template (TER) and the maintenance of functional telomeres. Replication has been described as a two-step process (Figure 4): (i) following replication of the double-stranded DNA, the C-strand is resected allowing the formation of G-strand overhang and (ii) the telomerase elongates the G-strand followed by the C-strand synthesis by the primase. Finally, the telomere ends are protected from exonucleases through t-loop formation as reviewed in (24).

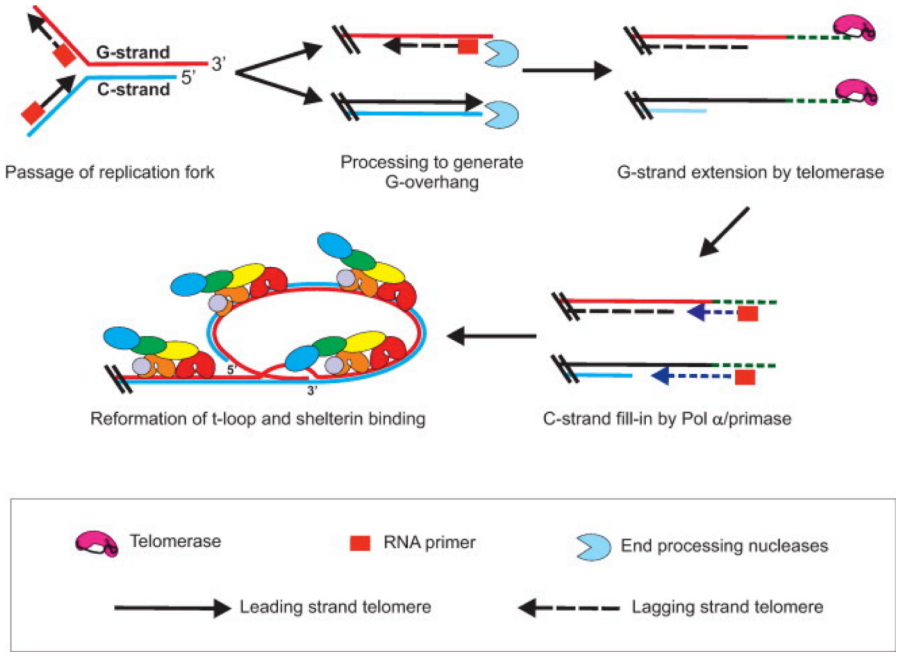


Figure 4: Telomere end replication
The C-strand is resected to form the G-strand overhang, then the telomerase elongates the G-strand followed by the C-strand synthesis by the primase. Finally, the telomere ends form t-loop to be protected from exonucleases. Figure from (24).

In *S. cerevisiae*, telomere length is remarkably uniform thanks to a balance between the DNA end resection and the $T(G)_{2-3}(TG)_{1-6}$ elongation process. Interestingly, the RNA templates (TER1) were found to be highly divergent between *C. albicans* and *S. cerevisiae* (65). In *C. albicans*, TERT and TER1, associated with Est1, Est2 and Est3 are also involved in telomere

capping, telomere maintenance, replication and regulation of the G-tail length (50, 51). Telomerase recruitment has not been characterized in *C. albicans* but it is noteworthy that the domain that is responsible for telomerase recruitment in *S. cerevisiae* has not been found in CaCdc13 (52). Thus, the telomerase recruitment process might not be conserved in *C. albicans*. While the functionality of the telomerase is highly conserved among organisms, some clear differences in the telomere regulation process have been observed between *S. cerevisiae* and *C. albicans*, marking the divergence between these two species.

2. Ploidy

Sexual reproduction is ubiquitous within eukaryotic cells. However, until recently, *C. albicans* was classified as an obligate diploid organism lacking any form of sexual reproduction. For review, see (53).

In *S. cerevisiae*, the mating type is determined by the mating-type (*MAT*) locus that encodes three transcription factors, Mat**a**1, Matα1 and Matα2 determining the **a**, α or **a**/α identity of a cell. Interestingly, in *MAT**a**/α* cells, Mat**a**1 and Matα2 transcription factors repress the haploid-specific genes, rendering the cell mating-incompetent but allowing meiosis. By contrast, in *MATα* cells, Matα2 binds the Mcm1 transcription factor to inactivate the transcription of **a**-specific genes while Matα1 and Ste12 bind Mcm1 to allow the transcription of α-specific genes (53). In addition, in *MAT**a*** cells, the **a**-specific genes are constitutively expressed in absence of Matα2 and Mat**a**1 inactivates α-specific genes. Both *MAT***a** and *MAT*α cells constitutively express haploid-specific genes via the Ste12 transcription factor favoring mating. As a consequence, in *S. cerevisiae,* Mat transcription factors regulate both cell type and meiosis (53).

In *C. albicans*, a mating-type like (*MTL*) locus has been identified on chromosome 5 (Chr5), although meiosis has never been described. Despite the fact that *C. albicans* and *S. cerevisiae* reproduction cycles share some similarities, their regulation shows significant divergence (54). Indeed, *C. albicans* seems to have retained an ancestral form of the *MAT* locus (53): its *MTL* locus is about 8 kb larger than from *S. cerevisiae MAT* locus and encodes not only 4 mating-related transcription factors, namely Mtl**a**1, Mtl**a**2, Mtlα1, Mtlα2 but also 3 additional proteins: Pik**a** or α (phosphinositol kinase), Pap**a** or α (poly A polymerase), Obp**a** or α (oxysterol binding protein) that have functions on their own, not related to mating (55). In *C. albicans MTL**a**/**a*** cells, Mtl**a**2 along with Mcm1 act as an activator of the **a**-specific genes, while in *MTLα/α* cells, no such activator is present (53). *MTL**a**/**a*** and *MTLα/α*

cells were engineered either by gene deletion (56) or by growth on sorbose-containing medium that induces Chr5 homozygosis (57, 58). Remarkably, *C. albicans MTL**a**/**a*** and *MTLα/α* cells were shown to be able to mate, generating tetraploid cells, both *in vitro* and *in vivo* (53). Later on, mating efficiency was shown to be positively linked to the ability of the cells to undergo a reversible phenotypic switch, from a "white" mating-incompetent to an "opaque" mating-competent cell (59, 60). Apart from specific mating capacities, white and opaque cells also differ in morphology (61), metabolic preferences (62) and their interactions with host immune cells (63). In *C. albicans MTL**a**/α* cells, Mtl**a**1 and Mtlα2 form a heterodimer that blocks the mating by preventing switching from the white to the opaque phase (53). As a consequence, only *MTL* homozygous cells would be able to switch and become mating competent (Figure 5). However, a recent study demonstrated that *MTL**a**/α* cells could undergo white-to-opaque switching under specific growth conditions (64). Indeed, when cells are grown in 5% CO_2 with N-acetylglucosamine as the sole source of carbon, *MTL**a**/α* cells can become opaque and exhibit similar phenotypes to the *MTL* homozygous opaque cells; however they are not able to mate. This phenomenon has been shown to be regulated by three transcription factors, namely Rfg1, Brg1, and Efg1 (64).

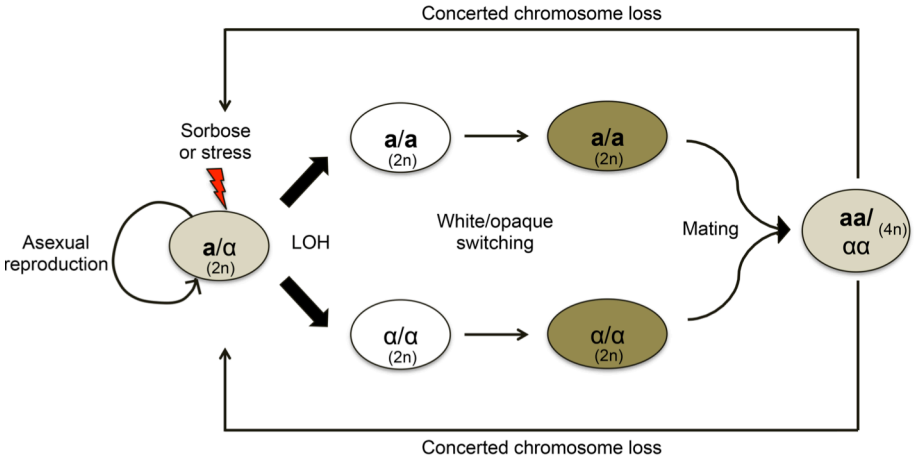


Figure 5: The parasexual cycle of *C. albicans*
C. albicans cells naturally harbor both *MTL**a*** and *MTLα*. When the cells are not exposed to a stress or to sorbose, they reproduce clonally. However, in stressful condition or when sorbose is found as the sole source of carbon, a LOH at the *MTL(**a**/α)* allows homozygosity at the mating locus (**a**/**a** or α/α). Upon white-opaque switching, the white cells become mating competent opaque cells. Then, the opaque cells mate and form tetraploid progenies, which undergo concerted chromosome loss and go back to a diploid or quasi-diploid state. Figure adapted from (65).

Regulation of the white-opaque switching in *C. albicans MTL* homozygous cells relies on several regulators. Wor1, the master regulator of the white-opaque switching (59, 66) is

repressed in *MTL**a**/α* cells by the action of the **a**1-α2 heterodimer and is responsible for a high frequency of opaque switching in *MTL* homozygous cells. In addition, the Wor2, Wor3, Wor4 and Czf1 transcription factors are necessary for maintaining the opaque state and are required to allow opaque switching at an elevated frequency (67-70). By contrast, Efg1 and Ahr1 are both repressors of the white-opaque switching, favoring the white state. Recently Hernday and colleagues (71) identified a new transcriptional regulator of this phenotypic switch, namely Ssn6 that indirectly interacts with DNA, revealing another layer of complexity in the regulation of the white-opaque switching by the repression of genes through interaction with the previously described DNA-binding regulators and favoring of the switch to the opaque state.

In eukaryotes, sexual reproduction involves meiosis as a program that consists of two successive divisions following a DNA replication event. Many genes have been found in *C. albicans* whose orthologs are known to have a role in meiosis in *S. cerevisiae*, such as the *RIM101* pathway genes, *LIG4, YKU70-80* etc… However the absence of *IME1*, the master regulator of meiosis in *S. cerevisiae,* or a gene having an analogous function in *C. albicans* might explain why meiosis has never been observed (53). Nevertheless, Bennett and Johnson (72) showed that tetraploid mating progenies could re-acquire a diploid state by a process called concerted chromosome loss. Indeed, upon culture in a pre-sporulation medium, a glucose-rich medium that is known to facilitate sporulation in *S. cerevisae*, the authors observed that cells with a downshifted ploidy had randomly lost chromosomes. This concerted chromosome loss is thought to be due to mitotic chromosomal segregation defects leading to unstable or transient aneuploid cells, that can go back to an euploid state through the additional loss of extra chromosomes (73, 74). This mode of reproduction is called the parasexual cycle (Figure 5). Nevertheless, the majority of *C. albicans* isolates are *MTL**a**/α* and population genetics studies have revealed that the parasexual cycle is not likely to occur very often in the *C. albicans* population (75, 76).

In addition, *C. albicans* aneuploid cells were also observed upon exposure to stresses such as heat shock or fluconazole treatment or when the replication or mitotic process is defective (21). Recently, Hickman and colleagues (10) have demonstrated that haploid cells are viable. These haploids were first detected upon exposure to fluconazole: among the resistant colonies, tiny colonies could be observed. DNA content analysis confirmed that cells derived from those tiny colonies are haploid resulting probably from a concerted chromosome loss. Hickman *et al.* (10) suggested that haploidization could be a mechanism to clear lethal mutations from the diploid population. Furthermore, some haploid cells were found to be able

to "auto-diploidize" probably due to mitotic defects. Interestingly, haploid and auto-diploidized cells grow significantly slower and are also less virulent (10) than the diploid reference strain SC5314. By contrast, tetraploids can grow as fast as or faster than diploids but are less able to survive and multiply *in vivo* (77). Together these studies show that a change in the whole genome ploidy may have a cost for the fitness and biological functions of *C. albicans*.

3. *Repeated sequences*

The *C. albicans* genome contains four classes of repeated sequences: the major repeat sequences (MRS), the rDNA locus, the retrotransposons and the subtelomeric regions.

i. Major Repeat Sequences (MRS)

MRS are 10 to 100 kb long repeat elements that represent 3% of the *C. albicans* genome. Each MRS is composed of three subunits, 5'-RB2-RPS-HOK-3', always located less than 390 kb away from the centromere and orientated in the 5'-3' direction towards the centromere (Figure 6).

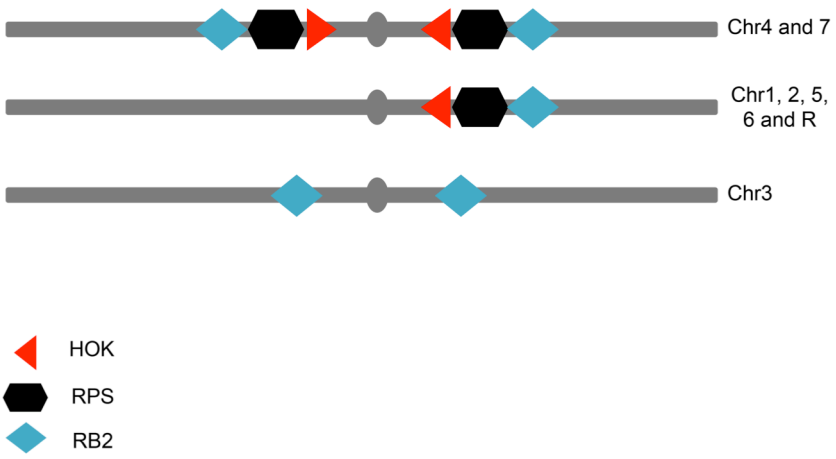


Figure 6: Major Repeat Sequences (MRS)
The MRS are repeated sequences composed of HOK2, RPS and RB2 regions, oriented from the centromere to the telomere. The MRS are found on each chromosome but one. While Chr1, 2, 5, 6 and R carry one MRS, Chr4 and 7 carry two inverted MRS located on both sides of the centromere and Chr3 carries partial MRS composed of RB2 repeats. Based on the work of (78, 79).

The repeated sequences (RPS) are 2.1 kb long and can vary in number (80), although some MRS can present only one RPS unit (81). The RPS carry an "alt" sequence of 172 bp that can be found in tandem leading to RPS size variation between the different MRS (82). The "alt" sequences contain 6 to 8 copies of a 29 bp sequence (called COM29) that includes the 8 bp

*Sfi*I recognition site (82). Outside of the RPS units, *Sfi*I cuts rarely, thus allowing the generation of the first macrorestriction maps to compare karyotype variation before *C. albicans* genome sequencing (83). The RPS units are flanked by two dissimilar sequences that are represented only once per MRS (84): (i) the HOK sequence that is about 8 kb and is on the centromere-proximal side of the MRS and (ii) the RB2 sequence that is of 6 kb. These repeated sequences are located on all chromosomes but one, Chr3: a single MRS is found on Chr1, 2, 5, 6 and R while two MRSs are organized by mirror-symmetry on Chr4 and Chr7 (78, 79) (Figure 6). Moreover, genome analyses of *C. albicans* led to the identification of 14 partial RB2 sequences with two RB2 flanking the centromere of Chr3, despite the absence of an MRS on this chromosome (Figure 6). Partial HOK sequences have also been located in the genome of *C. albicans*. In both cases, partial MRS constituted of HOK or RB2 sequences are localized around the centromere. Interestingly, MRS are found in *C. albicans* and *C. dubliniensis* (85-87) and have not been so far identified in other species. The origin and expansion of these MRS are still unknown and their proximity to the centromere raises questions about the functional relationship between the two DNA structures. Furthermore, the fact that no MRS is present on Chr3 demonstrates that this structure is not necessary to ensure proper chromosomal function. Nevertheless MRS have been suggested to play a role in chromosome disjunction (88) (see Section I.C.2.ii p.45) and have also been linked to translocations (89) (see Section I.C.1 p.43).

Surprisingly, the work done by Freire-Benéitez and collaborators (90) revealed that the MRS is not assembled in a classical heterochromatin structure. MRS repeats are rather found in a transcriptionally permissive state carrying both heterochromatin and euchromatin marks. One of the main roles played by heterochromatin is to maintain genome stability at DNA repeats by preventing chromosomal recombination. Hence, either the hypomethylation state of the MRS chromatin is sufficient to limit recombination events or this role is played by another mechanism (90). However, the lack of meiosis reduces theoretically genome variability and, thus genome instability at repeated sequences such as MRS might be beneficial to such an organism (90).

Furthermore, the analysis of repetitive sequences led to the identification of one ORF localized within the RB2 subunit (91). This ORF encodes a fungal growth regulator (*FGR6*), although this gene is present in eight copies (one copy within each MRS except on Chr5) and the deletion of one of them is enough to generate altered colony morphology (91). For review, see (92).

ii. rDNA

In *S. cerevisiae*, the rDNA region contains between 150 and 200 tandem repeats. The number of repeats depends on the *S. cerevisiae* strain and those repeats occupy about 60% of the ChrXII (93). Each repeat is composed of four interspaced ribosomal RNA (rRNA) genes: (i) the large 25S rRNA, (ii) the small 18S rRNA, (iii) the 5.8S rRNA – all forming the 35S rDNA – and (iv) the 5S rRNA genes. Additionally, different kinds of spacers can be found: two internal transcribed spacers (ITS1 and 2), two external transcribed spacers (ETS1 and 2) and an intergenic spacer (IGS). The overall size of the *S. cerevisiae* rDNA region is of 1.5Mb (94). It is well known that repeated DNA are silenced, in particular at rDNA regions (95). Despite the presence of large copy numbers of rDNA, half of the genes are still transcribed. The proportion of active versus inactive rRNA genes is not well understood yet, however, silencing can influence transcriptional activity. Silencing is mediated through Sir2, Net1 and Cdc14, three proteins forming a complex named RENT – REgulator of Nucleolar and Telophase exit (96, 97). As mentioned earlier, silencing is needed to limit recombination at repeated sites (98) such as rDNA. In *S. cerevisiae*, rDNA repeats can also be found in an extrachromosomal circularized DNA and those were linked to senescence (99).

In *C. albicans*, rDNA is found on ChrR and is constituted of tandemly repeated units of 12 kb sequences carrying the conserved 35S and 5S rRNA separated by two non-conserved and non-transcribed regions (78), NTS1 and NTS2. These repetitive regions are found in a heterochromatin state (90). The number of repeats depends, here again, on the *C. albicans* strain and on the growth conditions. The number can vary between 50 and 200 repeats and is responsible for ChrR size variation that mainly results from recombination between ChrR homologs (78). Hubert and Rustchenko (100) demonstrated for the first time the natural presence of extrachromosomal DNA in *C. albicans*. Indeed, they discovered that this extrachromosomal DNA carrying copies of rDNA could coexist in the circular and linear forms (100). Interestingly, the copy number of repeats in the linear DNA was shown to be dependent on the growth phase, with an increased number in dividing cells and a low number in stationary phase cells (100). While the circular rDNA might have arisen from the loop-out of rDNA repeats that were located on ChrR following a recombination event, the origin of the linear DNA is still a mystery: indeed, it has been shown that this linear extrachromosome does not result from ChrR truncation followed by telomere acquisition at the break site (100). Overall, the control of the total copy number of rDNA repeats is thought to be facilitated through the quantity of extrachromosomal DNA carrying rDNA units, which can be an advantage in some circumstances or growth conditions.

iii. LTR-retrotransposons

Retrotransposons are genetic elements that can autonomously replicate and integrate at various loci within the genome and are widespread in eukaryotes. There are several classes of retrotransposons: (i) retrotransposons without long terminal repeats (LTR) (101) and (ii) LTR-retrotransposable elements (102). As very little is known on non-LTR retrotransposons, I will only discuss about LTR-retrotransposons.

LTR retrotransposons have been extensively studied in *S. cerevisiae*, and are composed of several structures: a PBS (primer binding site) sequence for tRNA binding, *gag* (viral capsid), *pol* (encoding for a polyprotein with RNase-H, integrase, reverse transcriptase, protease catalytic domains) and a PPT (polypurine tract) sequence necessary for reverse transcription. All these sequences are flanked by two LTR sequences (250-600 bp) (103).

In *C. albicans,* 34 families of LTR-elements were identified and are reviewed in (104). The *C. albicans* LTR retroelements follow these 3 rules: (i) the size of the LTR elements ranges from 127 to 780 bp, (ii) LTR sequences start with a 5'-TG and end with CA-3', (iii) LTR elements are present in low copy number.

Eighteen families are found in the *C. albicans* genome as either solo LTRs that arose from the recombination between two LTRs of a LTR retrotransposon, or nonfunctional LTR retroelements. The 16 remaining families still carry LTR-retrotransposon coding sequences (104). Up to now, only three intact and active retrotransposons have been identified in *C. albicans*, namely *Tca2*, *Tca4*, and *Tca5* displaying low sequence diversity suggesting that they might have arisen recently (104).

Tca2 has two particularities: (i) this retrotransposon is part of dsDNA linear sequence that is not integrated into the genome and can be present at up to 100 copies per cell, and (ii) the genes located in the *pol* domain are expressed thanks to the bypass of a STOP codon located upstream its sequence via a secondary DNA structure called pseudoknot (105). Additionally, the *Tca2* promoter is heat shock-induced (104).

Similarly, *Tca4* contains intact *gag* and *pol* genes, and its retrotranscription depends on temperature (104).

The last well-studied retrotransposon is *Tca5* that also carries an intact *gag-pol* region and is present in low copy number. This retrotransposon has a conserved sequence when compared to *S. cerevisiae Ty5* retrotransposon family. Interestingly, *S. cerevisiae Ty5* integrates preferentially in heterochromatic region such as telomeres or subtelomeric regions (104).

LTR retrotransposons are known to have contributed to genome evolution (106) but have also been linked to gene expression variation (107) as well as genome instability (108) in other organisms, highlighting that, although their function is not completely understood, they are likely to play an important role in the biology of eukaryotes.

iv. Subtelomeric regions

In eukaryotes, telomeric and telomere-adjacent (subtelomeric) regions exhibit variability in size and sequence that tend to occur frequently in these genomic regions (for example, see (109)). Subtelomeres are characterized by two regions: (i) the telomere-proximal region that carries short tandem repeats and (ii) the telomere-distal region, harboring gene families, repetitive elements and unique genes (110). It is now well characterized that most subtelomeres in eukaryotes carry gene families, whose copy number varies and has been linked to a better adaptation to an environment. In *S. cerevisiae*, several subtelomeric gene families have been identified. For example, the number of subtelomeric genes involved in carbon source degradation, such as *MAL*, *MEL* or *SUC,* varies according to the selective pressures imposed by the carbon sources available for the yeast (111-113).

In *C. albicans*, the *TLO* genes are the only widespread subtelomeric gene family (78) and encode interchangeable subunits of the tail module of the mediator complex (Med2) (114). This complex allows interaction between transcription factors and RNA Polymerase II in order to activate transcription. The mediator is also known to play an important role in the regulation of virulence-associated factors in *C. albicans* (115-118). As reviewed in (119), the role of these Med2 subunits is not clear yet but, although dispensable, the *TLO* genes seem to be involved in several biological functions in *C. albicans* and are regulated in a heterochromatin-mediated way (119). The *TLO* gene family is composed of 14 genes (13 telomeric and 1 centromeric) in the reference strain SC5314. The telomeric genes are found at all chromosome arms except at the right arm of Chr2 and Chr6, and the left arm of Chr7, while one centromeric *TLO* gene was identified on Chr1 (119). The *TLO* gene family is characteristic of *C. albicans* as only two copies of the *TLO* genes are found in the very closely related *C. dubliniensis* and only one copy (*TLO2*) is found in other Candida species, thus, suggesting that the expansion of the *TLO* genes is recent and that this increase in copy number arises from the ancestral *TLO2* locus (119). In addition, recombination events between *TLO* genes can result in sequence changes or copy number variation at the origin of genotypic diversity (119).

The *TLO* gene family is divided into three clades (α, β, γ) based on the presence of LTR-retrotransposons within the 3' end of the genes (119). LTR retrotransposons can be found in subtelomeric regions in *C. albicans*. In this species, *rel-2* and *CARE-2* elements were found at subtelomeres (120, 121) in a region that contains a large quantity of LTRs between the subtelomeric repeats and the first centromere-proximal unique sequence, therefore suggesting that subtelomeres may be a hot spot for LTR retrotransposon integration (119, 122). Indeed, this LTR-rich region is present at most chromosome ends and all share a similar pattern of LTR retrotransposon insertion, implying that they derive from the same ancestral sequence via multiple duplication events (122).

4. Natural heterozygosity

The natural heterozygosity of *C. albicans* was first reported by Whelan and coworkers (123), showing that natural isolates were heterozygous for several genes required for the synthesis of amino acids, purine and pyrimidines. In 2004, the genome of the *C. albicans* reference strain SC5314 was sequenced by Jones and colleagues (11). The genome of *C. albicans* is composed of 6,198 ORFs, and among them, about 72% are still uncharacterized, while in *S. cerevisiae*, only 10% of the 6,604 identified ORFs are not characterized yet. Sequencing confirmed the natural heterozygosity in the *C. albicans* genome with the detection of 65,787 heterozygous SNPs in SC5314, corresponding to an average SNP density of 1 SNP every 217 bp within the *C. albicans* genome (12). In comparison, in the genome of *S. cerevisiae* diploid strains that have been sequenced by Magwene *et al.,* the number of heterozygous SNPs ranges from 337 SNPs (i.e. 1 SNP every 36,000 bp) for the laboratory strain Σ1278b to a maximum of 37,148 SNPs (i.e. 1 SNP every 324 bp) in the case of a clinical strain (124). Similarly, heterozygosity in humans is relatively low with 1 SNP every 1,900 bp (125).

Muzzey *et al.* showed that 54% of the *C. albicans* genes carry multiple SNPs and among these, 198 alleles contain a premature STOP codon in the heterozygous state localized mainly in the 5' and 3' ends of the genes (12). These premature STOP codons can be at the origin of altered phenotypes or lethality and therefore, the cause for biased homozygosity upon loss-of-heterozygosity (LOH). The presence of lethal alleles was suggested by the analysis of the progenies resulting from parasexual cycle in Forche *et al.*, and led to the conclusion that there was a strong bias towards the homozygosity of a specific haplotype for ChrR, 2, 4, 6 and 7 (74). Similarly, a *rad52Δ/rad52Δ C. albicans* mutant exhibited an increased frequency of spontaneous unidirectional LOH (126). This phenomenon has also been observed in haploids with the report by Hickman and colleagues of a bias in haploid progenies: indeed, only one of

the two haplotypes of Chr3, 4, 6, 7 and most of Chr1 could be observed in the homozygous state (10). Also, in Loll-Krippleber *et al.* (Appendix 1), the presence of one or several lethal alleles on one homolog of Chr4 generating a bias in the nature of the LOH events observed in SC5314 was suggested (20).

In addition to lethal alleles, heterozygous recessive deleterious alleles are also expected to be found in a relative high frequency. Indeed, the identification of the first heterozygous recessive alleles in *C. albicans* was made by Whelan and colleagues who identified the *MET* gene, involved in the biosynthesis of methionine, as carrying a heterozygous recessive deleterious allele along with two recessive lethal alleles located in the *LET1* and *LET2* genes (127, 128). Later, Gomez-Raja *et al.* identified a SNP on chromosome 4 haplotype A changing a glycine into a valine in the *HIS4* gene, which when found in the homozygous state, prevents growth on a medium lacking histidine (129). Therefore, these studies support the presence of recessive lethal alleles on many chromosomes of *C. albicans* that could influence the manifestation of LOH events following a DNA break and probably allow the maintenance of *C. albicans* heterozygosity.

The importance of the heterozygous status of *C. albicans* was further investigated by Hickman and collaborators by comparing haploid isolates or autodiploid homozygous to heterozygous strains (10). Interestingly, they discovered that heterozygosity is a necessary condition for virulence and fitness (10). This observation was supported by a recent study revealing that heterozygosity was positively correlated with faster growth (17). In addition, heterozygosity was found responsible for functional differences between two alleles bringing a selective advantage in specific growth conditions such as the mutations detected in the *CDR2* gene (130). Indeed, such *CDR2* mutations were found in 81% of clinical isolates and shown to confer a selective advantage probably linked to a better antifungal resistance; undeniably showing the relevance of heterozygosity for such a pathogen.

Although *C. albicans* exhibits unique genomic features characterized by unconventional centromeres, major repeat sequences that are specific to this species and an atypical mode of reproduction, this species presents a high number of similarities with *S. cerevisiae* notably regarding the global chromosome structures and repeated DNA, thus, making *C. albicans* a valuable candidate to study the biology of eukaryotes.

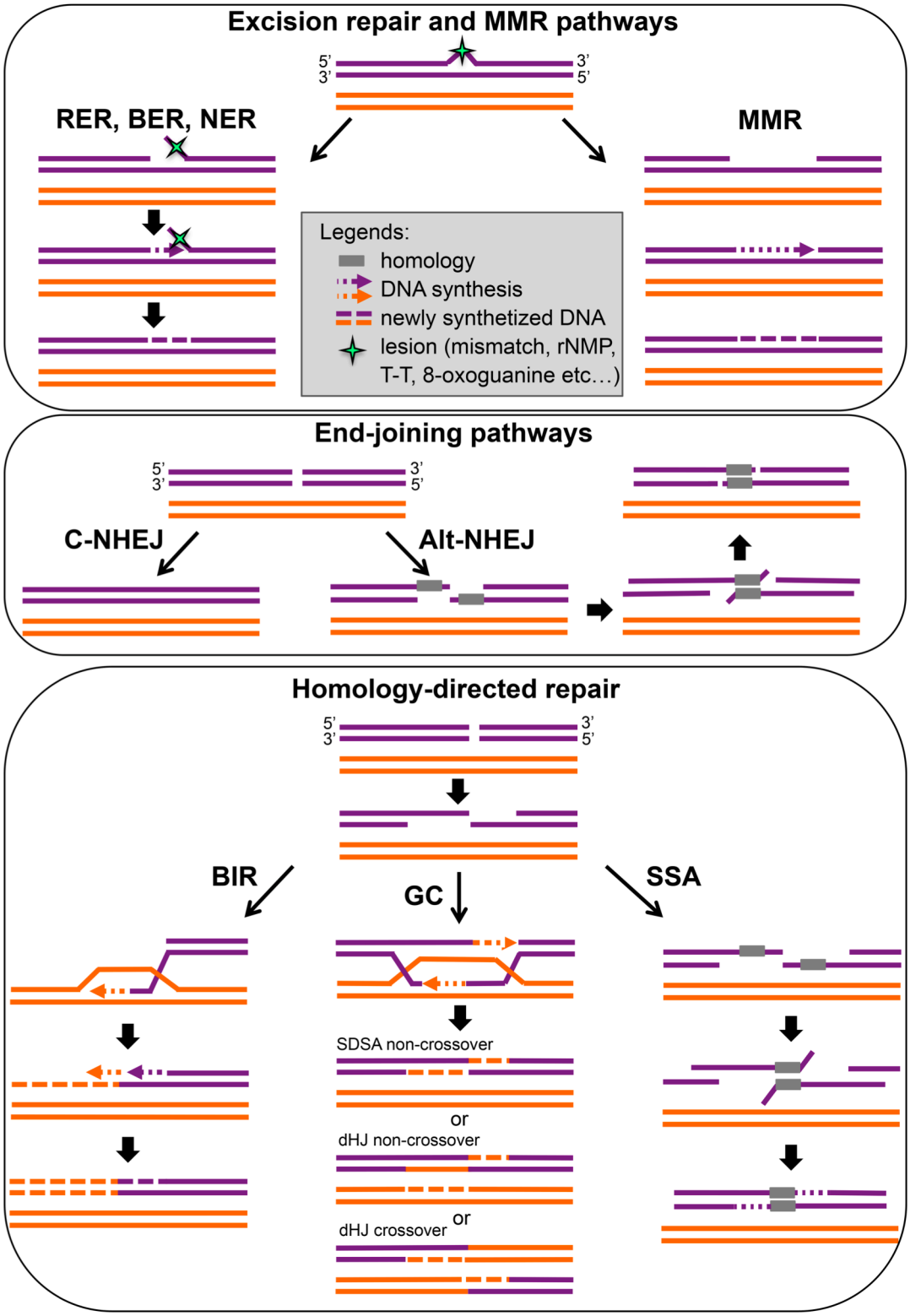


Figure 7: Schematic view of the DNA repair mechanisms, encompassing the excision repair and MMR, the end-joining and homology-directed repair pathways.
RER: ribonucleotide excision repair, BER: base excision repair, NER: nucleotide excision repair, MMR: mismatch repair, C-NHEJ: classical non-homologous end-joining, Alt-NHEJ: alternative non-homologous end-joining, BIR: Break induced replication, GC: Gene conversion (SDSA: synthesis-dependent strand annealing and dHJ: double Holliday junctions), SSA: single strand-annealing. Compiled from (131-135).

Table 1 – Proteins involved in DNA repair in *S. cerevisiae* and their orthologs in *C. albicans*

Repair pathways / Steps	Excision repair pathways					MMR	End-joining pathway		Homology-directed repair pathways			
	RER	BER		NER			C-NHEJ	Alt-NHEJ	BIR		GC	SSA
		SN-BER	LN-BER	TC-NER	GG-NER				Rad51-dependent	Rad51-independent		
Recognition of the lesion with or without scission	RNaseH1 or RNaseH2[1]	N-glycosylase		Rad4-Rad23-Rad33,Rdp9, Rad26,Rad14, Rad1-Rad10	Rad4-Rad23-Rad33, Rad14, Rad1-Rad10	Msh2-Msh3, Msh2-Msh6, Mlh1-Pms1, Mlh1-Mlh3	YKu70-YKu80	Mre11-Rad50-Xrs2	Mre11-Rad50-Xrs2		Mre11-Rad50-Xrs2, Tel1, Sae2, Rad52, H2A	
Resection	-	-		-		-	-	Mre11, Sae2, Exo1, Dna3	Mre11, Sae2, Exo1, Dna3, Sgs1-Top3-Rmi1[2]-Dna2		Mre11, Sae2, Exo1, Dna3, Sgs1-Top3-Rmi1[2]-Dna2	Mre11, Sae2, Exo1, Dna3, Sgs1-Top3-Rmi1[2]-Dna2
3' strand invasion and annealing	-	-		-		-	-	?	Rad51, Rad52	Rad52, Rad59	Rad51, Rad52	Rad52
3' flap end removal	Rad27	Lyase	AP-endo-nuclease or Rad27	Rad2		ExoI	-	Rad1-Rad10	-	-	-	Rad1-Rad10, Msh2-Msh3, Saw1
Strand displacement	PolIII, Pol30, RF-C	PolIV (+Pol30, RF-C ?)	PolIV, Pol30, RF-C	PolII or PolIII, Pol30, RF-C		PolIII, Pol30, RF-C	-		Mcm2-Mcm7[3], Rad51, RPA, Rdh54	?	-	-
Elongation	PolII or PolIII	PolIV	PolII, PolIII or PolIV	PolII or PolIII		PolIII	PolIV	PolIII, PolIV	PolI, PolII, PolIII	Polζ	PolIII	PolIII?
Ligation	Cdc9	Cdc9		Cdc9		Cdc9	Dnl4	Cdc9	?	?	?	?

[1] Rnh35
[2] Nce4
[3] Cdc47

Most of the proteins in *S. cerevisiae* have orthologs in *C. albicans*. The orthologous proteins having a different name in *C. albicans* are indicated in footnotes, and the proteins that do not have orthologs in this species are indicated in red. Compiled from (131-135).

B. DNA repair mechanisms

Despite its nuclear location, the DNA molecule is under constant attack from either exogenous or endogenous factors at the origin of chemical and/or physical damages. Thus, cells have evolved a large panel of repair mechanisms to face these constraints. These repair pathways are lesion-specific and regulated according to the cell-cycle phase. However misregulation of these processes or mutations in genes whose proteins play a role in the repair mechanisms were found to be involved in genome instability and linked to human diseases. Interestingly, very little is known on the repair pathways in *C. albicans*, despite the numerous reports on its high tolerance to genome rearrangements. Here, I propose to describe the different repair pathways from the knowledge gained in *S. cerevisiae* or mammalian cells and highlight the available data in *C. albicans*.

All repair mechanisms that are described in this section are represented in Figure 7.

1. Excision repair and mismatch repair mechanisms

The 2015 Nobel Prize in Chemistry was attributed to Tomas Lindahl and Aziz Sancar for their work on Base Excision Repair (BER) and Nucleotide Excision Repair (NER) mechanisms mainly allowing the repair of either endogenous or exogenous chemical damages of the DNA molecule. The work on the Mismatch Repair pathway was awarded by the same Nobel Prize to Paul Modrich. Thus, I decided to dedicate a few pages to these repair pathways along with a section on the Ribonucleotide Excision Repair (RER) pathway that repairs essentially ribonucleoside misincorporations.

i. Ribonucleotide Excision Repair

In eukaryotic cells, even though DNA polymerases are equipped with proof reading domains that ensure enzymatic fidelity by the recognition of incorrect base or sugar incorporation (136), polymerases are also known to induce replication errors. Such errors include the incorporation of ribonucleoside monophosphate, rNMPs, instead of deoxynucleotide triphosphates, dNTPs, probably explained by the fact that rNMPs are 100 times more frequent in the nucleus of a cell than dNTPs (137). An excessive level of rNMPs misincorporation can lead to a decrease in genome stability through the generation of nicks and an altered DNA structure caused by the distortion of the double-helix (138, 139). Cells have evolved the RER pathway where the RNaseH1 or RNaseH2 rNMP-specific endonucleases recognize stretches/unique misincorporated rNMP(s), respectively (140, 141). In *S. cerevisiae*,

following a nick at the 5' end of the rNMP by RNaseH2 or RNaseH1, either PolIII or PolII polymerases can proceed in strand displacement and DNA synthesis to fill in the gap (Figure 7 p.26 and Table 1 p.27). Then the *S. cerevisiae* flap endonuclease, Rad27, generates a second cut at the 3' end of the rNMP to remove the DNA overhang (reviewed in (142)). However, when the misincorporation is not repaired, PolII and PolIII can by-pass the rNMP with an efficiency of about 60% and this efficiency has been shown to decrease with the number of incorporated rNMPs (136).

Several studies showed that mutations in the *RNH2* gene (encoding RNaseH2) in *S. cerevisiae* diploid cells resulted in an elevated rate of gross chromosomal rearrangements, gene conversion and LOH events (143-145). To date, the RER pathway has not been studied in *C. albicans*.

ii. Base Excision Repair

Several endogenous chemical damages can often occur within eukaryotic cells (138). Base modifications such as oxidation, deamination, alkylation and hydrolysis are known to provoke DNA damage. The BER is reviewed in (131).

Oxidation by reactive oxygen species (ROS), molecules that play a role in signaling (reviewed in (146)), can induce a wide range of DNA damages in mammalian cells (131). Indeed, many groups have worked on the identification of oxidative DNA damage, demonstrating that the main two targets were: (i) purines or pyrimidines giving rise to 5-methylcytosine, uracil, urea or 8-oxo derivatives upon oxidation; and (ii) deoxyribose that can lead to either an abasic site or a single-strand break.

In the case of deamination-induced DNA damages, nitrogen is replaced by oxygen. Only 5-methylcytosine and cytosine carry a nitrogen atom and their deamination can lead to their replacement by either a thymine or a uracil, respectively. This DNA damage has been found to occur mainly in CpG repeats in human cells and has been linked to cancer (147, 148).

In addition, alkylation has also been found to generate mutations in human cells. Even if these kinds of lesions are quite rare, they have been shown to be highly mutagenic. Indeed, both purines and pyrimidines can undergo the addition of a methyl group. Alkylating agents were one of the first treatments used against some cancers. These drugs alter the DNA molecules either by damaging protein functions or by stopping the cancer cells from dividing due to the increased incorporation of mutations. However, the efficiency of DNA repair pathways is responsible for resistance to these drugs (For review, see (149)). Finally, hydrolysis of the

bond between the base and the phosphate-sugar backbone can occur and result in a single-strand break.

In *S. cerevisiae* and human cells, the classical BER pathway corrects, in both nuclei and mitochondria, DNA base lesions that cannot be repaired by direct base reversion. In *S. cerevisiae*, this repair pathway is the result of the action of multiple DNA binding proteins: (i) a DNA N-glycosylase that removes the base from the deoxyribose – depending on the nature of the lesions; different DNA-glycosylases can be involved in the repair (Ntg1, Ntg2, Ogg1…) and are reviewed in (131), (ii) endonucleases, either a lyase (AP-lyase) that initiates short patch-BER (SN-BER), resulting in one nucleotide gap or an AP-endonuclease (Apn1 and Apn2 in *S. cerevisiae*) or the flap endonuclease Rad27 initiating long patch-BER (LN-BER) that removes a segment of DNA, (iii) polymerases, PolIV in the case of SN-BER and PolII, PolIII, PolIV along with the proliferating cell nuclear antigen Pol30 in *S. cerevisiae*, the replication factor RF-C for LN-BER, releasing the remaining ribose-phosphate backbone and filling in the gap, and (iv) a DNA ligase, Cdc9, to seal the remaining gap between newly synthesized DNA and the backbone DNA ((131), Figure 7 p.26 and Table 1 p.27). The requirement for some proteins found in other repair pathways such as the RER pathway highlights the redundancy and overlap between DNA repair pathways.

In C. albicans, Legrand *et al.* (150) described the impact of disrupting genes known to be involved in BER in other species and that have been conserved. The authors demonstrated that mutations in *NTG1*, *OGG1* and/or *APN1* have no impact on oxidizing agents sensitivity and mutation/chromosomal instability rate nor alter the response of *C. albicans* to macrophages or various antifungal drugs (150). Hence, the BER pathway may be redundant with other pathways and does not play an important role in genome integrity in *C. albicans*.

iii. Nucleotide Excision Repair

NER is characterized in *S. cerevisiae* and human cells by the recognition and removal of DNA-distorting lesions and is particularly relevant to repair UV-induced DNA damages (151) as thymine dimers generate a distortion of the DNA molecule. NER is reviewed in (152, 153). More generally, NER can be divided into two subpathways: (i) GG-NER (global genome NER) that recognizes the initiating lesions regardless of the chromatin structure or the transcription status; and (ii) TC-NER (transcription-coupled NER) that is specific to transcription-blocking lesions (152). While both subpathways diverge in the recognition step of the lesions, they converge within the following steps. Indeed, upon lesion recognition, a single-stranded dual incision on both sides of the damage is generated by endonucleases

leaving a 25-30 nucleotides gap that is filled in by a DNA polymerase and sealed to the DNA backbone by a DNA ligase. In more detail, in *S. cerevisiae*, GG-NER recognizes the distorting lesion via the Rad4-Rad23-Rad33 complex that recruits the TFIIH transcription factor, usually required for the initiation of transcription at RNA polymerase II promoters, to generate an opened structure at the damaged DNA site (153-155). TFIIH along with the stabilizing proteins RPA (high affinity for single-stranded DNA) and Rad14 (high affinity for UV-induced DNA damages) form the pre-incision complex that recruits Rad1-Rad10 complex making an incision 15-24 bp upstream the 5' side of the lesion and Rad2, making a 2-6 bp incision dowstream the 3' end side of the lesion (153). Following the dual incision, the RF-C and Pol30 are positioned at the gap to allow the recruitment of PolII or PolIII DNA polymerases (153). Finally, the DNA sealing step following resynthesis requires Cdc9 in *S. cerevisiae* ((153), Figure 7 p.26 and Table 1 p.27).

In the case of transcription-stalling lesions, all the aforementioned proteins that are necessary for GG-NER are also required for TC-NER in addition to Rad26 and Rpd9 that are specific to TC-NER (152). *In vitro*, only *S. cerevisiae rad26* mutants have shown UV sensitivity (152). Studies have demonstrated that Rad26 can promote the bypass of a moderately blocking lesion that is at the origin of ribonucleotide misincorporations and consequently, aberrant proteins (152). In addition, if the RNA polymerase is stalled at strongly blocking lesions, the Rad26-dependent TC-NER subpathway is triggered, leading to the remodeling of the chromatin and the recruitment of the NER machinery. In the case of the Rad26-independent TC-NER subpathway, Rpd9 recruits the NER proteins to proceed to the repair (152).

In *C. albicans*, studies of the NER pathway have been carried by Legrand *et al.* (150) where deletion mutants of the conserved *RAD2* and *RAD10* genes, which encode endonucleases, were constructed. Upon UV exposure, these mutants were found to have acute growth defects. However, the double mutants neither were sensitive to oxidative stress, antifungal treatments, nor showed an increased mutation rate. As a result, similarly to BER, NER does not seem to have an important role in genome stability or drug sensitivity in *C. albicans*.

In addition, in other organisms, NER has also been found to act as a secondary repair pathway in case of dysfunction in BER or RER pathways (142).

Even if NER plays a role in the maintenance of genome integrity in dividing cells, it seems to be more importantly required for the repair of UV-mediated lesions in non-dividing cells (156).

iv. Mismatch repair

In eukaryotes and prokaryotes, mismatch repair (MMR) is a recognition mechanism for short insertions and deletions as well as nucleotide mismatches in a heteroduplex. MMR relies on a bidirectional excision/ resynthesis reaction via the removal of the mismatched nucleotide(s) and re-synthesis of the missing nucleotide(s) using the remaining strand as a template. Nucleotide mismatches can be incorporated during replication and homologous recombination processes both in meiosis and mitosis (157). The excision can start between hundred(s) or several thousands nucleotides upstream or downstream the mismatch and stops just past the mismatch (158). A clear distinction between the old and new strands of recently replicated DNA molecules is necessary for the cell to use the correct strand as a target for repair. This is easily achieved in Gram-negative bacteria such as *Escherichia coli*. Indeed, the methylated state of the parental strand, via the Dam methylation (GATC/GA^{me}TC), allows targeting a scission between the guanine and the non-methylated adenine (159). However, in Gram-positive bacteria and eukaryotes, the mechanism by which the enzyme makes the difference between the mutated and not mutated strand is not completely clear, although several studies showed that the incorporation of rNMPs during replication is a conserved evolutionary process that allows the MMR machinery to distinguish between the newly synthesized and the old strands (160). In *E. coli*, MMR requires the following proteins: (i) MutS and MutL, dimeric mismatch detector and binder, (ii) MutH, an endonuclease, (iii) UvrD – a helicase, (iv) Exo1 (3'-5'), ExoVII (bidirectional), ExoX (3'-5') and RecJ (5'-3') exonucleases, (v) PolIII, the DNA polymerase and finally, (vi) a DNA ligase (159).

In *S. cerevisiae*, Pms1, Mlh1, Mlh2 and Mlh3 are MutL homologues, while Msh1, Msh2, Msh3 and Msh6 have been found to be MutS homologues. In addition, other proteins have been proven *in vitro* to play a role in MMR, such as the exonuclease Exo1, the proliferating cell nuclear antigen Pol30, the replication factor C and protein A (RF-C and RPA) and the DNA polymerase PolIII. All these proteins are reviewed in (161, 162). Msh2-Msh3 and Msh2-Msh6 heterodimers form mispair recognition complexes undergoing a conformational change into sliding clamps that encircle the DNA molecules in presence of ATP. These sliding clamps recruit the accessory complex Mlh1-Mlh2 as well as the endonucleases Mlh1-Pms1 and Mlh1-Mlh3 generating a nick. Endonuclease activity is promoted by Msh2-Msh3, Msh2-Msh6, Pol30 and RF-C (3' excision) or Msh2-Msh6 and RPA (5' excision) (161, 162). Following the scission and Exo1 excision, PolIII and DNA ligase are recruited to fill in the gaps and finalize the repair ((161, 162), Figure 7 p.26 and Table 1 p.27). However, this has

been shown *in vitro* and a lot more remains to be understood about how these complexes work together *in vivo* and how they are coupled with DNA replication.

Other studies focused on the non-canonical actions of MMR in *S. cerevisiae*. Indeed, when a mispair occurs during replication, the distinction between the newly synthesized versus the old strand is easy to make thanks to methylation marks or rNMPs incorporation. However, when the MMR acts outside the context of replication, the distinction between the two strands is not possible and the MMR proteins can randomly use the strand containing the correct genetic information as much as the defect-carrying strand and be, in this case, at the origin of a mutation in non-dividing cells (For review see (163)).

Additionally, studies have been made to understand the impact of DNA mismatches on homologous recombination. A study in *S. cerevisiae* revealed that one mismatch in a 300 bp recombination region led to a 3-fold reduction in the recombination rate (164). One way to decrease the recombination rate between homeologous sequences is by heteroduplex rejection. When there is a high level of mismatches in a region undergoing homologous recombination, the heteroduplex can get destabilized through the action of Msh2-Msh6 and the Sgs1 helicase that would unwind the homeologous recombined DNA sequences (165), resulting in either no repair or if repaired, no strand exchange using one of the non-crossover recombination pathways (166-168). This raises questions regarding the efficiency of the MMR repair pathway in *C. albicans*: indeed, with a relatively high level of heterozygosity, recombination rate would be low, and thus LOH would be rare.

In *C. albicans*, the role of Msh2 and Pms1 proteins has been investigated (169). Legrand *et al.* (169) showed that the knockouts of *MSH2* and *PMS1* genes resulted only in an increase of the mutational rate that was interestingly associated with antifungal resistance. A link between MMR and antifungal resistance was recently found in *C. glabrata* clinical isolates in which *MSH2* mutations promote an elevated rate of antifungal resistance (170).

2. DNA DSB repair mechanisms

As mentioned earlier, DNA DSBs can arise through the action of various cellular stresses. The repair of such DNA DSBs is essential for the maintenance of genome integrity. The two main repair mechanisms are homology-dependent repair and end-joining pathways. Both cell cycle status and DNA end resection – a conserved process generating single-stranded DNA ends – have been shown to initiate the choice for DNA DSB repair pathways.

i. End-Joining

End-joining pathways are conserved among bacteria, archea and eukaryotes with the exception of *C. albicans* in which no proof of the existence of this mechanism has been provided yet. However, most of the key players in end-joining pathways are conserved in this species (171, 172).

- Classical or canonical-NHEJ (C-NHEJ)

In *S. cerevisiae* and mammalian cells, C-NHEJ has been identified as a repair mechanism that joins either two blunt DNA DSB ends or two compatible broken DNA ends, followed by their subsequent ligation. No resection step is required to allow the sealing of the two broken DNA ends. In *S. cerevisiae*, this repair mechanism is known to rely on Ku proteins (173) – a heterodimer of YKu70 and YKu80 that encircles the DNA molecules with a high affinity for DNA DSB ends – and the DNA ligase 4, Dnl4, that has an inherent tolerance to mismatches, allowing the sealing of the DNA broken ends brought together as reviewed in (174-176). These proteins are also associated with other factors such as nucleases, DNA-dependent protein kinases, DNA polymerases and the Mre11-Rad50-Xrs2 complex (MRX). In *S. cerevisiae*, C-NHEJ is a four-step reaction: (i) following a DNA DSB, the chromatin is modified to allow the removal of histones and facilitate the recruitment of the C-NHEJ core proteins. This step is processed by chromatin remodeling complexes, i.e. Ino80 and RSC, after a hypothetical DNA damage signaling (177, 178). Additionally, Ku proteins, known to be localized at telomeres, can relocate at the DNA DSB site followed by DNA-dependent protein kinases that interact with H1 histones and promote DNA release. (ii) Then, the two broken ends are brought together, forming end synapsis, a conformation also called MRX globular head. (iii) The DNA polymerase, PolIV, is subsequently recruited at the DNA DSB ends whose role is to fill in the gaps. Finally, (iv) Dnl4 is recruited to seal the DNA extremities together ((174), Figure 7 p.26 and Table 1 p.27). Both in yeast and humans, the choice between HR and C-NHEJ depends on several factors. First, the loss of HR genes does not increase the efficacy of the C-NHEJ pathway (179), suggesting a regulation of these two mechanisms instead of a competition. Second, the resection level has been shown as one of the main regulators that dictates the choice for the repair mechanism to use (For review see: (180)): indeed, C-NHEJ is working mainly in the G1 phase as resection of the 5' ends is under the control of the cell cycle and is active only during S-G2, allowing HR to be efficient at this period of the cell cycle (181) even though C-NHEJ can also be active in the S-G2 phase (182). In *S. cerevisiae* Sir proteins have been shown to be required for this process by forming a heterochromatin-like structure at double-stranded breaks (183). In addition, Åström

et al. described that the heterozygous state of the *MAT* locus represses NHEJ through the formation of the repressor composed of Mat**a**1 and Matα2 in G2. This step allows the cells to adapt the repair pathway to the availability of the template: in G1, no homologous template is available, therefore requiring the use of NHEJ to repair a DNA DSB while in G2, the cells can repair the damage by homologous recombination (184). The nature of the DNA DSB (blunt or cohesive ends) can also determine the choice of the repair pathway used. In mammalian cells, C-NHEJ has rapid kinetics (~30 minutes) whereas the HR pathway displays much lower kinetics (up to 7 hours) (185). However, in the G2-S phase, complex DNA DSB structures (186) and heterochromatin can delay C-NHEJ from acting and HR can take over in case C-NHEJ is too slow to repair (187).

However, C-NHEJ has not been described in *C. albicans.* Indeed, several studies assessed the effect of the deletion of the gene encoding the key protein involved in homologous recombination, Rad52, or studied the role of Lig4 in NHEJ (126, 171, 172, 188-190). The authors observed that the homozygous and heterozygous deletions of the *LIG4* gene and the homozygous deletion of *RAD52* did not affect the level of translocations, suggesting the absence of NHEJ repair mechanisms. Hence, the hypothesis set in *S. cerevisiae* might be also relevant in *C. albicans* i.e. the NHEJ pathway is inhibited by the heterozygosity at the mating type locus, thus raising questions about the existence of NHEJ in *MTL* homozygous or haploid *C. albicans* cells. In addition, NHEJ core proteins play a different role in *C. albicans*, notably in morphogenesis, virulence and telomere length regulation (171, 172, 189).

Apart from DNA DSB repair mechanisms, the C-NHEJ pathway is involved in V(D)J recombination – a physiologic system aiming at intentionally generating DNA DSB to create a variety of antigen receptors in early B or T cells – and class switch recombination at the origin of immunoglobulin heavy chain variability (For review see: (191)).

Despite translocation events, C-NHEJ is thought to be a relatively error-free repair pathway thanks to the ligation of blunt ends or complementary broken DNA ends. However, an alternative error-prone mechanism of C-NHEJ has been identified in the late 1980's and is associated with deletions and insertions of the surrounding regions to the break site. This mechanism is called alternative-NHEJ (alt-NHEJ) and is discussed below.

- Alt-NHEJ

Early studies in *S. cerevisiae* demonstrated the existence of a mutagenic alternative end-joining pathway called alt-NHEJ when C-NHEJ is disabled. It is still not clear how many

distinct mechanisms alt-NHEJ encompasses but one mechanism, the microhomology-mediated end-joining (MMEJ) pathway, has been widely studied over the last two decades.

This pathway was first observed in *S. cerevisiae* in which transformed linearized plasmids were religated even in absence of YKu proteins, though, at a low rate (192). The recovered religated plasmids were found to exhibit deletions surrounding 5 to 18 bp of microhomology (193). In mammalian cells, the MMEJ pathway is not only a back-up pathway when HR or NHEJ is not functional, but is also active when the other repair mechanisms are fully efficient. Truong and colleagues demonstrated that 10 to 20% of the cells would repair an induced DNA DSB by MMEJ in presence of functional HR and C-NHEJ pathways and that the use of this pathway increases at collapsed replication fork (194).

Unlike C-NHEJ, MMEJ shares the end resection step with the HR pathway in both yeast and mammalian cells. Indeed, in *S. cerevisiae*, the repair of a DNA DSB by MMEJ is initiated by the exonuclease activity of Mre11, Exo1 and Dna2 performing a 5' to 3' resection at the origin of a 3'-OH single-stranded tail. This resection is usually short except when microhomologies are separated by more than 2 kb (194). In humans, this resection step is initiated by the Poly-(ADP) ribose polymerase 1 (PARP-1) and conducted by the Mre11-Rad50-Nbs1 complex and CtIP proteins. In *S. cerevisiae*, the second step is the annealing of the microhomologous sequences: in *S. cerevisiae*, the microhomology region has to be at least 6 bp long (195) and the larger the sequence (12-17 bp), the better the repair by MMEJ (196). In addition, the heterologous 3' tail resulting from the annealing of nucleotides that are localized a few bp from the 3'OH end has to be removed to allow the next step. In *S. cerevisiae,* this task is processed by the Rad1-Rad10 proteins (197). Once the microhomology regions are properly annealed, the single-stranded side-regions can be filled-in (Figure 7 p.26 and Table 1 p.27). The predominant DNA polymerases involved in this process are PolIII and PolIV in *S. cerevisiae* (195, 196, 198) and Polθ in multicellular eukaryotes (199). The final stage of MMEJ repair is the ligation of the newly synthesized DNA to the DNA backbone thanks to Cdc9 in *S. cerevisiae* (200).

In *S. cerevisiae* and mammalian cells, the choice between MMEJ and HR does not depend on the cell cycle. Indeed, MMEJ is active throughout the G2-S phase, similarly to HR (181). MMEJ and HR share the end resection step and are both active at the same cell-cycle phase: thus, either the absence of extensive resection favors the MMEJ pathway or the presence of RPA and Rad51 inhibits MMEJ, thus promoting HR repair mechanisms (200).

Similar to C-NHEJ and due to the poor DNA homology required to repair a DNA DSB by MMEJ, this repair mechanism can lead to very dramatic genomic rearrangements such as

deletions of variable length, inversions and chromosomal translocations, shown to be at the origin of many cancers through the formation of neo-oncogenes as a result of gene fusion. Indeed, 300 gene-fusions have been identified and are thought to be responsible for about 20% of human cancers (201). On the other hand, MMEJ might also play an important role in maintaining the stability of repetitive DNA sequences that can be at the origin of dangerous genomic rearrangements following DNA DSB repair by HR (202).

MMEJ can also be thought to be an acquired process that allows genetic variation and evolution. Indeed, several studies in metazoans revealed a significant amount of microhomologies that have favored intron loss or that are present in the genome at former break points (203, 204). In addition, MMEJ is more frequently used in mammalian cells than in yeast. Furthermore, MMEJ is not a back-up pathway in mammals. Taken together, these data suggest a more important role of MMEJ in multicellular eukaryotes than in yeasts.

ii. Homology-directed repair (HDR)

In *S. cerevisiae* and mammalian cells, all HDR pathways, namely break-induced replication (BIR), gene conversion (GC) and single strand annealing (SSA), occur in both meiosis and mitosis. They are initiated by the resection of the 5' broken ends, allowing a 3' tail to be coated by RPA and Rad51 proteins involved in the protection of the single-stranded DNA and search for a homologous template sequence allowing repair. In BIR and GC pathways, Rad51 initiates homology search and subsequently, strand invasion of the homologous template by the 3'-tail within 30 to 45 minutes (For reviews, see: (205, 206)). The DNA synthesis then differs between the different repair pathways.

In *C. albicans*, detailed studies have been conducted on the molecular basis of homologous recombination (126, 207). In particular, the recent work of Bellido and colleagues reported for the first time results on interactions between HDR genes and identified Rad51 and Rad52 as proteins having a partially redundant function in *C. albicans* (208). Almost all, if not all, proteins involved in HDR pathways are conserved in *C. albicans* and are thought to work in a similar manner as in other eukaryotes (172) (See section I.C.3 p.53).

- Break-induced replication

BIR has been best characterized in *S. cerevisiae*. Under some situations such as collapsed replication fork, a DNA DSB presents only one end for repair. In this case, NHEJ cannot be involved. Thus, if this DNA end shares a minimum of 63 bp of homology with a template DNA, the repair operates through homologous recombination (HR) allowing the repair from

the break site to the proximal telomere (209). The outcome of this repair pathway depends on the nature of the DNA template. Indeed, if the template is located at an ectopic locus, the resulting repair would give rise to non-reciprocal translocations; however, if the template is the homologous chromosome or the sister chromatid, then the repair would be at the origin of extensive loss-of-heterozygosity (LOH).

In the context of DNA repair in *S. cerevisiae*, DNA synthesis can be a slow process as it requires a replication execution checkpoint to sense the presence of one or two ends at the DNA DSB site in order to orientate the repair towards the BIR or GC pathway: indeed, while GC is completed within 2 or 3 hours, BIR can be delayed up to 4-6 hours (210).

From a mechanistic point of view, in *S. cerevisiae*, BIR involves the Mcm2-7 DNA helicase complex, Rad51 and RPA to form an opened helix structure called a displacement loop or D-loop (also called bubble) (211). In addition, Pif1 is required as an accessory protein for the recruitment of PolIII in the D-loop (212). Pol32, a dispensable subunit of PolIII for DNA replication in S phase, Pol30 and possibly PolII are essential for the synthesis of the leading strand of the broken end using another DNA sequence as a template (213). PolI is involved in the synthesis of the lagging strand (213).

The model of replication during BIR has remained unclear until the recent work of the Malkova and Symington groups demonstrating that DNA is synthesized in a conservative manner (133, 214) and that the lagging strand is synthesized asynchronously to the leading strand (133): once the leading strand synthesis is completed, the lagging strand will be replicated using the leading strand as a template (Figure 7 p.26 and Table 1 p.27). However, during BIR, DNA synthesis is highly mutagenic due to (i) the lower fidelity of PolIII, (ii) an increased dNTP pool known to decrease the fidelity of DNA polymerases and (iii) a lower efficiency of the MMR pathway to correct BIR-mediated errors, leading to a frameshift rate that is 2,800 times higher than during normal replication (215). This mutagenic state is emphasized by the presence of template switching during DNA synthesis (216). Indeed, the transition between PolII-independent and PolII-dependent DNA synthesis is equivalent to the transition between an unstable to a mature stable replication fork (216). Thus, during the synthesis of the first 10 kb, template switching proceeds through consecutive rounds of dissociation and re-invasion between the newly synthesized DNA strand and the template, and is facilitated by the Mph1 helicase that allows the displacement of the D-loop following Rad51-mediated strand invasion (217). Despite the widespread action of BIR in various cellular processes, BIR can also be related to human diseases (218). Indeed, BIR can lead to

non-reciprocal translocations in both humans and yeast when a DNA DSB occurs near transposons or DNA repeats (219).

Although most of the BIR-mediated repair events are Rad51-dependent, some Rad51-independent events were observed. Indeed, this pathway acts preferentially in presence of short homology sequences (220) and requires less than 33 bp of homology (220) while for Rad51-dependent pathway, the longer the homologous sequence, the more efficient the repair pathway (221). Rad51-independent BIR requires Rad52, Rad59, Rdh54 and the MRX complex (222). Nevertheless, how strand invasion occurs in absence of Rad51 remains one of the biggest mysteries of the BIR repair pathway. In addition, the Rad51-independent microhomology-mediated BIR (MMBIR) pathway was found to have major implications in complex genomic rearrangements in humans (223-225). This subpathway can be initiated by the interruption of repair-specific DNA synthesis, by the presence of DNA breaks in the template blocking the leading-strand synthesis or by the occurrence of secondary structures that obstructs the DNA synthesis (226). All these circumstances force the 3' DNA end to be dissociated from its template. The repair is then reinitiated by the coupling of the 3' end with microhomologous regions (226). Then, PolIII is replaced by the hyper mutagenic translesion polymerase, Polζ. The polymerase Polζ either synthetizes short DNA fragments at the 3' end upon the presence of microhomologous regions, or extends the 3' end in absence of microhomology. However, the latter process is not well understood (226). The MMBIR repair pathway is responsible for copy number variations that can be associated with genomic disorders in humans (224). Rad51-independent BIR is also involved in telomere maintenance in absence of telomerase and is named alternative lengthening of telomeres (227).

In *C. albicans*, the studies that have been made to assess the outcomes of random DNA DSBs showed that BIR was frequently observed, although BIR, GC with CO, MCO and chromosome truncation could not be distinguished (18, 20). In addition, BIR events were observed *in vitro* resulting from either spontaneous DNA DSBs or upon overexpression of *RAD51* and *RAD53* (20) (Appendix 1), as well as upon deletion of *RAD50*, *MRE11*, *RAD53* and *DUN1* (19), upon treatment with H_2O_2, fluconazole or temperature (18); or *in vivo*, during commensalism (14) and passage through a mouse (15).

- Gene conversion

GC has been observed for the first time during meiosis with the occurrence of aberrant allele segregation and proved later to result from meiotic DNA DSB in *S. cerevisiae* (228). The bases of this mechanism were in agreement with the DNA DSB repair mechanism proposed

by Szostak and colleagues (229). In eukaryotic cells, gene conversion is defined as the nonreciprocal exchange of genetic information from one template to its homolog and leads to localized LOH events. Initiated by end resection and followed by strand invasion, the heteroduplex composed by the invading strand and its template forms an intermediate, which can be repaired through pathways possibly involving crossovers. Following DNA synthesis and ligation, double Holliday junctions (dHJ) are formed and can be resolved either with or without crossovers (for review, see: (230)); or synthesis-dependent strand annealing (SDSA, reviewed in (231)) pathway minimizes crossovers and can be associated with MMR heteroduplex correction, leading to LOH events during gene conversion (232). Unlike BIR, GC involves both ends of the DNA DSB that have arisen from the 5'-3' DNA end resection. These 3' tails need 337-456 bp of homology in humans (233). If several DNA DSBs occur at the same time, GC can appear either in *cis* (ends belonging to the same DNA DSB) or in *trans* (ends from two different DNA DSBs).

As mentioned earlier, in *S. cerevisiae*, the replication execution checkpoint mediates the choice between GC and BIR based on the topology of the DNA ends (210). Very recently, Haber's group demonstrated in *S. cerevisiae* that the replication execution checkpoint favors GC over BIR, when possible (234). In addition, during GC, DNA replication is observed within 30 minutes following strand invasion and DNA replication machinery components used by BIR are not required for GC completion (235). Indeed, the PolI-primase is not necessary for the synthesis of the lagging strand because of the presence of a 3'-OH end from the second end and neither are Pol32 nor Pif1 (212).

The length of GC tracts has been assessed in several studies using *S. cerevisiae*. GC tracts are usually short (236). However, long tract GC events could not be observed due to technical limitations. Therefore, the recent studies led by Yim and collaborators could circumvent this limitation and revealed for the first time two classes of gene conversion sizes: 1 to 21 kb and 26 to 121 kb (237). Such variations in size can be explained by the recognition and repair of heteroduplex formed between either the invading strand and the template DNA or by the newly synthesized strand and its complementary strand by the MMR pathway (232, 238). For very long tracts, it is either the MMR pathway that generates long gaps on the invading strand thanks to Mlh1-Msh6 that recognizes mismatches (239) or a double-BIR event (237): the BIR pathway starts by repairing the DNA DSB using the homologous chromosome but instead of finishing the DNA synthesis on the same template, BIR ends by copying DNA from the broken chromosome (237).

In *S. cerevisiae*, although GC is the least error-prone repair pathway, it is known that GC is associated with a high mutation rate (240). As seen with BIR, DNA synthesis during GC can undergo interchromosomal template switching (ICTS) (239) resulting from errors by PolIII (235). However, the GC-mediated ICTS machinery seems to be different from the one associated with BIR: neither Rad51, Mus81, Yen1, Mph1 nor Pol32 are involved in the GC-associated template switching; nonetheless, Rdh54 is known to be involved in ICTS but its precise role still needs to be identified (239). Despite its mutagenic activity, one role of GC is to maintain homogeneity in the number of the tandem repeats found in ribosomal DNA in *S. cerevisiae* (241).

Furthermore, biased GC have been widely studied and are known to have a high impact on genome evolution in humans (242) as well as in other eukaryotes (243). It has been shown that gene families with genes sharing a high degree of homology are more prone to undergo GC and fix guanine and cytosine instead of adenine and thymine (244), thus increasing the level of substitution, a hallmark of positive selection in these species.

GC has also been linked to human diseases when GC uses pseudogenes as templates and transfers DNA from a non-functional to a functional gene, resulting in pathogenesis but GC can also have a therapeutic effect if GC exchanges the non-functional gene with the functional one.

Despite the importance of this repair mechanism, little is known in *C. albicans*. In this species, GC events were shown to arise *in vitro* either spontaneously, upon *RAD53* and *BIM1* overexpression (20) or upon H_2O_2 (18), or *in vivo*, during commensalism (14). GC is also known to be involved in centromere protection by conserving its function and location in *C. albicans* (36) as mentioned in section I.A.1.i. p.10. The role of GC in loss-of-heterozygosity in *C. albicans* will be further detailed in section I.C.3 p.53.

- Single-Strand Annealing

SSA is involved in DNA DSB repair only when the DNA DSB is flanked by direct repeat sequences and has been best described in *S. cerevisiae* (245). Unlike the two other HDR pathways, SSA does not require strand invasion by Rad51, thus relying only on strand annealing between the resected ends of the DNA DSB. In mammalian cells, NHEJ is the favorite pathway to repair DNA DSB (246). By contrast, *S. cerevisiae* uses preferentially SSA (247). In addition, while SSA is mainly active during G1 in yeasts, it is repressed in mammalian cells (248). If the DNA DSB occurs between repeated sequences, the DNA end resection would uncover the repeats as single-stranded DNA, resulting in the formation of a

heteroduplex upon annealing by Rad52 (247). However, SSA can have dramatic consequences for the cell such as being responsible for translocation events. Indeed, Sugawara *et al*. found that 3% of divergence between two repeated sequences lead to an antirecombination mechanism such as SSA; a process called heteroduplex rejection (see section I.B.2 p.33) (247). This is dependent on Msh2-Msh6 and Top3-Rmi1 complexes as well as the Sgs1 helicase. Nevertheless, the Mlh proteins and the MMR specific proteins seem to be dispensable (249). Additionally, Chakraborty and colleagues quantified that 80% of cells undergo heteroduplex rejection in wild-type cells (249).

Upon SSA, the unique sequences surrounding the DNA DSB (>30 nt) and found between the repeats become nonhomologous sequences, are removed by the Rad1-Rad10 complex, Msh2-Msh3 and Saw1 (250, 251). This results in sequence deletion making this repair pathway highly mutagenic. However, in the case of flap ends <30 nt, the removal of the nonhomologous DNA sequences is accomplished either by Rad1-Rad10 without the action of Saw1 or by the exonuclease activity of PolIII (250). The gaps resulting from the flap ends removal are filled in by a polymerase probably associated with Pol30 (249). The ends between the newly synthesized DNA and the DNA backbone are finally bound together by a ligase (Figure 7 p.26 and Table 1 p.27).

Several studies reported that repetitive DNA sequences are widespread throughout the genome of humans (252) and other eukaryotes, and thus are prone to recombination events. SSA-mediated non-allelic homologous recombination leads to copy number variations (253) and possibly to diseases. Nonetheless, SSA has never been reported nor studied in *C. albicans*. Indeed, apart from the redundant proteins playing a role in almost all HDR pathways, no role of the orthologous genes of *S. cerevisiae* has been investigated in *C. albicans*, despite the presence of repeated sequences such as MRS, rDNA or telomeric repeated genes and a high level of heterozygosity that would trigger an elevated rate of heteroduplex rejections, and thus favoring SSA in this species.

C. Genome instability in *Candida albicans*

C. albicans displays a high tolerance to genome plasticity. The fact that this fungus maintains a high level of heterozygosity while displaying a high number of genomic rearrangements is intriguing. Following the discussion on the natural heterozygosity of *C. albicans* (Section I.A.4 p.24), we will now discuss on the genome dynamics and their effect on the biology of this species. For review, see (21, 254).

1. Rare dynamic events: translocation events

Due to their repetitive structure, MRS are hotspots for recombination and are associated with translocation events, when a reciprocal recombination occurs between two MRS repeats from nonhomologous chromosomes (83, 255). For instance, Chu *et al.* described 3 translocation events in the vicinity of a MRS in WO-1, the strain in which white-opaque switching was first observed (83). Additional studies made by Lephart and colleagues aimed at better understanding the role of the MRS on translocation by investigating the impact of the presence of MRS on mitotic recombination between two chromosome homologs (89). To do so, the authors inserted two selection markers located on both side of the MRS on Chr5, with one on the *MTLα* proximal side and the other, on the opposite side of the MRS, on the *MTL**a***-bearing chromosome. They measured a recombination rate of 2.82×10^{-6} events per generation, equivalent to the known recombination rate in *C. albicans* (256). The measured rate is not far from the spontaneous mitotic recombination rate in *S. cerevisiae* (2×10^{-6} to 1×10^{-5} events per generation) and depends on the locus that is studied (257). Additionally, in *C. albicans*, Lephart *et al.* showed that recombination events at MRS are stress-induced and that the recombination rate is not higher at MRS than at non-repetitive sequences under normal growth conditions (89).

Translocation events were observed in *C. albicans* clinical isolates, notably one event involving three chromosomes (Chr2, 4 and 7) (258, 259). These studies revealed a link between the occurrence of a translocation event and a phenotypic change. Indeed, upon translocation between Chr4 and 7, the cells exhibited high ploidy and an unstable number of nuclei: this phenotype was called Sps- (suppressor of ploidy shift) (258). Moreover, a clinical strain harboring multiple translocation events (Chr1 and 4, Chr2 and 4, Chr4 and 7, Chr6 and 7, and Chr2, 4 and 7) displayed incapacity to form chlamydospores (259). In addition, it has been shown that translocations are linked to virulence. Indeed, in *C. dubliniensis* and *C.*

glabrata, whose genome is rich in RPS and megasatellites respectively, both reciprocal and non-reciprocal translocations were reported and associated with an attenuated virulence (260, 261).

2. *Partial or complete aneuploidies*

An aneuploidy is characterized by an abnormal number of chromosomes, such that either whole chromosomes or fragments of chromosomes are deleted or present in a higher number. In aneuploid eukaryotic cells, a change in the level of expression of a large amount of genes may occur, possibly resulting in a stoichiometric imbalance at the protein level. Interestingly, in *C. albicans,* specific aneuploidies have been shown to be under positive selection in particular environments. For recent reviews, see (262, 263).

i. Chromosome truncation

Chromosome truncations are found in both laboratory strains and clinical isolates. The laboratory strain BWP17 (264) and some of its parental progenitors such as RM1000 (265) have undergone a chromosome truncation that shortened the right arm of Chr5 during *HIS1* disruption (266, 267). The truncated Chr5 is approximately 35 kb-shorter, lacking a region containing 17 genes and rendering the region distal to the *HIS1* gene monosomic (267). Interestingly, a sequence of 9 bp sharing similarities with telomere repeats was identified nearby, probably helping in the formation of a de novo telomere at the truncated end.

Another study demonstrated that the deletion of the *RAD52* gene in the *C. albicans* laboratory strain CAI-4 (268) was responsible for a 100-fold increase in the number of both chromosome truncations and losses (126). In this study, the authors took advantage of the presence of a heterozygous allele of *HIS4*, presenting a non-functional copy of the gene on Chr4A and a functional allele on the other haplotype (129). The authors screened for cells that became auxotrophic for histidine and analyzed their karyotype (126). Interestingly, while chromosome truncation-mediated LOH could be observed in 40% of the *C. albicans rad52* mutants that became his-, truncations were never observed in *S. cerevisiae rad52Δ/rad52Δ* cells (269). By contrast, while translocations were not detected in *C. albicans*, they were the predominant events occurring in *S. cerevisiae rad52* mutants. This can be explained by the fact that the DNA damage checkpoint favors de novo telomere addition at broken ends in absence of Rad52 in *C. albicans* (126).

Moreover, as mentioned earlier, truncations were also observed in strains isolated from patients. Indeed, Selmecki and colleagues identified a 200 kb truncation on the left arm of

Chr5 (270) in a strain isolated from a patient treated with fluconazole following a bone marrow transplant (271). Further analyses revealed that the breakpoint at the origin of the truncation is located in the vicinity of several genes belonging to the same fungi-specific gene family and presenting 73-75% of sequence similarities (270). The high resemblance between the sequence of these genes, along with the presence of telomeric repeat-like sequences, might have facilitated the recombination and led to chromosome truncation.

ii. Chromosome loss

Many studies in *C. albicans* revealed the presence of homozygous chromosomes present in a unique copy, a state called monosomy associated with a long range LOH. The first studies demonstrating the occurrence of chromosome loss as a response to a stress were conducted by the Rustchenko group. The authors showed that when grown in presence of sorbose or arabinose as sole source of carbon, the cells that were able to form colonies had lost one copy of Chr5 or Chr6 respectively (57, 272). This is due to the presence of one or several genes on Chr5 or Chr6 that negatively regulate(s) the genes allowing growth in presence of these carbon sources. When present in two copies, the suppressors prevent growth of *C. albicans* on sorbose or arabinose containing media. Lephart *et al.* (88) took advantage of this property of *C. albicans* to assess the impact of MRS on Chr5 loss (see Section I.A.3.i p.19). The authors plated *C. albicans* cells on sorbose containing media in order to select cells that have lost one of the two Chr5 homologs. By doing so, the authors were able to link MRS size to the ability of a chromosome to be lost: they observed that between the two homologs of a chromosome pair, the one carrying the smallest MRS tends to be maintained, while both tend to be lost with the same frequency when they carry a MRS of equal size (88). One hypothesis to explain this phenomenon is that the MRS size affects chromosome disjunction: the larger the MRS, the more difficult the separation of the two homologs. Therefore, both chromatids of the homolog carrying the large MRS along with one chromatid of the short MRS-bearing homolog stay in the mother cell, while the daughter cell carries the second chromatid of the homolog harboring the short MRS (88).

Later on, a study on the Rad52 protein showed that chromosome loss was the major event observed in the absence of the *RAD52* gene. This has been previously described in several studies conducted in *S. cerevisiae* (for example: (269)). In this species, *rad52Δ/rad52Δ* cells that had undergone chromosome loss rarely experienced a chromosome reduplication event (269). In *C. albicans*, almost all chromosome loss events were followed by the reduplication of the remaining chromosome (126). Indeed, monosomy has been reported as highly unstable

in *C. albicans*: monosomic cells return rapidly to a euploid state upon standard growth conditions (57, 273). Moreover, *in vitro* experiments also showed that transformation following a heat shock induces chromosome losses in *C. albicans* (274).

At the molecular level, in *S. cerevisiae*, an abnormal number of chromosomes often results from a defect in the mitotic chromosome segregation process by either non-disjunction or unequal segregation of chromosomes (275). This has been sustained by the work of several groups showing the involvement of the chromosome segregation machinery in the maintenance of genome stability in *C. albicans*. Loll-Kribpleber and colleagues demonstrated, based on the consequences of overexpressing *CDC20*, that the proper regulation of the Anaphase Promoting Complex is crucial for chromosome stability (20). Furthermore, a study used (i) oxidative stress, mimicking the attack of *C. albicans* by macrophages, (ii) elevated temperature, reflecting fever and finally (iii) fluconazole, one of the most commonly used antifungal treatment targeting the biosynthesis of a membrane component (Erg11) against *C. albicans* infections to investigate genome instability (18). The authors not only observed that *C. albicans* cells displayed a significant increase in LOH rates in response to these stresses (1 to 40-fold for temperature, 3 to 72-fold for oxidative stress and up to 285-fold for antifungal) (Figures 8A and B), but also that the molecular mechanisms at the origin of these LOH events change according to the nature of the stress the cells are exposed to (18). Indeed, while oxidative stress increases the number of GC and BIR-mediated LOH events, both heat stress – that involves a limitation in heat shock proteins playing a role in kinetochore assembly (276) – and fluconazole – that generates trimeras at the origin of aneuploid cells (277) (see the section I.C.2.iii p.47) or haploid cells (10, 278) – influence the segregation machinery, and thus, amplify the number of chromosome losses (18).

Additionally, UV also induces chromosome loss events (274). Forche *et al.* (74) also observed that concerted chromosome loss during the *C. albicans* parasexual cycle results in a high number of aneuploidies, including cases where some chromosomes have become monosomic before being reduplicated, as suggested by the presence of two copies of homozygous chromosomes.

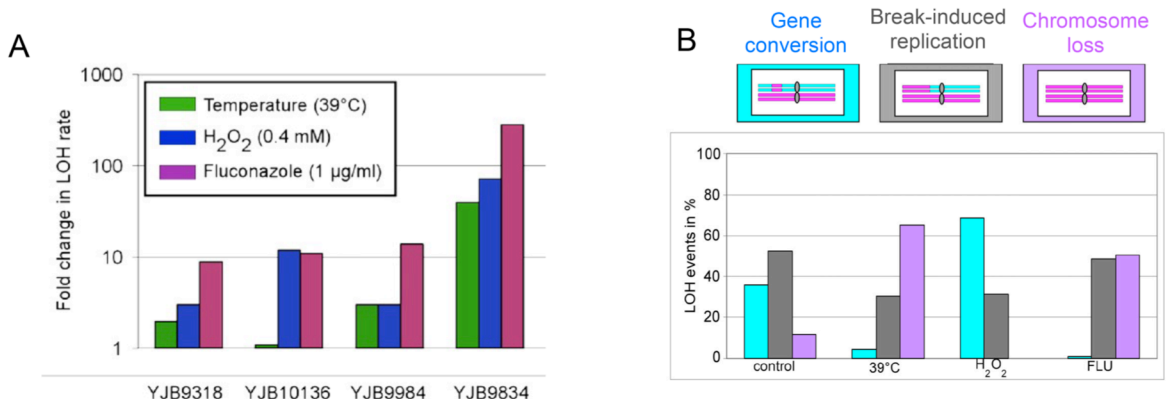


Figure 8: Influence of stresses on the frequency and types of LOH events
(A) Exposure to elevated temperature, oxidative stress and fluconazole increases the LOH rate of the different strains. (B) According to the type of stress the cells were exposed to, the outcomes varied: while chromosome loss are the main LOH events occurring upon high temperature and fluconazole, gene conversions were mainly observed upon oxidative stress. Figure from Forche *et al.* (118).

Chromosome loss has also been reported *in vivo*. The study of commensal isolates of *C. albicans* demonstrated one occurrence of chromosome loss for one of Chr6 homologs (279). Similarly, the impact of a single passage through a mammalian host was assessed in a study made by Forche and coworkers. These researchers discovered that, among the recovered colonies, some had undergone chromosome loss followed by reduplication for both haplotypes of ChrR and Chr2 (15).

Chromosome loss events are often associated with fitness cost. The striking examples illustrating this phenomenon are the rare occurrence of haploid cells (10) or the higher doubling time displayed by the half haploid laboratory strain WO-2 derived from WO-1 (280). The fitness cost associated with the loss of one copy of a chromosome is thought to be a signal triggering chromosome reduplication, probably explaining the instability of haploids (126).

iii. Extra chromosomes

Aneuploidies are one of the characteristics of cancer cells and are widespread within eukaryotes, notably in *C. albicans*, often conferring drug resistance. As discussed in the previous sections, *C. albicans* can undergo a parasexual cycle generating tetraploids able to return to a diploid state through a concerted chromosome loss process. However, diploidy is not fully restored and the progenies resulting from the parasexual cycle are often aneuploid, as shown in the study made by Forche *et al.* (74). The authors demonstrated that there was a

bias towards trisomy of Chr4 in the progenies derived from the parasexual cycle: all strains carrying at least one aneuploidy were trisomic for Chr4. Trisomies of Chr2, 5, 6, 7 or R were also detected in at least one of the progeny strains. This work has been supported by a recent study showing that ploidy reduction in tetraploid progenies is highly stochastic, passing by different aneuploid intermediates until reaching the euploid state (73). Trisomies were also reported in the case of selective growth on different carbon sources: a gain of Chr2 or an extra copy of the right arm of Chr4 confers the capacity to grow on arabinose and sorbose-containing media, respectively (57, 272).

In addition, trisomies were reported in a study using 70 fluconazole resistant and sensitive strains from both clinical and laboratory environments (281). Different antifungal drugs can be used against fungal infections: (i) echinocandins target the cell wall, (ii) polyenes act on the ergosterol-containing plasma membrane, (iii) flucytosine inhibits DNA synthesis, (iv) azoles block ergosterol biosynthesis at the endoplasmic reticulum (282). Importantly, treatment with azoles of patients infected by *C. albicans* can result in the appearance of azole-resistant isolates. Among the pool of resistant strains reported by Selmecki *et al.* (281), 23 were found to be aneuploid. These aneuploidies were mainly trisomies and segmental aneuploidies, mostly occurring on Chr5. Additionally, the authors revealed that the number of aneuploidies was seven times higher in fluconazole resistant (21/42 strains) than in fluconazole sensitive (2/28 isolates) strains (281). In another study where a combination of azole and calcineurin was used as a treatment against the *C. albicans* cells, the researchers observed multiple aneuploidies resulting in extra copies of chromosomes or fragments of chromosomes (283). The relatively high occurrence of aneuploid cells upon treatment with antifungal drugs can be partially explained by the recent work of Harrison and colleagues (277) who discovered the formation of trimeras when cells were exposed to antifungals (Figure 9).

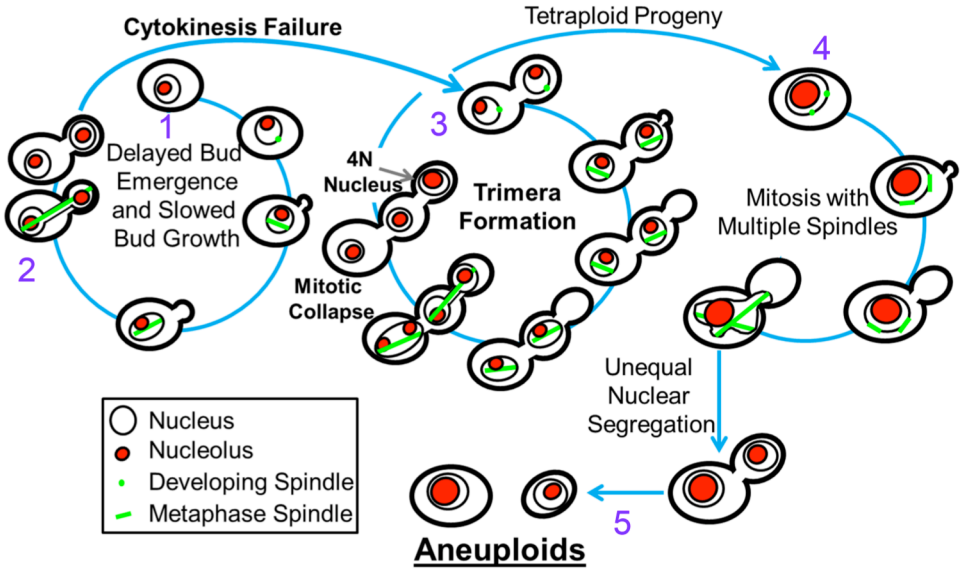


Figure 9: Model for aneuploid cell formation upon exposure to fluconazole
1) Exposure to azole, 2) mitosis collapse, 3) dimeras followed by trimeras formation due to mitosis collapse, 4) formation of a tetraploid cell that can divide, 5) unequal chromosomal segregation giving rise to aneuploid daughter cells. Figure from Harrison *et al.* (440).

The hypothesis that the authors proposed is the following (277): upon fluconazole treatment, bud formation would be delayed as compared to both cell cycle progression and mitotic spindle formation, leading to a failure to divide (collapsed mitosis). Then, the "daughter" cell still attached to its mother cell would form a bud, producing a trimera structure. The newly formed bud would result in either a tetraploid or a dikaryotic cell via a mitotic collapse. The tetraploid cell would be able to give rise to viable aneuploid progenies through chromosome missegregation (Figure 9).

Work by Bouchonville and coworkers demonstrated that aneuploidies were unstable upon transformation with the lithium acetate protocol, displaying a shift from trisomic towards disomic strains (274). Interestingly, aneuploidies resulting in extra chromosomes are also frequent in strains having undergone a single passage through a mouse (15) or in clinical isolates (16, 17, 260, 284) with a bias towards the aneuploidy of the smallest chromosomes (17, 274, 281), as well as the presence of Chr4 and 7 aneuploidies in multiple isolates, either being well tolerated or bringing a selective advantage (17). However, many aneuploidies were observed in clinical isolates without being necessarily associated with antifungal resistance. Consequently, it is important to keep in mind that aneuploidies such as a gain of chromosome(s), can be advantageous by driving positive selection and yet intermediate until the acquisition of beneficial mutations that would allow the cell to grow with a lower fitness cost than the one associated with aneuploidy. This phenomenon has been reported in Ford *et*

al., who showed that aneuploidies are transient events sometimes not linked to antifungal resistance and thus, suggesting that these aneuploidies might facilitate survival until the cells reach a more stable state (16). In addition, trisomy, as monosomy, correlates significantly with a slow growth (17). Thus, tackling fungal infections could become possible. Indeed, cells that, upon a drug treatment, acquire an aneuploid karyotype adapted to their environment and leading to a confine aneuploid population (or evolutionary trap), can be targeted using another drug for which the cells are not well fitted (285). Despite the absence of new drug targets, the study of Chen *et al*. (285) opens the door to new hopes towards the development of novel strategies to treat fungal infections as well as cancer.

iv. Supernumerary chromosomes: isochromosomes

One of the first studies that showed a clear link between genome rearrangements and acquisition of antifungal resistance was perfomed by Selmecki *et al.* (281). A detailed analysis of resistant strains by Comparative Genome Hybridization revealed the occurrence in some isolates of an extra chromosome called isochromosome that consists of either two inverted copies of single chromosome regions flanking a centromere, in particular isochromosome of the left (L) arm of Chr5 (isochromosome Chr5L, i.e. i(5L)), isochromosome Chr5L fused to the right (R) arm of Chr3 (i.e. i(5L)+3R) or isochromosome 5L fused to an intact Chr5, (i.e. att-i(5L)) (see figure 10A) or chromosome fragments (270, 281). An isochromosome of Chr5R (i.e. i(5R)) were also observed in sensitive strains. Noticeably, Chr5 left arm harbors *ERG11*, whose gene product catalyzes a critical step in the ergosterol biosynthesis pathway and is the target of azole agents; and *TAC1* that encodes a transcription factor which controls the expression of the *CDR1* and *CDR2* genes, encoding azole efflux pumps. Amplification of *ERG11* and *TAC1* genes through the formation of i(5L) is sufficient to increase azole resistance (Figure 10A). It is noticeable that many isolates carrying i(5L) had acquired other aneuploidies on chromosomes that harbor genes involved in antifungal resistance: the higher the concentration of fluconazole, the more aneuploid chromosomes (286). The occurrence of isochromosomes is transient: once the stress condition disappears, the extra chromosome is lost (281). This is one proof of adaptation to a stressful environment.

Interestingly, isochromosomes have been observed in the fission yeast, *S. pombe* (287-289). For example, in the context of a failed homologous recombination-mediated repair upon deletion of the *RAD3* and *RAD17* genes, Blaikley and coworkers observed that extensive end processing led to Chr16 loss and was associated with the formation of an isochromosome

(287), resulting from centromere rearrangements (289).

From a mechanistic point of view, eukaryotic isochromosomes can arise by either (i) the impaired disjunction of chromosomes during anaphase giving rise to one chromosome with two short arms and a second chromosome with two long arms (290); (ii) U-type exchange between two chromatids belonging or not to homologous chromosomes and resulting in the appearance of isochromosomes upon replication (291); or (iii) upon a DNA DSB close to the centromere, a BIR involving the inverted repeats surrounding the centromere allow the repair by the duplication of the undamaged arm (288) (Figures 10B, C and D). Contrarily to the advantageous effect that isochromosomes can have in *C. albicans*, isochromosomes have also been observed in humans but were linked to many diseases, highlighting the dramatic effect of such a genetic abnormality if recurrent.

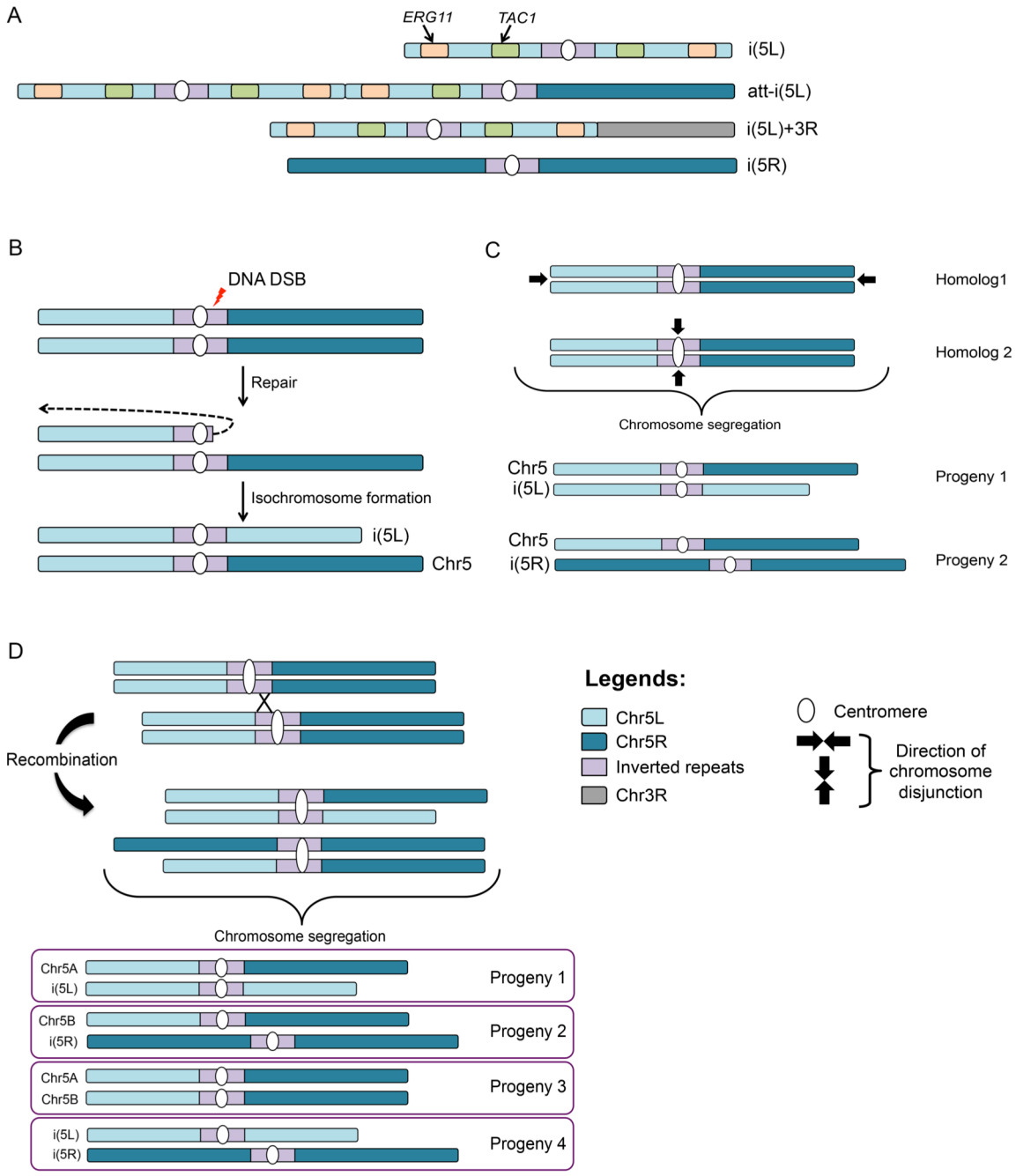


Figure 10: Isochromosomes in *C. albicans* and mechanisms of appearance
(A) In *C. albicans*, several observations made by Selmecki and colleagues revealed the formation of isochromosomes resulting from left or right arm of Chr5 duplication (i(5L) or i(5R)), Chr5 left arm duplication and fusion to Chr5 (att-i(5L)) or Chr5 left arm duplication and fusion to Chr3 right arm (i(5L)+3R). Isochromosomes can arise from (B) DNA DSB in the inverted repeats on Chr5 right arm and repair using inverted repeats located on the left arm, (C) Misdisjunction of chromosomes or (D) U-type (recombination event for example exchange between repeated regions belonging to two chromosomes) which upon chromosome segregation will give rise to isochromosomes. Figure compiled from (21, 288, 291).

3. *Homologous recombination (HR)-dependent LOH*

In addition to chromosome loss and truncation that generate LOH, gene conversions, BIR and mitotic crossovers are also responsible for LOH events. These HR-mediated LOH have been well characterized by Forche *et al.*, (18). Indeed, as mentioned above (see section I.C.2.ii p.45), the molecular mechanisms leading to LOH events differed according to the nature of the stress that is encountered (Figure 8 p.47).

The authors demonstrated a 2-fold increase of short LOH tracts upon oxidative stress, while both fluconazole and high temperature triggered a large decrease in the number of LOH events resulting from gene conversions (Figure 8B p.47). The work performed by Loll-Krippleber *et al.* aimed at identifying genes, which upon overexpression, increase the LOH rate (20). To this end, the authors used a fluorescent reporter of LOH to detect LOH events by flow cytometry (Figure 11A). This system consists of a heterozygous locus on Chr4, with one homolog carrying the blue fluorescent protein-encoding gene and the other homolog, the green fluorescent protein-encoding gene. Thus, upon a LOH event at the BFP/GFP locus, the cells switch from a doubly fluorescent state to a monofluorescent state.

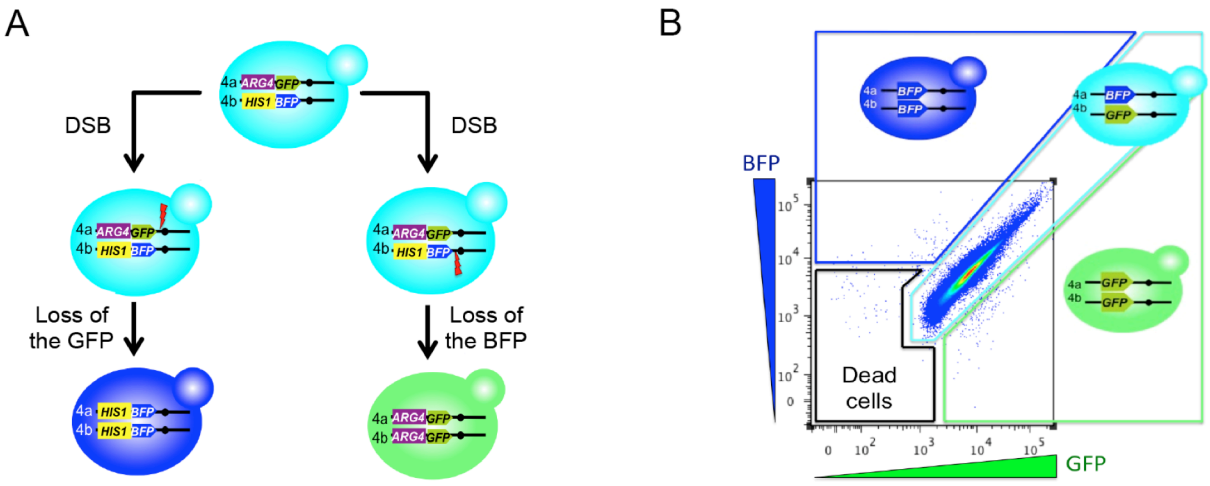


Figure 11: A FACS-optimized LOH reporter system.
The LOH reporter system. (A) This system consists of an artificial heterozygous locus on Chr4 with the BFP-encoding gene placed on one homolog of Chr4 and the GFP-encoding gene introduced at the same locus on the other homolog. Upon an LOH event at the BFP/GFP system integration locus, the cell can go from a double fluorescent state to either a mono-BFP (LOH event on the GFP-bearing chromosome) or a mono-GFP state (LOH event on the BFP-bearing chromosome). (B) Cells undergoing an LOH event at the BFP/GFP locus are revealed by flow cytometry. On a flow cytometry output, the mono-fluorescent cells are localized in the side gates and the double fluorescent cells are found in the middle gate. (Feri *et al.*, *accepted for publication*, mBio).

Interestingly, the authors showed that the overexpression of some genes not only increased significantly the LOH rate but also influenced the outcome of the LOH events that were observed. Indeed, Loll-Krippleber and colleagues reported that the overexpression of *RAD51* triggered mainly BIR and chromosome loss, while upon *BIM1* and *RAD53* overexpression, both GC and BIR were observed in addition to chromosome loss events. Hence, this study demonstrates that LOH outcomes can be influenced by the protein involved in repair, a process that is probably regulated by environmental cues and thus, signaling pathways.

Moreover, many studies reported that LOH events play an important role in the acquisition of antifungal resistance by the homozygosis of hyperactive mutations giving an advantage to *C. albicans* for its survival as a pathogen. For example, homozygosis of hyperactive alleles of *TAC1* (Figure 12), *ERG11*, *MRR1* and *UPC2* are associated with increased azole resistance (292-296). Indeed, all the previously cited genes encode either a direct target of antifungals (Erg11) or transcription factors controlling genes encoding drug efflux pumps (Tac1 upregulating *CDR1* and *CDR2* or Mrr1 increasing the expression of *MDR1*); or genes involved in ergosterol biosynthesis (Upc2 controlling *ERG2* and *ERG11*). Other allelic homozygoses have been linked to resistance to antifungals. In the case of the *PAP1* gene, which is located within the *MTL* locus, its homozygosis (*PAPα/PAPα*) is linked to hyperadenylation; enhancing the stabilization of the *CDR1* transcripts, thus, leading to increased azole resistance (297).

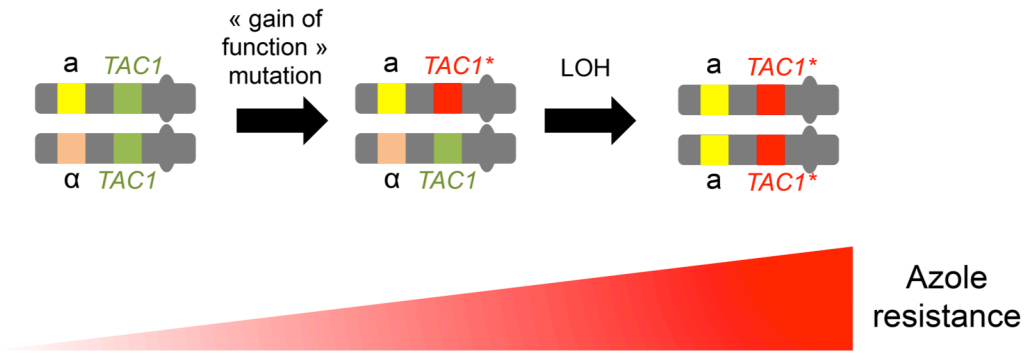


Figure 12: LOH and resistance to antifungals
TAC1 gene product is a transcription factor which controls the expression of the *CDR1* and *CDR2* genes, encoding azole efflux pumps. Gain-of-function mutations in the *TAC1* gene have been reported in clinical isolates and were associated with a resistance to azoles. Upon LOH event rendering the hyperactive allele homozygous, the cells become fully resistant to azole antifungals. Based on the work presented in (292).

Another example is the homozygosis of the hyperactive allele of *FUR1*, which is the first description of a mutation conferring resistance to 5-flucytosine (298). This mutation was

identified in *C. albicans* strains from Clade 1 only. Antifungal resistance has also been reported upon echinochandin treatment where a homozygous mutation in the *GSL1* gene (ortholog of *ScFKS1*) conferred an increased resistance to micafungin (299). All together these studies suggest a positive effect of LOH on the biology of *C. albicans*. However, LOH can have a negative effect on the biology of *C. albicans*. Indeed, as mentioned earlier in Section I.A.4 (p.24), several studies suggested the presence of recessive deleterious alleles in the genome of *C. albicans.* For example, Ciudad and colleagues (300) identified the recessive allele of *MBP1*, located on the left arm on Chr2, as responsible for methyl methanesulfonate (MMS) sensitivity when found in the homozygous state in CAI-4 (268), a SC5314 derivative that is naturally homozygous on the left arm of Chr2. Homozygosis of the *MBP1* deleterious allele coupled with a LOH on Chr3 right arm resulted in an enhanced sensitivity to MMS, the first report of "sign epistasis" in *C. albicans*. More importantly, some recessive alleles have been reported as possibly lethal (10, 20, 74, 126, 129) and if found in the homozygous state following a LOH event, these alleles will prevent cells from growing on rich medium.

LOH are also involved in telomere and chromosome size expansion/contraction at subtelomeric repeats (TLO genes), MRS or rDNA (see Section I.A.3 p.19). Indeed, recombination between subtelomeres from different chromosome ends were often followed by LOH, resulting in the complete loss of one TLO gene and the expansion of the other (301). This phenomenon was quantified and occurred every 5,000 generations in average and contribute to the genome variability of the *C. albicans* population. Furthermore, expansion or contraction of MRS repeats is caused by unequal HR at sister chromatids contributing to variability in chromosome length for all but one chromosome within the *C. albicans* population (88, 302). The occurrence of ChrR size polymorphisms is also observed in *C. albicans*, notably through the passage in a mammalian host, and is the result of HR at rDNA sequences between the two homologous ChrR (303, 304). However, no link between repeat size variation and virulence was found. Together, these examples demonstrate the importance played by LOH events in *C. albicans* biology. Despite the fact that most of the LOH events found within clinical isolates or in cells exposed to stresses bring selective advantages for its survival, the homozygous regions generated by HR-mediated LOH are irreversible and conserved over time. The only way to regenerate heterozygosity in homozygous regions is to undergo point mutations, a relatively slow process. Alternatively, the parasexual cycle could contribute to restore heterozygosity as proposed by Ciudad *et al.* (300) (see section I.A.4 p.24).

D. Tools used to study DNA repair mechanisms

1. *Triggering DNA lesions to study DNA repair*

Cells are under constant exposure to stresses, some yielding DNA damages. To investigate the molecular mechanisms involved in the repair of stress-induced DNA lesions, a variety of agents can be used to mimic cellular stresses, efficiently triggering DNA breaks or lesions that have to be repaired to ensure survival. Among those, genotoxic drugs (for examples: rapamycin, bleomycin or hydroxyurea (HU)), heat shock, UV, DNA damaging agents (for instance: H_2O_2 or MMS) or antifungal treatments in the case of fungal organisms affect genome maintenance. Indeed, genotoxic agents such as HU inhibit the ribonucleoside reductase, thus blocking the synthesis of dNTP and consequently DNA synthesis (305); rapamycin is an antibiotic, which acts on the TOR pathway by inhibiting its cellular proliferation function (306); and bleomycin, another antibiotic, generates single stranded breaks directly to the DNA (307). Heat shock, as outlined in the previous sections, induces genome instability and notably aneuploidies through a limitation of heat shock proteins that were found to play a role in kinetochore assembly (276); while antifungals, such as fluconazole, were proven to be at the origin of mitosis collapses caused by a delay in bud formation (277). In the case of DNA damaging agents, H_2O_2 mimics an oxidative stress, i.e. the production of reactive oxygen species that can generate lesions such as 8-oxoguanine, 5-methylcytosine, urea, uracil or abasic site at the origin of single-strand breaks (308). MMS, an alkylation agent, is also widely used in DNA repair studies as it adds methyl groups to guanine preferentially, inducing DNA breaks (309). Moreover, UV induces DNA lesions by triggering the formation of thymine dimers, thus inducing a torsion of the DNA molecule (310). Although these and many other agents have been extensively used in many organisms and for multiple studies, their action is inaccurate, generating stochastic lesions in the genome. Hence, for a more precise study of DNA repair mechanisms, the control of the localization, type and amount of DNA breaks is necessary.

2. *Genome engineering tools*

The genome editing field has made major advances over the last 30 years. Indeed, the modification of DNA through the action of an enzyme introducing a site-specific DNA DSB, triggering the DNA repair machinery, has brought significant progresses in genetics. In addition, these genome-engineering tools have the potential to overcome the most ambitious

therapeutic challenges, such as curing severe genetic diseases. These nucleases are reviewed in (311).

i. Meganucleases

Meganucleases, also called homing endonucleases, are classified into 5 groups but the LAGLIDADG family, named after the consensus amino acid sequence found mid-way in the protein sequence, is by far the best-studied family. The LAGLIDADG family is reviewed in (312). These proteins are intron-encoded selfish genetic elements that can be found in most prokaryotes (bacteria, archea) as well as in eukaryotes (plants, fungi). These enzymes have one function: once produced, the meganuclease facilitates the propagation of its own intron through the generation of a DNA DSB at an accurate localization. However, no biological role has been identified yet (312).

Meganucleases are rare-cutting enzymes as they recognize relatively long sequences (18-40 bp). A correct but modular base composition of the target site is necessary to ensure the distortion of the DNA double helix favoring its hydrolysis (312). The originality of these enzymes lies in that both recognition and cleavage functions are carried by a unique unit. Several teams have explored the engineering of new meganucleases, either by generating hybrid proteins (313) or by modifying the recognition site of the meganuclease by mutation combinations (314). Indeed, using the I-*Cre*I meganuclease, the authors could construct more than 100 new nucleases recognizing different targets. However, there are limitations in using these engineered endonucleases as their efficiency is not only driven by the efficacy by which they recognize their target sequence but also depends on the chromatin structure where the target site is localized (315).

The most widely used meganuclease is I-*Sce*I (316, 317). I-*Sce*I, a 28kDa meganuclease, is encoded by group I intron of the mitochondrial 21S rRNA gene. I-*Sce*I generates a DNA DSB at a 18 bp target sequence (5'-TAGGGATAACAGGGTAAT-3'), thus triggering gene conversions and intron propagation into intron-less copies of the 21S rRNA gene in the mitochondrial genome. I-*Sce*I has been used in *S. cerevisiae* (318, 319) and other organisms (for examples: (320, 321)) and has been fused to a nuclear localization sequence for efficient targeting to the nucleus. However, I-*Sce*I or other rare-cutting enzymes have limitations such as the need to integrate the target site into the genome, which requires efficient gene targeting.

ii. Zinc-Finger nucleases

One way to overcome the limitations imposed by the high specificity of the meganucleases is to separate the cleavage and recognition domains. The Type II S restriction enzyme *Fok*I has physically separated domains (311). The substitution of the DNA-binding domain by another sequence recognition domain allows generating nucleases with new substrate specificity and efficiency similar to that of *Fok*I (322). Zinc fingers are DNA-binding domains that can be found in a large number of transcription factors usually composed of 3 fingers, with each finger recognizing 3 nucleotides. Dimerization of the *Fok*I cleavage domain is necessary for its function (311). Thus, to facilitate its dimerization, two sets of 3 zinc fingers are tethered and spaced by 5 to 7 amino acids in order to allow the correct binding of DNA and its cleavage by the dimerized *Fok*I leaving a 4-5 bp long 5' overhang (Figure 13) (323).

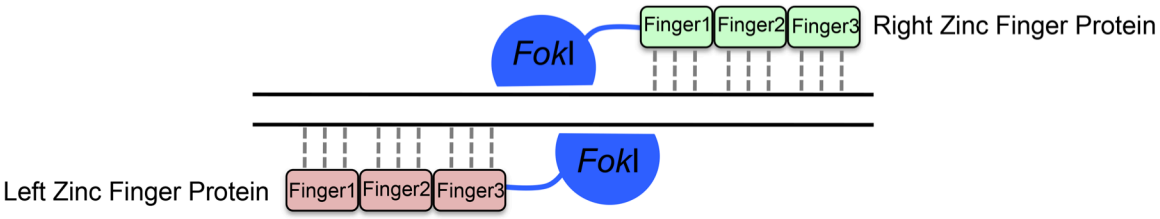


Figure 13: Zinc finger nucleases
Zinc finger nucleases are composed of 3 DNA-binding domains (fingers), with each finger recognizing 3 nucleotides, and a dimerized *Fok*I cleavage domain is necessary for its function. Figure adapted from (311)

Consequently, the addition of supplementary fingers allows modulating the size of the recognition sequence, until having a unique target site (311, 324). A study in Drosophila showed that DNA DSB triggered by zinc finger nucleases are mainly repaired by homologous recombination (HR), with a significant proportion that are repaired by end-joining (325). However, the purpose to generate a DNA DSB is often to modify the genome by stimulating HR. To circumvent this obstacle, zinc finger nucleases can be used in a cellular background where the genes involved in NHEJ are deleted.

Zinc finger nucleases have been used in clinical trials since 2009 in order to target the CCR5 gene to cure HIV: indeed, CCR5 encodes the main chemokine receptor used by HIV to enter and infect cells. The nucleases are delivered either by adenoviral vectors to T cells or upon electroporation of mRNAs into cells, both methods requiring the delivery of two different coding sequences for nuclease dimerization (326). Although zinc finger nucleases seem convenient, many zinc finger dimers fail to form (311, 327). In addition to failed zinc finger

dimers, off-target cleavage and limited target choice were reported and prompted a new hunt to find a more reliable tool.

iii. TALEN

The transcription-activator like effectors (TALE) were first identified as virulence factors in the Gram-negative plant pathogen, *Xanthomonas* (328), and orthologs have also been found in *Ralstonia solanacearum* (329). In *Xanthomonas*, the TALE are secreted through a type III secretion system into the plant, in which they are targeted to the nucleus thanks to a nuclear localization sequence. In the nucleus of the plant cells, the TALE alter the expression of the host genes (328).

The DNA binding domain of TALE is constituted of 34 amino acid repeats, but varying at the 12th and 13th positions, called repeat variable di-residues or RVD (Figure 14). Each RVD recognizes with a relatively high specificity one of the four bases (A, T, C or G) (311, 330). Therefore, the assembly of successive rearranged RVD would allow for the recognition of any sequence. As for zinc finger nucleases, TALE do not have a cleavage activity, thus requiring the fusion to the *Fok*I restriction domain (311, 331) (Figure 14). These engineered nucleases are called TALE-nucleases or TALEN. TALEN are made of two arms composed of 12 to 24 RVD, arranged in the opposing direction and linked by a spacer to allow *Fok*I dimerization (Figure 14).

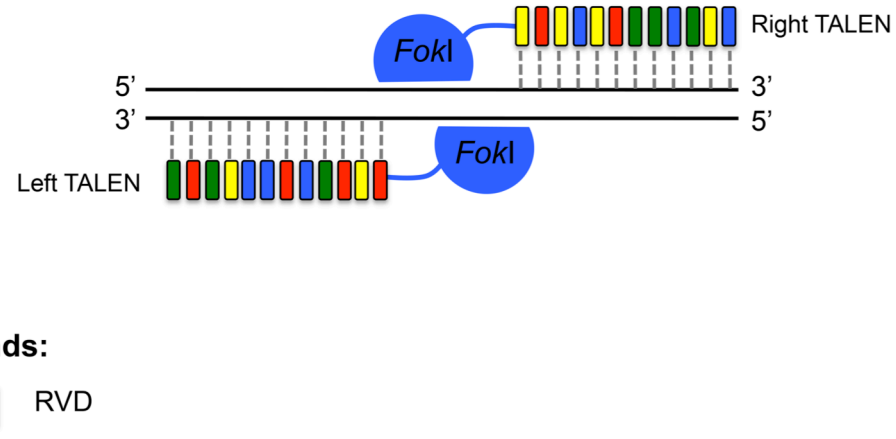


Figure 14: TALE nucleases
TALEN are made of two arms which carry 12 to 24 RVD, with each RVD recognizing one nucleotide and a *Fok*I core that generates a DNA DSB upon dimerization. Figure adapted from (311, 332).

In comparison to zinc finger nucleases, TALEN seem to be efficient and causing less cytotoxicity in human cells (333). Regarding the high potential of such a nuclease, Kim and

coworkers developed a library of TALEN able to target 18,740 protein-encoding genes and 274 miRNA coding sequences in the human genome (334). Furthermore, improving the efficiency of the DNA DSB repair by HR could help increasing the mutagenesis rate in the different organisms used to study DNA DSB repair or genome editing. Hence, the association of a TALEN with an exonuclease (Exo1 or Trex2) favors HR and limits NHEJ (335). In addition, due to the necessity for *Fok*I dimerization, the nuclease is twice bigger (about 3 kb per arm) than it could have been without dimerization. Recently, new TALEN were developed by fusing TALE with the cleavage domain of a meganuclease. By hybridizing TALE with I-*Ani*I (336), I-*Tev*I (337) or I-*Sce*I (338) cutting-domain, the off-target rate is reduced and the TALEN size smaller. However, these engineered nucleases necessitate cleaving sequences close to the natural recognition site of the meganuclease restriction domain.

Interestingly, TALEN were very recently used in the clinic to treat one patient (339) to allow the production of T-cells harboring universal chimeric antigen receptor, (CAR)19 T-cells, known to positively impact health of patients with leukemia (340). This new clinical trial suggests the high potential of TALEN to treat some diseases such as malignancies.

iv. CRISPR-Cas9

Initially identified as the immune system of bacteria and archea to fight against phages and plasmid transfers (341), CRISPR-Cas9 constitutes the major breakthrough of the last few years in the genome edition field. The CRISPR-Cas9 system is made of clustered regularly interspaced short palindromic repeats (CRISPR) of 20-50 bp separated by spacers (342), called protospacers, matching with viral DNA and serving as records for phage infection (343, 344) and immunity (345) mediated by CRISPR associated proteins (Cas). There are three types of CRISPR-Cas systems in bacteria/archea, among which the Type II system found in *Streptococcus pyogenes,* requires a unique Cas protein, Cas9, and has led to the development of a powerful genome editing tool (346) (Figure 15). Briefly, upon infection by a phage, viral DNA is injected inside the bacterium and detected by the cell. The foreign DNA, that is thought to undergo spontaneous DNA breaks following replication, is integrated into the CRISPR locus as spacers. This step is called immunization (347). The second step is named immunity (348). The CRISPR locus is transcribed and matured into short RNA molecules, crRNA, via complementarity with a transencoded crRNA (tracrRNA) (349). The crRNA serves as an antisense guide to the Cas9 protein thanks to its characteristic to match with the viral DNA. The crRNA also needs to carry a protospacer adjacent motif (PAM) to promote

Cas9 to distinguish the invader from the bacterium DNA and to bind its target (350). Once the DNA matches the crRNA sequence, the two cleavage domains of the Cas9 proteins cut each DNA strand from the bacteriophage (346), resulting in the non-propagation of the viral DNA.

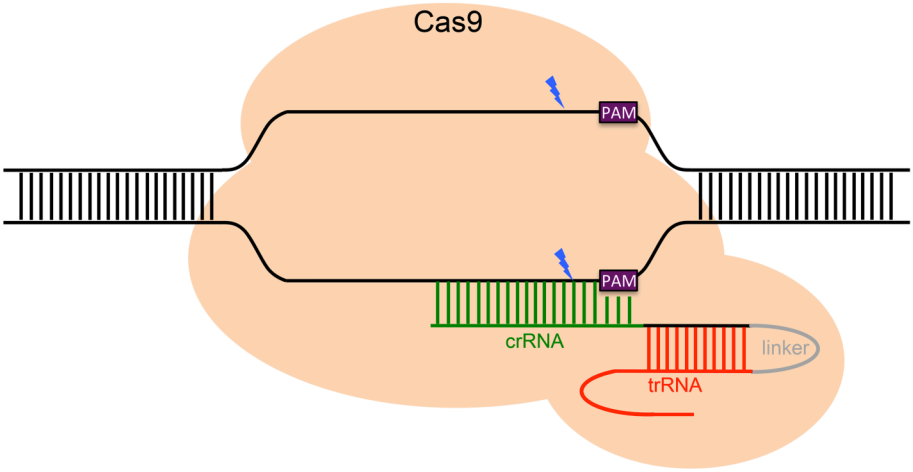


Figure 15: CRISPR-Cas9 – a revolutionary tool to edit genomes
CRISPR-Cas9 is an RNA-dependent DNA nuclease that generates a blunt DNA DSB (blue arrow) 3 nucleotides upstream a PAM motif (NGG) and requires a 20 bp long crRNA fused to a trRNA that serves as a guide RNA (sgRNA) that is complementary to a targeted genomic region. Figure adapted from (311, 332).

Recently, Jinek and colleagues modified the natural immune system of bacteria to adapt it to engineer genomes, a tool easily modulated, generating a blunt DNA DSB and that has very little functional requirements (346) (Figure 15). Indeed the CRISPR-Cas9 system that these authors developed made use of a chimeric RNA, fusing the crRNA and the tracrRNA into a secondary-structured single guide RNA (sgRNA), offering the advantage of working with a simple tool that can be easily adapted to various biological applications. The only requirement is the synthesis of a 20 bp DNA sequence that, once transcribed, serves as sgRNA that is complementary to a targeted region of interest in the genome (Figure 15). However, as aforementioned, the selected target region has to carry a PAM sequence (NGG in the case of the CRISPR-Cas9 system from *S. pyogenes*), impeding the use of this system to generate DNA DSB in trinucleotide repeats (351).

As compared to the other cited nucleases, CRISPR-Cas9 is unique by its capacity to bind DNA through a double interaction: (i) RNA-DNA and (ii) DNA-protein. This system is now under constant evolution to decrease at maxima the frequency of off-target cleavage (as reviewed in (352)) for an application in medicine and agriculture. The off-target issue can also be solved for example by the modification of Cas9 into a single-strand break nuclease:

thus, the use of two paired single-strand break-inducing Cas9 requires the use of two sgRNA, increasing the specificity associated with the break, but the need of two PAM sequence limits the choice of the target regions (353). In addition, a CRISPR-Cas9 system has been developed to target RNA instead of DNA (354), once again demonstrating the power of this tool. CRISPR-Cas9 is now available in a very large number of organisms, notably in *C. albicans*, allowing among other applications, a significantly facilitated construction of double-mutants that was previously knotty to make due to the absence of a haploid phase in this species (355).

v. CRISPR-Cas9 – Already out-dated?

On the same basis as CRISPR-Cas9 systems, new tools have been recently developed that make use of nucleic acid-guided nucleases, being alternatives to the CRISPR-Cas9 system. Up to now, two types of nucleic acid-guided nucleases have been published after the CRISPR-Cas9 breakthrough: (i) Cpf1, a single RNA-guided nuclease that lacks tracrRNA (356) and (ii) a programmed DNA-guided Argonaute that does not require a PAM sequence (357).

In more detail, Cpf1 can be classified as a type II CRISPR system, as it requires the use of a unique endonuclease in bacteria. Zetsche and collaborators (356) showed that (i) this CRISPR associated nuclease does not require a tracrRNA to have crRNA processed, (ii) when associated with crRNA, Cpf1 cleaves its target upstream of a short PAM rich in thymine and (iii) the cleavage by Cpf1 introduces a DNA DSB with a 4-5 bp long 5' overhang. By screening 16 Cpf1 protein families from various bacteria species, the authors identified two Cpf1 proteins from *Acidaminococcus* and *Lachnospiraceae* species as valuable tools to mediate genome editing in human cells. Indeed, the use of Cpf1 increases the chance to insert genes through the action of NHEJ by allowing a correct orientation of the gene (thanks to the generation of overhang ends), as HR is rare in non-dividing cells. Additionally, the T-rich PAM requirement can also be an advantage in organisms whose genome content is AT-rich (356).

The second discovery took advantage of the Argonaute protein families. Argonautes are well known for their role in RNA interference, a process that makes use of single-stranded RNA to guide Argonaute proteins to their complementary single-stranded RNA target. Interestingly, Argonautes have been recently identified as a natural immune system of bacteria shown to cleave different nucleic acid targets such as mobile genetic elements (358) or foreign DNA (359). Hence, Gao *et al.* (357) developed a programmable DNA-mediated DNA interference Argonaute protein from *Natronobacterium gregoryi* (NgAgo) to edit human genomes. The

two main advantages of this system are that no specific sequence is necessary to NgAgo to cut its target efficiently and while a RNA-RNA secondary conformation is necessary in the CRISPR-Cas9 system for correct hybridization with Cas9, no specific secondary structure is required with NgAgo. Moreover, this nuclease is highly specific as one mismatch in the target almost completely prevents the cleavage and the cleavage efficiency is the same as CRISPR-Cas9 system (about 32% of cut targets at the tested locus). With a fine-tuning, this tool can become a major player in genome engineering within the next years.

3. Tools to detect DNA DSB repair

To facilitate the study of DNA repair, many tools have been developed to detect repair events. In yeasts, the simplest way to analyze DNA repair mechanisms is to take advantage of auxotrophic markers. This technique is widely used in *in vitro* studies on genome stability. In *S. cerevisiae*, the insertion of auxotrophic markers can be associated with an HO-endonuclease recognition sequence, the nuclease necessary for *MAT* locus switching acting through the generation of a DNA DSB and leading to homozygosity at the *MAT* locus (360); or I-*Sce*I meganuclease target site (319). Auxotrophic markers can also be used to detect DNA DSB triggered by genotoxic, chemical or physical DNA damaging agents (361). Indeed, upon the induction of a DNA DSB, the presence of one or several auxotrophic markers in the heterozygous state will allow to detect if the repair has occurred and if so, to assess the extent of the repaired region. This technique has been widely used to demonstrate NHEJ or HDR repair mechanisms upon DNA DSB, for instance, through the appearance of auxotrophic cells (loss of an auxotrophy marker via the repair using a template that does not carry the marker) (for some examples: (200, 217, 319, 362, 363)). In addition, the loss of auxotrophic markers can be associated with counter-selection. Indeed, *URA3* encodes the orotine-5-monophosphate decarboxylase that allows the synthesis of uracil. However, in the presence of 5-fluoroorotic acid (5-FOA), strains harboring the *URA3* gene transform 5-FOA into a toxic component: 5-fluorouracil (268). Thus, upon loss of the *URA3* marker, 5-FOA is not toxic and cells can form colonies. This experiment allows making a rapid screen on a cell population by the counter-selection of events associated with the loss of the *URA3* gene following a DNA DSB repair. In *C. albicans*, counter-selection on 5-FOA or 2-deoxygalactose (2-DG) containing-medium (364) (*GAL1/gal1*Δ strains are sensitive to 2-DG while cells that have undergone a LOH at the *GAL1* locus and become *gal1*Δ*/gal1*Δ are 2-DG-resistant) have been used to assess the frequency of spontaneous versus induced DNA DSB (18-20, 150, 169).

Another method to detect LOH events following a DNA DSB uses SNP-Restriction Fragment Length Polymorphism (SNP-RFLP) analysis. This method makes use of the presence of heterozygous nucleotide polymorphisms that abolish a restriction site on one chromosome homolog. Hence, SNP-typing can only be used in a heterozygous or partially heterozygous (by using minichromosomes also known as chromosome fragments) organism. Forche and colleagues have developed this technique in *C. albicans*, mapping all SNP-RFLP sites in the genome of the laboratory *C. albicans* SC5314 strain (365), and this has been applied in a number of studies to assess the length of a LOH event (18-20).

Upon stress-induced DNA DSB, whole ploidy change can occur. Such events can be quantified by measuring whole DNA content, where DNA is stained with an intercalating dye and the quantity of DNA is measured by flow-cytometry (366). In the case of aneuploidy, the presence of an extra chromosome or the loss of a chromosome can be evaluated by diverse methods. For example, karyotyping allows the detection of gross chromosomal rearrangements by the separation of chromosomes by pulsed-field gel electrophoresis (PFGE) (367). An alternative to this technique is the use of a contour-clamped homogeneous electrophoretic field (CHEF) (368) that permits to separate the chromosomes by conserving a homogeneous potential across the electric field. Furthermore, comparative genome hybridization (CGH) can also be used to detect aneuploidies. This technique allows quantifying the copy number variation (CNV) of genes. To do so, two genomes are used, a test and a control, which are labeled with different fluorescent dyes (369). The genomes are hybridized and the relative fluorescent signal intensity allows the identification of CNVs. Although this technique has a limited resolution, array-based CGH (370) have been developed making use of DNA probes of different lengths covering the genome and the fluorescent dosage between a control and a test genome permits CNV detection. Nowadays, next generation sequencing is the best tool to visualize DNA repair, allowing the mapping of SNPs as well as the detection of CNV. However, although the price associated with the sequencing of one genome tends to decrease, the analysis of a large number of clones remains expensive.

In addition to the previous methods used to detect DNA repair, other techniques have been developed and take advantage of fluorescent markers to follow DNA repair events. For example, Loll-Krippleber and coworkers (20) (see appendix 1) developed a FACS-optimized LOH reporter system in *C. albicans* that relies on an artificial heterozygous locus on Chr4, consisting in the combination of flow cytometry and two fluorescent markers (Figure 11 p.53). While one homolog of Chr4 carries the gene encoding the Blue Fluorescent Protein (BFP),

the other homolog harbors the gene encoding the Green Fluorescent Protein (GFP). Cells that have undergone a LOH at this locus will express either the BFP or the GFP, thus allowing the LOH events to be detected and quantified by flow cytometry (Figure 11 p.53). Further characterization of the type of LOH event is completed by SNP-RFLP or whole genome sequencing (19, 20).

Another technique that is also used to follow the repair of a DNA break is the fusion of repair proteins with a fluorescent marker (For example: (371)). Hence, upon DNA DSB, the accumulation of repair proteins at the break site would form foci, easily detected by microscopy. The kinetics and intensity of the repair response can be thus measured by time-lapse experiments.

The previously cited tools are the most commonly used to detect DNA repair or genome rearrangements, notably in *C. albicans*; nevertheless, many other techniques exist, making yeasts valuable organisms to carry such studies.

E. The use of genetic screens to investigate genome stability in yeasts

Significant advances in the understanding of yeast biology have been made through genetic screens that used physical or chemical DNA damaging agents to generate random mutations in the genome. This approach allows screening for phenotypes of interest and is followed by the identification of the mutated gene that is responsible for the observed phenotype. However, because mutability varies among genes (variation that rely notably on gene size and base composition) and because the identification of the mutated gene is laborious, necessitating back-crossing between the mutated and the parental strains to clear the genome from unrelated mutations, random mutagenesis presents limitations that have been circumvented by the development of new approaches. Indeed, the construction of collections of mutants has led to the genome sequencing and the expansion of screens whereby each gene is individually tested for its construction to a phenotype possibly in a high-throughput manner. Indeed, in *S. cerevisiae*, several collections are available: deletion collections (372, 373), conditional mutant libraries (374, 375) and overexpression collections (376-378). These mutants can be used either in pools or alone and crossing of different mutants are possible, thus, greatly facilitating the study of the *S. cerevisiae* gene pool. Nowadays, next generation sequencing facilitates the study of the mutant(s) of interest allowing the characterization of genetic variation. Genome sequencing was also used to go back to the mutants performed by random mutagenesis and identify mutated genes responsible for a phenotype and that could not be characterized by the former methods. For review on genetic screens, see (379).

These approaches allowed us to raise questions on a large number of biological processes, and notably genome stability. Genetic screens by random mutagenesis have led to the identification of many genes that we described earlier (see section I.B p.28) and that play important roles in chromosome segregation, centromere and telomere maintenance, DNA repair and cell cycle, whose proper regulation is necessary to ensure the maintenance of genome integrity and cell viability. Therefore, we will discuss in more details, the use of mutant libraries to study genome stability in yeasts.

1. *Study of genome stability in Saccharomyces cerevisiae, using mutant collections*

Several studies made use of deletion collections to investigate the role of genes in genome dynamics. In Huang *et al.*, the authors used a collection of deletion mutants and investigated

the mechanism of resistance to cisplatin, a commonly used cytotoxic anticancer drug (380). They identified 20 additional genes to MMR genes that were associated with resistance upon deletion and are involved in mRNA catabolism, nucleotide metabolism, RNA Polymerase II-dependent gene regulation, vacuolar and membrane transports, cell wall, respiration, genome stability, and other uncharacterized genes. Many of the identified genes have mammalian orthologs, thus suggesting that this study might have found a new target to counter cisplatin resistance (380). Another group making use of the deletion mutants collection focused their work on determining genes involved in genome instability, increasing the number of LOH events upon knock-out (381). The authors identified 132 genes increasing LOH events when deleted. To follow the occurrence of LOH events, Andersen *et al.* made use of auxotrophic markers whose loss resulted in a change in the color phenotype of the colony. The researchers observed a bias on ChrXII, the chromosome carrying the rDNA repeats. Indeed, deletion of *SIR2*, responsible for the maintenance of genome integrity at rDNA loci, deletion of *RPA34*, *RPA14*, *RIF1*, *RLF2*, *NPT1*, *MSI1*, *LRS4* and *CSM1* known to reduce *SIR2* recruitment or activity, resulted in an increase in the LOH rate. Interestingly, the author could also identify new genes involved in genome integrity at rDNA repeats: deletion of *TOP1*, a topoisomerase, and *DUN1*, *CCR4*, *POP2*, *ADK1*, *GLY1* notably playing a role in the regulation of nucleotide pools were also associated with a high occurrence of LOH events at the rDNA region (381). However, the existence of essential genes and gene families limits the study of gene function and pathways. Thus, the overexpression or conditional expression constitutes an alternative or a complementary tool for the study of gene functions.

Indeed, to study essential genes, a conditional temperature-sensitive mutant collection was constructed (374). Such a collection provides a simple way to screen for essential gene function; thus allowing a fine-tuning of gene expression by enabling permissive (wild-type phenotype), semi-permissive and restrictive (growth defect or lethality-associated phenotype) conditions. In Li *et al.* (374), the authors focused on the regulation of mitotic spindle disassembly by combining temperature-sensitive mutants with high content microscopy-based analyses. This technique allows a high-throughput phenotypic screening of strains based on the spatial organization of the proteins of interest (harboring a fluorescent tag) in multiple mutant backgrounds. By applying this method, Li and colleagues demonstrated that cohesin and condensin were playing a role in the correct localization of the chromosome mid-zone via an interaction with the chromosomal passenger complex (CPC – reviewed in (382)), at the origin of proper mitotic spindle disassembly.

In addition, genetic interactions can be visualized when multiple alleles are mutated in the same strain and generate a phenotype that differs from the phenotype observed in the single mutant strain. Hence, since 18 million gene-gene combinations can be made, synthetic genetic arrays (SGA) were developed to automate the study of genetic interactions. It consists in generating a collection of haploid double mutants via mating and meiotic recombination, followed by replica pinning on different media. The phenotypes associated with the genetic interaction are quantified by the size of the colony, which is used to score the cellular fitness. Using overexpression collection, the SGA allows performing a high-throughput screening of synthetic dosage lethality (SDL). SDL occurs when the overexpression of a wild type gene is associated with lethality when occurring in a deletion or temperature-sensitive mutant background, while neither the overexpression alone nor the mutant strain exhibited such a phenotype. This approach helped for instance to uncover interactors of kinases, phosphatases or histone deacetylases. Bian and colleagues used this method to find genes whose deletion induces a synthetic lethality upon *MAD2* (mitotic arrest deficiency 2) overexpression (383). *MAD2* gene is a key player in the spindle checkpoint pathway, which ensures faithful chromosome segregation during mitosis (384). In human cells, Mad2 has been shown to prevent from the loss or gain of chromosomes during mitotic divisions (385). However, the overexpression of *MAD2* in mice led to chromosome rearrangements and aneuploidies (386). Additionally, *MAD2* was found to be overexpressed in numerous malignant tumors. SDL in *S. cerevisiae* identified 13 candidate genes that were subsequently studied in humans cells, and lethality was found associated with the deletion of one gene, *PPP2R1A*, ortholog of *ScTPD3* (383) and encoding a subunit of the protein phosphatase 2A (PP2A). This phosphatase participates in the regulation of the cell cycle by the dephosphorylation of key proteins. Hence, *PPP2R1A* can be an interesting target to treat tumors overexpressing *MAD2* as it was also shown to occur in 7% of the ovarian carcinomas (387).

2. *Candida albicans – what are the possible approaches to study genome dynamics?*

The genome of *C. albicans* carries 72% of uncharacterized genes, thus the study of gene function and the assessment of the number of essential genes need to be improved by the development of tools allowing large-scale studies. Random mutagenesis cannot be performed in *C. albicans* because of its ploidy and the absence of meiosis that does not allow backcrosses. Thus, the only methods that can be applied to study biological processes in this species are the generation of mutant collections. Indeed, over the last fifteen years, a large

effort has been made to develop collections of mutants to better characterize *C. albicans* cellular processes. Collections of heterozygous or homozygous deletion mutants are available and have been applied to study several aspects of *C. albicans* biology, notably, morphogenesis (388), biofilm formation (256, 389), virulence (390) or regulatory networks (22, 391). As compared to *S. cerevisiae*, the homozygous disruption of a gene is laborious, requiring a two-step gene inactivation.

In *C. albicans*, Noble and Johnson generated a collection of homozygous knock-out mutants, first by constructing parental strains with a combination of auxotrophies for *HIS1*, *ARG4* and *LEU2* markers (266). The collection gathered 666 mutants and is available to the *C. albicans* community. Moreover, derived from the parental strains constructed by Noble and Johnson, a collection was generated to study regulatory networks in *C. albicans* by the construction of 317 mutant strains (22). However, the study of deletion mutants in *C. albicans* is limited by the heavy work required for the generation of homozygous double mutants and the needed verification as the genome of *C. albicans* is highly plastic.

Additionally, as in *S. cerevisiae*, the study of essential genes is hampered by the generation of deletion mutants. Thus, techniques have been developed to allow their study through the generation of temperature sensitive mutants, inducible mutants and reduced expression mutants (392, 393). However, no large-scale screen was made with these techniques. Another technique mixing gene disruption with an auxotrophic marker and the use of conditional promoter was set up by Roemer and colleagues and was named Gene Replacement And Conditional Expression (GRACE) (394). The disruption of the first allele is mediated by gene replacement with a *HIS3* marker flanked by two unique barcodes while the second copy of the gene was placed under the control of a tetracyclin inducible promoter (P_{TET}). The TET-OFF system was used here: in absence of a tetracyclin derivative, the constitutive expression of the gene of interest is achieved; upon addition of a tetracycline derivative, the association between the transactivator and P_{TET} is disrupted, leading to the weakening or complete loss of the gene function. The collection gathered 1,152 mutants.

In addition, a method has been set up in *C. albicans* making use of transposons for the construction of heterozygous deletion strains, leading to haploinsufficiency. Even if SGA and SDL screens cannot be performed in *C. albicans* due to the lack of stable haploid strains and meiosis, synthetic genetic studies were made. Uhl *et al.* initiated this approach by generating a mutant collection carrying 18,000 transposon insertions and 146 genes were found to play a role in the morphogenetic switch (yeast-to-hyphae transition), and among them no ortholog in *S. cerevisiae* was found for about 50 genes (395). This technique was implemented by

Bharucha *et al.* who screened a transposon-mediated collection in a mutant background and looked for synthetic phenotypes (396). Despite the encouraging data obtained from these screens, haploinsufficiency limits the obtention of interesting phenotypes that could have been observed upon drastic gene expression modification.

Moreover, replicative plasmids in *C. albicans* have been reported as unstable, thus not suitable for overexpression studies (397). Hence, no matter the technique used, every large-scale study has to go through integrative transformation steps. Overexpression was used at small (393, 398), medium (399-401) and large scales (402) in *C. albicans*. Indeed, in collaboration with the Munro group, the d'Enfert laboratory developed the ORFeome project that consists in the construction of a modulable collection of overexpression plasmids and a collection of overexpression strains encompassing over 85% of the 6,198 ORFs of *C. albicans*. This project was initiated by the establishment of a primary collection of strains overexpressing genes encoding proteins involved in signaling pathways (402). Several screens were performed to study morphogenesis, virulence and biofilm formation using this collection or part of this collection (402-404).

Up to now, only one study made use of a strain collection to uncover new genes involved in genome instability in *C. albicans*. In Loll-Krippleber *et al.*, a collection of 124 overexpression mutants for genes involved in DNA processes such as replication, repair and recombination was constructed using an inducible promoter (20) (see appendix 1). This collection was associated with the LOH reporter system that was presented in the above sections (Figure 11 p.53). This system allows high-throughput studies of LOH events and their detection by flow cytometry through the loss of fluorescent markers. By doing so, *CDC20*, *RAD53*, *BIM1* and *RAD51* were identified as major players in genome instability. Additionally, as the mutants that had undergone a LOH event could be sorted, we could assess the molecular mechanisms at the origin of LOH: while both BIR and chromosome loss were observed upon *BIM1*, *RAD51* and *RAD53* overexpression, LOH events occurring upon *CDC20* overexpression, a gene enabling the metaphase-anaphase transition, were mainly the results of chromosome loss (20) (see appendix 1). This study is the first to have used this approach to demonstrate the effect of overexpression on LOH and extending the study to a larger set of overexpression mutants might highlight signaling pathways involved in genome instability.

F. Outlooks

LOH and aneuploidies have been reported in human cells and associated with diseases when occurring in dividing cells. Indeed, LOH can unmask recessive deleterious alleles or mutations in promoter regions whose homozygosis results in the misregulation of some biological processes, either by depleting a tumor-suppressive gene or enhancing a proto-oncogene; thus leading to cancer. A well-known instance of LOH-mediated cancer is the case of retinoblastoma. One of the causes leading to this form of cancer is a mutation in the *RB1* gene followed by a LOH (405). *RB1*, a tumor-suppressor gene, regulates more than hundreds of genes, and among them transcription factors that control the transcription of genes required in S-phase. Cancer cells are also characterized by aneuploidies, as a change in the gene copy number is accompanied by an up- or a down-regulation of the targeted genes, thus altering their function. Aneuploidies and LOH also arise from the treatment of tumor cells by chemotherapy, resulting in resistance.

C. albicans is a commensal yeast but is also classified as the major fungal pathogen of humans. *C. albicans* raises major concerns in the context of disseminated infections, and their treatment can be associated with drug resistance, often resulting from LOH and aneuploidies (270, 277, 281, 292, 293, 299). Given the importance played by heterozygosity, LOH and chromosome ploidy, *C. albicans* has become an organism of captivating interest to study genome stability. Despite the fact that no NHEJ pathway has been described in *C. albicans* (171, 172), the homologous recombination pathways are very efficient and are responsible for most of the LOH events detected within the natural population of *C. albicans*, at the origin of intra-species variations (18-20). The number of molecular tools that have been developed in *C. albicans* allows extensive in-depth work to be performed in order to elucidate the molecular mechanisms involved in the control of LOH events, leaving us with important and unsolved questions such as: how does *C. albicans* respond to a site-specific DNA DSB? How efficient and faithful is the DNA repair machinery in *C. albicans*? Does NHEJ exist in *MTL* homozygous cells? Do signaling pathways control the repair of a DNA DSB? Can we develop new high-throughput tools to facilitate the study of the *C. albicans* biology? Can we characterize more precisely the genomic population structure of the *C. albicans* species?

Although a lot remains to be discovered and new tools are needed to extend genetic studies to a large scale, using *C. albicans* as a model to understand the mechanisms regulating genome dynamics may bring new insights to genome instability in eukaryotes.

II. Results and discussion

In the previous section, I have highlighted several studies depicting the importance of the maintenance of genome integrity in human and yeast cells. In addition, the same studies shed light on the beneficial role played by some genomic rearrangements in defined circumstances. *Candida albicans* is a diploid yeast, whose genome is highly plastic and characterized by aneuploidies and relatively frequent loss-of-heterozygosity (LOH) events. The ability of *C. albicans* to undergo large-scale genome rearrangements and its apparent tolerance to such changes are likely to be crucial for its survival upon exposure to new environments or for proper colonization of new niches. LOH events are likely to result from segmental or total chromosome loss or recombination events such as Break Induced Replication (BIR), gene conversion (GC) or mitotic crossover (MCO) following a DNA DSB. Yet, very little is known about the mechanisms controlling these events in *C. albicans*.

In this context, my PhD work was organized following two axes: up to now, to study genome instability in *C. albicans*, DNA breaks were generated using either physical or chemical DNA damaging agents, at the origin of random location and uncontrolled amount of DNA DSB in the genome, thus biasing the interpretation of the DNA DSB repair outcomes (i.e. LOH events). Hence, we carried (A) (i) a study to better characterize the molecular mechanisms involved in the repair of a targeted DNA DSB in the *C. albicans* genome, which led to the identification of recessive deleterious alleles in this species and (ii) a high resolution characterization of repair mechanisms at the site of a targeted DNA DSB in the genome of *C. albicans*, which offers a sufficient number of SNPs to follow recombination and mutation events during repair.

Very few genes were identified as playing a role in genome dynamics in *C. albicans*, due to the difficulty to achieve knock-out mutants, thus (B) we aimed at identifying such genes with an original approach and based on the study made by Forche *et al.* (18) who showed that different stresses trigger different LOH outcomes, we led investigations to characterize signaling pathways that can be involved in genome rearrangements in *C. albicans*.

A. Analysis of repair mechanisms following an induced double-strand break uncovers recessive deleterious alleles in the *Candida albicans* diploid genome.

1. Context and aim of the work

DNA double-strand breaks (DSBs) have been shown to be very potent initiators of recombination in yeast and other organisms (406, 407) and the use of DSB repair assays based on rare-cutting endonucleases has been applied to many organisms (408-410). Until recently, the study of genome instability in *C. albicans* has used genotoxic agents or physical stresses in order to generate DNA damages (for examples: (18-20)) This approach is limited by the lack of accuracy in the nature of the lesions and their location. In addition, LOH events have generally been screened through loss of auxotrophy or more often using the *GAL1* marker (364). In the latter case, *GAL1/gal1*Δ strains are sensitive to 2-deoxygalactose (2-DG) while cells that have undergone a LOH at the *GAL1* locus and become *gal1*Δ*/gal1*Δ are 2-DG-resistant. While convenient, this LOH-monitoring system introduces a selective pressure as it requires spreading cells on selective medium.

In order to circumvent the constraints imposed by the available tools, we have developed a DNA DSB-inducing system, using the *Saccharromyces cerevisiae* I-*Sce*I meganuclease that we adapted to the *C. albicans* codon usage. I-*Sce*I has already been used to generate DNA DSB *in vivo* in multiple organisms (320, 321) and its 18 bp long recognition site is not found in the *C. albicans* genome. This DSB-inducing system was combined to a LOH reporter system already available in the laboratory (see section I.D.3 p.63 and Figure 11 p.53). This system relies on the construction of an artificial heterozygous locus on Chr4 at the *PGA59-PGA62* locus: while one homolog of Chr4 carries the gene encoding the Blue Fluorescent Protein (BFP), the other homolog harbors the gene encoding the Green Fluorescent Protein (GFP). Hence, LOH events can be detected by flow cytometry since a cell that has undergone a LOH at this locus will express only the BFP or only the GFP (20) (See appendices 1 and 2).

The manuscript presented below presents (i) the development and validation of an I-*Sce*I-dependent DSB-inducing system that allows generating a targeted DNA DSB in the genome of *C. albicans* and (ii) the characterization of the molecular mechanisms involved in DNA repair upon the induction of a targeted DNA DSB in the *C. albicans* genome. As noted above, a bias towards homozygosity of one of the two haplotypes of *C. albicans*

chromosomes has been previously reported, notably on Chr4 (10, 20, 74, 126, 129, 300). This supports the presence of recessive lethal alleles on chromosome homologs that cannot be found in the homozygous state. Hence, in this manuscript, we also present (iii) the identification of recessive deleterious allele(s) responsible for unidirectional LOH events observed on Chr4 in the reference laboratory strain, SC5314.

Finally, in a final section, I present a detailed characterization of the genomic rearrangements occurring upon repair of an I-*Sce*I DNA DSB by BIR or GC with CO. This characterization raises hypotheses regarding the functioning of repair pathways such as BIR and GC with CO in *C. albicans* with possible relevance to other organisms such as *S. cerevisiae*.

2. *Research article*

This article is *accepted for publication* in mBio.

Analysis of repair mechanisms following an induced double strand break uncovers recessive deleterious alleles in the *Candida albicans* diploid genome

Adeline Feri[a,b], Raphaël Loll-Krippleber[a,b,*], Pierre-Henri Commere[c], Corinne Maufrais[d], Natacha Sertour[a], Katja Schwartz[e], Gavin Sherlock[e], Marie-Elisabeth Bougnoux[a,f], Christophe d'Enfert[a,§] and Mélanie Legrand[a,§]

Institut Pasteur, INRA, Unité Biologie et Pathogénicité Fongiques, Paris, France[a]; Univ. Paris Diderot, Sorbonne Paris Cité, Cellule Pasteur, rue du Docteur Roux, Paris, France[b]; Institut Pasteur, Imagopole, Plate-Forme de Cytométrie, Paris, France[c]; Institut Pasteur, Centre d'Informatique pour la Biologie, Paris, France[d]; Department of Genetics, Stanford University, Stanford, CA 94305-5120, USA[e]; Unité de Parasitologie-Mycologie, Service de Microbiologie clinique, Hôpital Necker-Enfants-Malades, Assistance Publique des Hôpitaux de Paris (APHP), Univ. Paris Descartes, Paris, France[f]; [§]Corresponding authors: Mélanie Legrand & Christophe d'Enfert, Institut Pasteur, Unité Biologie et Pathogénicité Fongiques, Département Mycologie, 25 rue du Docteur Roux, F-75015 Paris, France ; Phone : +33 (1) 40 61 31 26 ; E-mail : melanie.legrand@pasteur.fr, christophe.denfert@pasteur.fr

[*]Present address: Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada

Running title: Recessive lethal alleles and DSB repair in *C. albicans*

Abstract

The diploid genome of the yeast *Candida albicans* is highly plastic, exhibiting frequent loss-of-heterozygosity events. To provide a deeper understanding of the mechanisms leading to loss-of-heterozygosity, we investigated the repair of a unique DNA double-strand break in the laboratory *C. albicans* SC5314 strain using the I-*Sce*I meganuclease. Upon I-*Sce*I induction, we detected a strong increase in the frequency of loss-of-heterozygosity events at an I-*Sce*I target locus positioned on chromosome 4, including events spreading from this locus to the proximal telomere. Characterization of the repair events by SNP-typing and whole genome sequencing revealed a predominance of gene conversions but we also observed mitotic crossover or break-induced replication events, as well as combinations of independent events. Importantly, progeny that had undergone homozygosis of part or all of chromosome 4 haplotype B were inviable. Mining of genome sequencing data for 155 *C. albicans* isolates allowed the identification of a recessive lethal allele in the *GPI16* gene on chromosome 4 haplotype B unique to *C. albicans* strain SC5314, which is responsible for this inviability. Additional recessive lethal or deleterious alleles were identified in the genomes of strain SC5314 and two clinical isolates. Our results demonstrate that recessive lethal alleles in the genomes of *C. albicans* isolates prevent the occurrence of specific extended loss-of-heterozygosity events. While these and other recessive lethal and deleterious alleles are likely to accumulate in *C. albicans* due to clonal reproduction, their occurrence may in turn promote the maintenance of corresponding non-deleterious alleles and, consequently, heterozygosity in the *C. albicans* species.

Importance

Recessive lethal alleles impose significant constraints on the biology of diploid organisms. Using a combination of I-*Sce*I meganuclease-mediated DNA double-strand break, a FACS-optimized reporter of loss-of-heterozygosity and a compendium of 155 genome sequences, we were able to unmask and identify recessive lethal and deleterious alleles in isolates of *Candida albicans*, a diploid yeast and the major fungal pathogen of humans. Accumulation of recessive deleterious mutations upon clonal reproduction of *C. albicans* could contribute to the maintenance of heterozygosity despite the high frequency of LOH events in this species.

Introduction

Candida albicans is a quasi-obligate diploid yeast (10) whose 32Mb genome is organized in eight pairs of chromosomes, with, on average, one heterozygous position every ~250 bp (11, 17, 123). Genomic studies have shown that the *C. albicans* genome displays a high degree of plasticity. Indeed, aneuploidies, gross chromosomal rearrangements and loss-of-heterozygosity events (LOH) of variable length and location were observed in both commensal and clinical isolates and upon commensalism or passage of a *C. albicans* laboratory strain in animal models (15, 279). Importantly, the ability of *C. albicans* to undergo genome rearrangements and its apparent tolerance to such changes can be critical for its survival upon exposure to changing conditions, such as antifungal treatments (16, 17, 286, 411). In this respect, LOH events contribute to the expansion of hyperactive mutations leading to antifungal resistance (281, 293, 294, 299, 412). More generally, allelic differences within a *C. albicans* strain can result in variations in gene expression, protein production or function (12, 413). Hence, LOH events in *C. albicans* have been associated with phenotypic variation, such as amino acid auxotrophy or drug sensitivity (129, 300), white-opaque switching upon mating-type like locus homozygosis (58, 59, 414), and adaptation to growth on alternative carbon sources (57).

LOH events can arise from the mechanisms used by *C. albicans* in response to DNA double strand breaks (DSBs), or can be the consequence of chromosome non-disjunction events during mitosis. Repair of DNA DSBs by gene conversion without crossover (GC) explains short-range LOH, while repair by either Break Induced Replication (BIR) or mitotic crossover (MCO) leads to LOH that extends from the DNA DSB site to the telomere. In the absence of DNA DSB repair or upon chromosome non-disjunction, segmental or whole chromosome losses (SCL/WCL) are observed and the loss of a chromosome is often followed by a reduplication event (21). Interestingly, LOH events in strains derived from the *C. albicans* SC5314 strain, from which the reference genome sequenced is derived, appear to be biased towards one of the two haplotypes for several chromosomes. For instance, Forche *et al*. (74) have observed that homologous recombination-mediated LOH in progeny resulting from the *C. albicans* parasexual cycle (72), had a strong bias towards one of the two haplotypes for chromosomes R, 2, 4, 6 and 7. A similar bias was observed in a *C. albicans rad52Δ/rad52Δ* mutant that shows an increased frequency of spontaneous unidirectional LOH (126). The recent finding that *C. albicans* can exist in a haploid form also led to the observation that one of the two haplotypes for chromosomes 3, 4, 6 and 7 and for most of chromosome 1 is never observed in the homozygous state under laboratory growth conditions (10). Finally, an

investigation of the events associated with LOH at a specific locus on *C. albicans* chromosome 4 (Chr4) revealed that chromosome loss events leaving haplotype B as the sole remaining haplotype were never observed (20). Taken together, these studies have led to the hypothesis that recessive lethal alleles are present on *C. albicans* chromosome homologs and prevent some LOH events from being detected.

DNA DSBs have been shown to be very potent initiators of recombination in yeast and other organisms and consequently of LOH (406, 407). The mechanisms by which DNA DSBs are repaired can greatly influence the nature of the LOH events that affect the *C. albicans* genome. In this respect, genotoxic agents have been used to trigger DNA breaks (305, 309, 415) and study DNA repair mechanisms in *C. albicans* (150, 169, 172, 207, 416). However, the use of genotoxic agents (18, 20, 169, 417) or physical or chemical stresses known to induce LOH (18, 20, 418) do not allow precise control of the nature or the location of a DNA break. In order to circumvent this limitation, DNA DSB repair assays based on rare-cutting endonucleases such as I-*Sce*I have been developed in many organisms (320, 408-410, 419-422).

Here, we show how the combination of I-*Sce*I-induced DNA DSB and a recently developed LOH reporter system (20) allowed us to precisely study the mechanisms involved in DNA DSB repair at a specific genomic location in the *C. albicans* genome. Importantly, our detailed analysis of LOH events resulting from an induced DNA DSB in strain SC5314 allowed us to identify recessive deleterious alleles on the *C. albicans* Chr4 haplotype B that explain why the haplotype A for this chromosome cannot be lost. Furthermore, we have expanded this work to clinical isolates of *C. albicans* by the identification of recessive lethal alleles on Chr5. Taken together our results suggest that recessive deleterious alleles could play a role in the maintenance of heterozygosity in the *C. albicans* species.

Results

Development and validation of an I-SceI-dependent DNA DSB-generating system in Candida albicans

To study the mechanisms involved in the repair of a single DNA DSB in *C. albicans*, we took advantage of the *I-SceI meganuclease*, an intron-encoded homing endonuclease isolated from the yeast *Saccharomyces cerevisiae*. I-*Sce*I recognizes an 18bp-long sequence (316, 317) *absent from* the *C. albicans* genome. We also used an LOH reporter system located at the *PGA59-PGA62* locus on Chr4 (19, 20) that consists of the combination between flow cytometry and two fluorescent markers (Figure 11A p.53).

Briefly, while one homolog of Chr4 carries the gene encoding the Blue Fluorescent Protein (BFP), the other homolog harbors the gene encoding the Green Fluorescent Protein (GFP) (Figure 11A p.53). Hence, LOH events can be detected by flow cytometry, as cells that have undergone an LOH at this locus will express either the BFP or the GFP (Figure 11B p.53). Further characterization of the LOH can be achieved either by SNP-typing or whole genome sequencing (WGS) after cell sorting (20, 365).

We thus generated a *C. albicans* strain that carries (i) a tetracycline-inducible, codon-optimized gene encoding the rare-cutting endonuclease I-*Sce*I modified to harbor a SV40 nuclear localization signal (NLS) (320, 423-425), (ii) a gene encoding a tetracycline-dependent transactivator (402), (iii) the I-*Sce*I target sequence along with the *URA3* marker on the left arm of Chr4, and (iv) the FACS-optimized LOH reporter system with the *BFP* gene linked to the *HIS1* gene on the left arm of Chr4 haplotype B (Chr4B) which also bears a functional allele of the *HIS4* gene closer to the telomere (129); and the *GFP* gene linked to the *ARG4* gene on the left arm of Chr4 haplotype A (Chr4A) which bears the non-functional $his4^{G310V}$ allele closer to the telomere (Figure 16) (19, 20, 129).

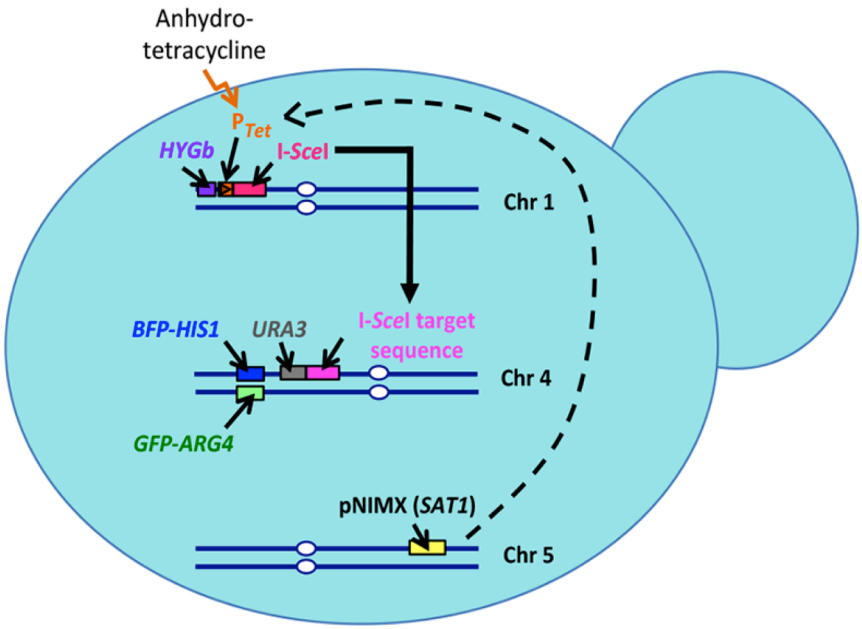


Figure 16: The DNA DSB-inducing system.
The double strand break inducing system consists of (i) the gene encoding the rare-cutting endonuclease, I-*Sce*I, placed under the control of the tetracycline inducible promoter (P_{TET}) integrated at the *XOG1-HOL1* locus on Chr1, (ii) the gene encoding the tetracycline-dependent rtTA transactivator of the P_{TET} promoter placed at the *ADH1* locus on Chr5, and (iii) the I-*Sce*I target sequence integrated at the *CDR3-tG(GCC)2* locus on Chr4 on haplotype A or B, between the centromere and the FACS-optimized reporter system of LOH at the *PGA59-PGA62* locus. Upon binding of anhydrotetracycline to the rtTA transactivator, the I-*Sce*I gene is expressed, the endonuclease is directed to its target sequence generating a DNA DSB which can be in particular repaired by mitotic crossover or break-induced replication yielding mono-fluorescent cells that are detected by flow cytometry.

In this setting, the I-*Sce*I target sequence is ~215kb distant from the Chr4 centromere, while the LOH reporter system is ~300kb further towards the telomere (Figure 17A p.80). The resulting strain is referred to as "I-*Sce*I+TargetB" as WGS showed that the I-*Sce*I target site and the *URA3* gene were inserted on Chr4B. WGS also showed that the "I-*Sce*I+TargetB" strain had not experienced gross chromosomal rearrangements (aneuploidies, LOH) upon the successive transformation events needed for its construction (data not shown). Control strains lacking the I-*Sce*I gene or the I-*Sce*I target sequence were referred to as "Target only" and "I-*Sce*I only", respectively, and were used to assess the occurrence of I-*Sce*I-independent DNA DSBs at the I-*Sce*I target site or I-*Sce*I-induced Chr4 off-target DNA DSBs.
In the "I-*Sce*I+TargetB" strain, induction of the I-*Sce*I gene through addition of a tetracycline derivative should result in I-*Sce*I endonuclease production and targeting to the nucleus,

followed by the generation of a DNA DSB at the I-*Sce*I target sequence (Figure 16). DNA DSBs can be repaired either by gene conversion (GC) thus leading to double-fluorescent cells that have lost the *URA3* gene and are 5-FOA-resistant (268), or by Break Induced Replication (BIR)/Mitotic Crossing Over (MCO) leading to the loss of the *URA3* gene and BFP reporter and thus, appearance of 5-FOA-resistant, arginine prototroph, histidine auxotroph, mono-GFP cells (19, 20) (Figure 17A p.80). WCL/SCL should also lead to progeny with uridine and histidine auxotrophies and GFP fluorescence (Figure 17A p.80). Importantly, since the *HIS1* gene is linked to the *BFP* gene and all mono-GFP cells should be histidine auxotrophs, an unexpected crossover between the BFP/GFP locus and the heterozygous *HIS4/his4*^{G310V} locus should not impact the phenotypes of cells that have undergone BIR/MCO, SCL or WCL (Figure 17A p.80).

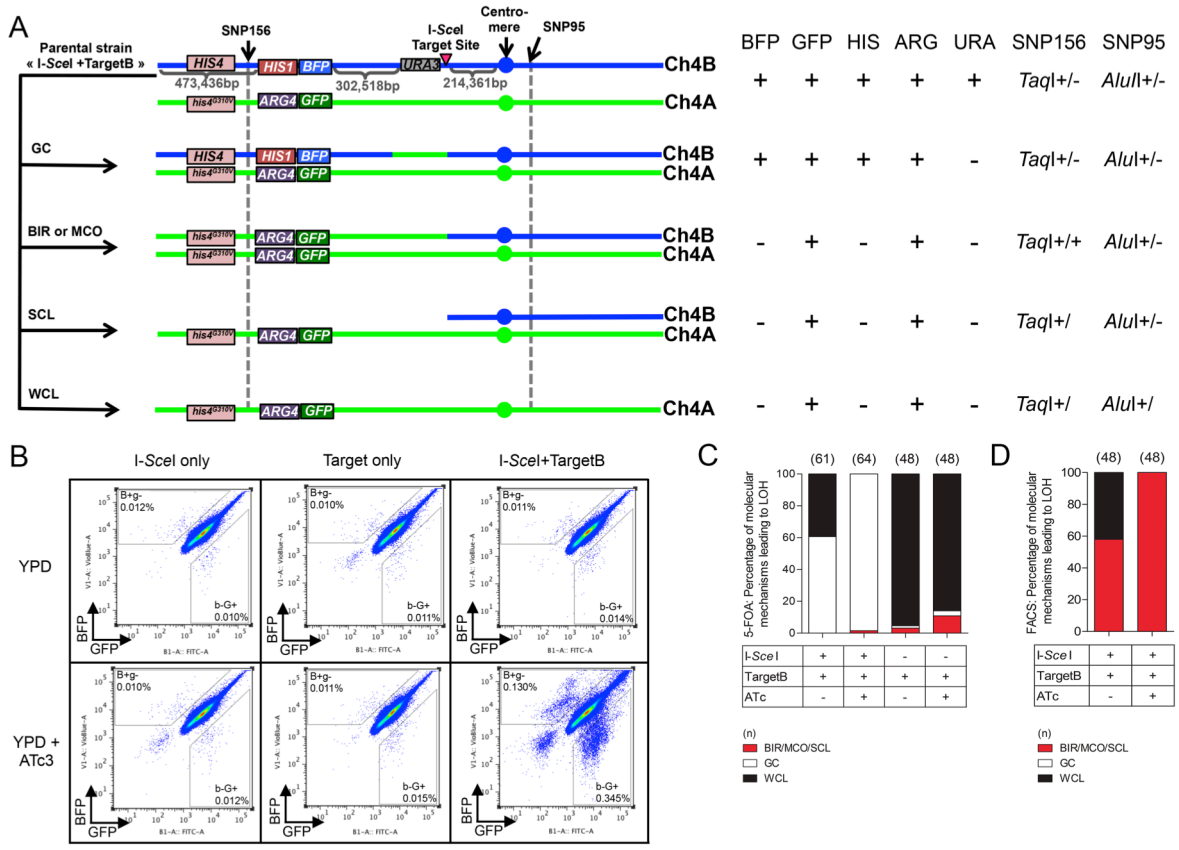


Figure 17: Effect of I-*Sce*I induction in « I-*Sce*I + TargetB » and control strains.
(A) Different LOH events on Chr4B can arise upon I-*Sce*I induction. The heterozygous SNPs used for RFLP characterization are indicated with black arrows. SNP156, close to the left arm telomere (at position 367,295 on Chr4A and 367,352 on Chr4B), is part of a *Taq*I restriction site while SNP95, on the right arm of Chr4 and close to the centromere (at position 1,310,251 on Chr4A and 1,310,274 on Chr4B), is located in an *Alu*I restriction site. One haplotype carries the restriction site while the other does not. The combined heterozygosity or homozygosity of these SNPs gives insights about the molecular mechanisms leading to LOH events. The *HIS4* gene presenting a non-functional allele on haplotype A as shown in Gomez-Raja *et al*. (2008) is also represented here. GC: gene conversion, BIR: break-induced

replication, MCO: mitotic crossover, SCL: segmental chromosome loss, WCL: whole chromosome loss. BIR/MCO or SCL events are indistinguishable. (B) Co-occurrence of I-*Sce*I and its target sequence triggers a predominant increase in mono-GFP cells. The cultures were analyzed on a MACSQuant cytometer. 10^6 events are displayed. The B+g- and b-G+ gates were defined arbitrarily. (C and D) SNP-RFLP analysis shows that I-*Sce*I dependent DNA DSBs on Chr4B are mainly repaired by GC. Histograms presenting the proportion of BIR/MCO or SCL, GC and WCL events in the population having undergone a LOH either recovered from 5-FOA counter-selection (C) or from FACS (D). BIR/MCO or SCL events correspond to mono-GFP cells that displayed a homozygous SNP156 but have maintained a heterozygous SNP95. WCL events correspond to mono-GFP cells in which both SNP95 and SNP156 became homozygous. GC events correspond to doubly fluorescent cells in which both SN156 and SNP95 remained heterozygous. L: large-sized colonies and S: small-sized colonies.

To validate the functionality of the I-*Sce*I system, the "I-*Sce*I+TargetB" and the control strains were grown in the presence or absence of anhydrotetracycline (3 μg/ml; ATc) and plated on YPD and 5-FOA agar plates (20). For both control strains, no increase in the number of 5-FOA^R colonies was observable between the non-induced and induced conditions, with a rate below 1.0 x 10^{-7} events/cell/generation in both conditions (Table 2). By contrast, I-*Sce*I expression yielded a 372-fold increase in the rate of 5-FOA^R colonies appearance as compared to the non-induced condition when using the "I-*Sce*I+TargetB" strain (Table 2).

Table 2 – 5FOA resistance quantification on Chr4

Strains	Growth conditions	5FOA^R acquisition rate ($\times 10^{-8}$)[1] (events/cell/generation)	Fold change[1]
I-*Sce*I + TargetB	YPD	11	**372**
	YPD + ATc	4,100	
I-*Sce*I + TargetA	YPD	0.4	**2,450**
	YPD + ATc	980	
I-*Sce*I + TargetA + *GPI16*	YPD	2.44	**574**
	YPD + ATc	1,400	
I-*Sce*I only	YPD	6.1	**1.2**
	YPD + ATc	7.4	
Target only	YPD	6.3	**0.5**
	YPD + ATc	3.4	

[1]Values are representative of 2 independent experiments

The "I-*Sce*I+TargetB" and the control strains were also grown for 8 hours in the presence or absence of ATc and analyzed by flow cytometry. This allows detecting long-range LOH events only, *i.e*. BIR, MCO, SCL or WCL. As expected, a 30-fold ATc-dependent increase in mono-GFP frequency was observed for the "I-*Sce*I+TargetB" strain and no change was detected in the controls (Figure 17B p.80, Table 3 p.83), consistent with the I-*Sce*I recognition sequence being located on the BFP-bearing chromosome, Chr4B. Additionally, an increase was noticeable in the number of non-fluorescent cells, likely to be dead cells, as previously shown in Loll-Krippleber *et al*. (20). Strikingly, we also observed a 17-fold increase in the frequency of appearance of mono-BFP cells in the ATc-treated "I-*Sce*I+TargetB" cells only (Figure 17B p.80; Table 3 p.83). The basis for this unexpected population of mono-fluorescent cells will be revisited below.

Taken together, these results indicated that I-*Sce*I is functional and induces a target specific DNA DSB in *C. albicans*. In addition, the different increase in frequency of 5-FOA^R (372-fold, including both long and short-range LOH events) and mono-GFP (30-fold, including long-range LOH events only) cells upon induction of the I-*Sce*I gene suggested that long and short-range LOH events occur at different frequencies.

Table 3 – LOH quantification on Chr4 by flow cytometry

Strains	Cell population	Growth conditions	N[1]	LOH frequency ± SEM (×10^{-4})	Fold change	Mann-Whitney test P-value
I-*Sce*I + TargetB	**mono-GFP**	YPD	37	1.35 ± 0.1	**30**	**≤ 0.0001**
		YPD + ATc	40	40.0 ± 0.9		
	mono-BFP	YPD	37	1.0 ± 0.1	**17**	**≤ 0.0001**
		YPD + ATc	40	16.8 ± 0.6		
I-*Sce*I + TargetA	**mono-GFP**	YPD	36	2.0 ± 0.1	**9**	**≤ 0.0001**
		YPD + ATc	36	18.7 ± 0.3		
	mono-BFP	YPD	36	1.8 ± 0.1	**48**	**≤ 0.0001**
		YPD + ATc	36	85.5 ± 1.6		
I-*Sce*I + TargetB + *GPI16*	**mono-GFP**	YPD	36	0.3 ± 0.02	**105**	**≤ 0.0001**
		YPD + ATc	36	31.6 ± 2.2		
	mono-BFP	YPD	36	0.5 ± 0.04	**58**	**≤ 0.0001**
		YPD + ATc	36	29.0 ± 1.3		
I-*Sce*I + TargetA + *GPI16*	**mono-GFP**	YPD	36	0.3 ± 0.02	**43**	**≤ 0.0001**
		YPD + ATc	35	14.8 ± 0.9		
	mono-BFP	YPD	36	1.0 ± 0.1	**100**	**≤ 0.0001**
		YPD + ATc	35	101.8 ± 1.6		
I-*Sce*I only	**mono-GFP**	YPD	21	1.1 ± 0.1	**1**	**0.9424**
		YPD + ATc	21	1.1 ± 0.1		
	mono-BFP	YPD	21	1.0 ± 0.0	**1**	**0.4605**
		YPD + ATc	21	0.9 ± 0.0		
Target only	**mono-GFP**	YPD	21	1.0 ± 0.1	**1**	**0.0514**
		YPD + ATc	21	0.9 ± 0.1		
	mono-BFP	YPD	21	1.1 ± 0.1	**1**	**0.2251**
		YPD + ATc	21	1.0 ± 0.1		

[1]Values represent the biological replicates analyzed

I-SceI-induced DNA DSBs are predominantly repaired by GC

-FOA^R colonies obtained following I-*Sce*I induction can arise from point mutation in the *URA3* gene or as a consequence of DNA DSB-triggered GC, BIR, MCO or WCL/SCL events (Figure 17A p.80) (21). We used PCR to assess whether 5-FOA resistance was a consequence of *URA3* loss or point mutation and SNP-typing to assess the heterozygous or homozygous state of SNPs of interest allowing to deduce the length of the LOH and to distinguish between GC, BIR, MCO or WCL/SCL events (365) as illustrated in Figure 17A p.80. We used SNP156 (*Taq*I restriction site on left arm of Chr4A) located between the telomere and the BFP/GFP locus and SNP95 (*Alu*I restriction site on the right arm Chr4A), located close to the centromere (365). PCR of 64 5-FOA^R clones derived from "I-*Sce*I+TargetB" revealed that none of these clones had acquired 5-FOA resistance by point mutation in the *URA3* gene, as *URA3* itself could no longer be detected. Furthermore, SNP156 and SNP95 remained heterozygous in 98.5% (63/64) of the tested clones, consistent with a GC event. The remaining clone was homozygous for SNP156 and heterozygous for SNP95, suggesting a BIR/MCO/SCL event. Consistently, all 5-FOA^R clones with GC-mediated LOH events were still expressing both BFP and GFP and were HIS+ ARG+ while the unique 5-FOA^R clone with a BIR/MCO/SCL-mediated LOH event was only expressing GFP and found to be HIS- ARG+.

To assess whether the high number of GC-mediated LOH events is specifically linked to I-*Sce*I expression, we tested 61 5-FOA^R colonies arising from the "I-*Sce*I+TargetB" strain grown in absence of ATc and from the "Target only" strain that lacks the I-*Sce*I gene. We found that in the "I-*Sce*I+TargetB" strain, 60.7% of the 5-FOA^R clones (37/61) had undergone LOH through GC and 39.3% (24/61) through WCL (Figure 17C p.80) in absence of ATc. When testing "Target only" 5-FOA^R clones, WCL appeared as the main mechanism leading to LOH events (95.4%) on Chr4 both in absence and presence of ATc. Differences observed between the results obtained for the "I-*Sce*I+TargetB" strain grown in absence of ATc and the "Target only" strain might reflect leakage of the P_{TET} promoter in the absence of inducer. Thus, taken together, our results indicated that GC is the predominant mechanism for the repair of an I-*Sce*I-induced DNA DSB on Chr4B in *C. albicans*.

In order to determine the frequency at which BIR/MCO, SCL and WCL might occur when GC was not the mechanism of repair, we used fluorescence activated cell sorting (FACS) to isolate the mono-GFP cells observed by flow cytometry that are likely to have undergone long-range LOH events. 48 confirmed mono-GFP clones were analyzed for the loss of auxotrophic markers, SNP156 homozygosity and SNP95 heterozygosity. All tested

clones were URA- HIS- ARG+. As expected, SNP-typing revealed that all clones obtained from the induced culture had repaired the I-*Sce*I-induced DNA DSB by BIR, MCO or SCL (Figure 17D p.80). Furthermore, genome sequencing of a subset of these clones identified no cases of SCL (data not shown). By contrast, in the non-induced conditions, LOH events were the result of BIR, MCO or SCL (58%) but also WCL (42%) (Figure 17D p.80).

Taken together, our results revealed that a majority of *C. albicans* cells repaired an I-*Sce*I-induced DNA DSB on Chr4B by GC but that BIR or MCO could also be used, although at lower frequency.

GC-independent repair of an I-SceI-induced DNA DSB on chromosome 4 haplotype A leads to inviable progeny

Results presented above were obtained with a strain that harbored the I-*Sce*I target site on Chr4B. We and others have shown that LOH events on Chr4 are biased towards haplotype A suggesting that Chr4B may bear one or more recessive lethal alleles that could influence the frequency at which DNA DSB repair mechanisms are detected in our assay (10, 20, 74, 126). Therefore, we constructed the "I-*Sce*I+TargetA" strain, which carries the I-*Sce*I target site on the GFP-bearing Chr4A homolog (Figure 18A). WGS of "I-*Sce*I+TargetA" confirmed the location of the I-*Sce*I site on Chr4A and gross chromosomal rearrangements were not observed (data not shown). We observed that growth of strain "I-*Sce*I+TargetA" in the presence of ATc resulted in a large increase (2,450-fold) in the number of 5-FOA^R clones as compared to the non-induced condition (Table 2 p.81). In addition, we observed a 48-fold increase in the number of mono-BFP cells upon induction. Again, we also observed an unexpected 8-fold increase in the number of cells expressing only the other fluorescent protein (mono-GFP) in the induced cultures of the "I-*Sce*I+TargetA" strain (Figure 18B; Table 3 p.83; see below for further investigation of this observation).

SNP-typing of 62 5-FOA^R clones revealed that 98.4% (61/62) arose from a GC event, in agreement with the cells being doubly fluorescent and HIS+ ARG+ (Figures 18A and D).

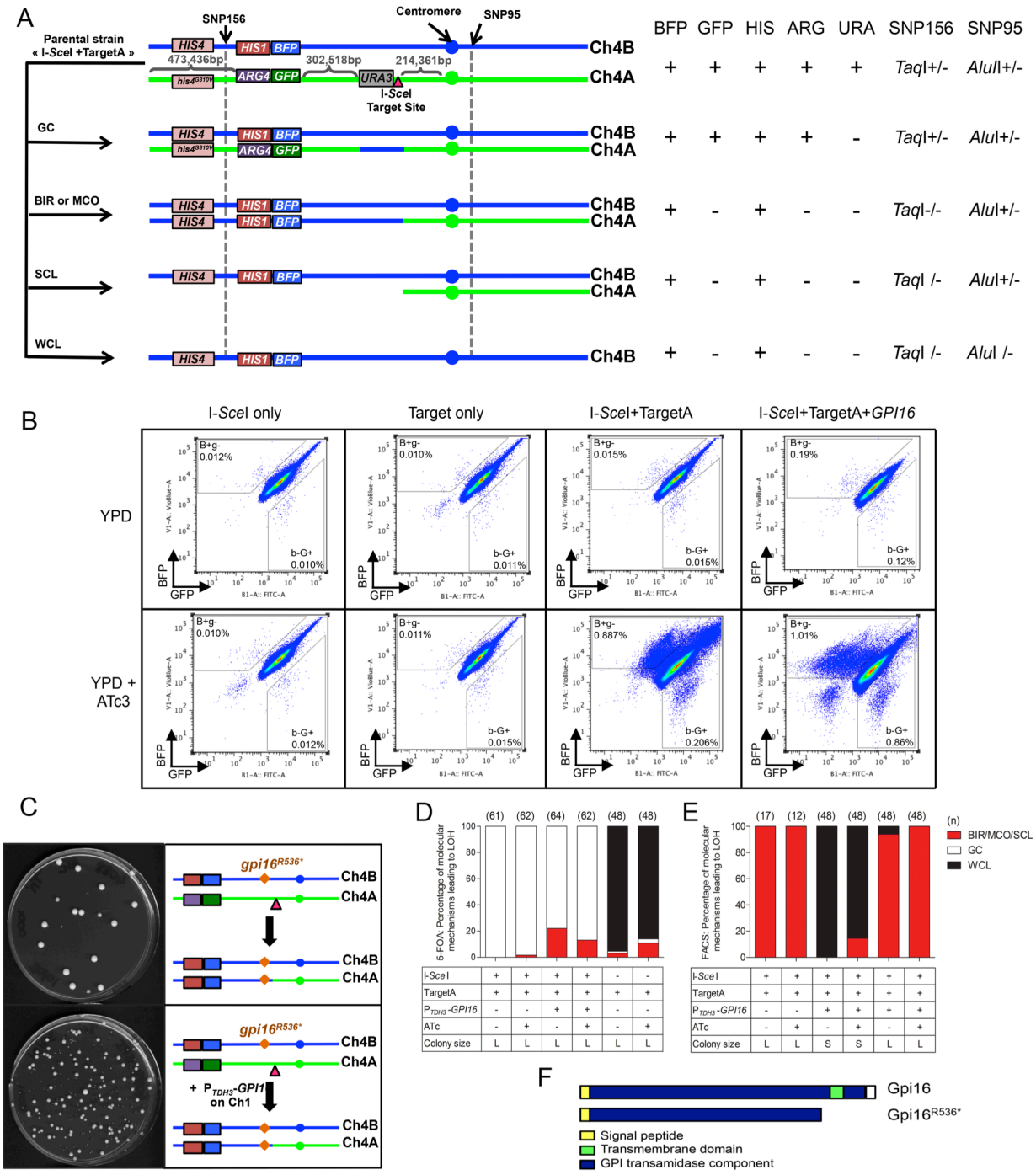


Figure 18: Effect of I-*Sce*I induction in « I-*Sce*I + TargetA » and control strains.
(A) Different LOH events on Chr4A can arise from I-*Sce*I. As seen in Figure 17A, the combined heterozygosity or homozygosity of the SNP156 and SNP95 gives insights about the molecular mechanisms leading to LOH events. GC: gene conversion, BIR: break-induced replication, MCO: mitotic crossover, SCL: segmental chromosome loss, WCL: whole chromosome loss. (B) Co-occurrence of I-*Sce*I and its target sequence triggers a predominant increase in mono-BFP cells. 10^6 events are displayed. The B+g- and b-G+ gates were defined arbitrarily. (C) Integration of the full-length allele of *GPI16* on Chr1 allows recovery of viable mono-BFP cells after cell sorting. While cells obtained from strain « I-*Sce*I+TargetA » showed poor viability due to homozygosis of the gpi^{R536*} allele, complementation with a wild-type *GPI16* allele in strain «I-*Sce*I+TargetA+*GPI16* » restored viability. The largest colonies

observed in both cases are doubly fluorescent, having not undergone I-*Sce*I cleavage on Chr4A. (D and E) SNP-RFLP analysis shows that I-*Sce*I dependent DNA DSBs on Chr4A are mainly repaired by GC. Histograms presenting the proportion of BIR/MCO or SCL, GC and WCL events in the population having undergone a LOH either recovered from 5-FOA counter-selection (D) or from FACS (E). BIR/MCO or SCL events correspond to mono-BFP cells that displayed a homozygous SNP156 but have maintained a heterozygous SNP95. WCL events correspond to mono-BFP cells in which both SNP156 and SNP95 became homozygous. GC events correspond to doubly fluorescent cells in which both SN156 and SNP95 remained heterozygous. L: large-sized colonies and S: small-sized colonies. (F) The $gpi16^{R536*}$ allele might result in the truncation of the Gpi16 protein carboxy-terminal transmembrane domain, part of the conserved GPI transamidase domain.

Hence, GC also appears to be the predominant mechanism for the repair of an I-*Sce*I-induced DNA DSB on Chr4A in *C. albicans*. Unexpectedly, the remaining 5-FOA^R clone appeared as mono-GFP by flow cytometry. This clone was homozygous for SNP156 but heterozygous for SNP95 and HIS+ ARG-. This suggested that this 5-FOA^R clone belonged to the rare mono-GFP cell population observed by flow cytometry as described above and that are likely to have arisen by other recombination events (see below).

In order to determine the frequency of the molecular mechanisms giving rise to mono-BFP cells, we enriched them by FACS and plated them onto YPD medium. Strikingly, only a subset of the plated cells was able to form colonies (~4%, Figure 18C) whose characterization highlighted two populations: (i) doubly fluorescent and HIS+ ARG+ URA+ colonies – suggesting that they were wild type cells, illegitimately recovered in our sorting procedure – and (ii) mono-BFP and HIS+ ARG- URA+ colonies – likely to result from an I-*Sce*I-independent LOH (Figure 18E). This result suggested that all mono-BFP cells that had arisen by repair of the I-*Sce*I target site on Chr4A were inviable, possibly due to homozygosis of one or more recessive lethal alleles on Chr4B.

A heterozygous null mutation in the **GPI16** *gene is responsible for the inviability of* **C. albicans** *cells homozygous on the left arm of chromosome 4 haplotype B*

Results presented above implied the presence of at least one recessive lethal allele on Chr4B between the left arm telomere (position 1) and I-*Sce*I target site (position 778,082). We reasoned that (i) the genotype for this allele should be heterozygous in *C. albicans* strain SC5314, as homozygosis of the Chr4A allele is viable, (ii) the recessive lethal allele should never be found in the homozygous state in the *C. albicans* population, and (iii) it should not affect a gene previously shown to be dispensable in *C. albicans*. In order to identify such an allele, we took advantage of sequencing data obtained from 155 *C. albicans* isolates including

the reference strain SC5314 (M.E.B., G.S., N.S., K.S., C.M., and C.d'E., manuscript *in preparation*) and searched for SNPs generating a stop codon in ORFs on Chr4B and meeting the above criteria. Strikingly, only one such SNP was identified in SC5314 at position 659,191 on Chr4B (equivalent to position 659,155 on Chr4A), which resulted in a change from CGA (arginine) on Chr4A to TGA (STOP codon) on Chr4B in the C4_03130W gene. The premature STOP codon resulted in a protein shorter by 87 amino acids, deleting a C-terminal membrane-spanning domain (Figure 18F C4_03130W is the ortholog of the essential *S. cerevisiae GPI16* gene encoding a membrane-bound component of a glycosylphosphatidylinositol (GPI) transamidase complex necessary for GPI anchor biosynthesis. Notably, no disruptant could be obtained for the C4_03130W gene (now referred to as *GPI16*), suggesting that this gene is essential in *C. albicans* (426). Thus, our results suggested that the truncated *GPI16* allele (referred to as *gpi16*^{R536*}) might be responsible for the inviability of *C. albicans* cells that experienced a long-range LOH on Chr4B.

To test this hypothesis, the *GPI16* wild-type ORF available in the *C. albicans* ORFeome (402) was placed under the control of the constitutive P_{TDH3} promoter and integrated at the *RPS1* locus on Chr1 in the "I-*Sce*I+TargetA" strain, yielding strain "I-*Sce*I+TargetA+*GPI16*". As observed for the "I-*Sce*I+TargetA" strain, growth of the "I-*Sce*I+TargetA+*GPI16*" strain resulted in a 100-fold increase in the number of mono-BFP cells in presence of ATc (Figure 18B; Table 3 p.83). Interestingly, FACS-sorted mono-BFP cells derived from strain "I-*Sce*I+TargetA+*GPI16*" grown in the presence or absence of ATc showed 100% viability on YPD agar plates, although variability in colony size was observed with 54% of small- and 46% of large-sized colonies (Figures 18C and 19A). These results contrasted with those obtained for strain "I-*Sce*I+TargetA" and indicated that overexpression of *GPI16* complemented the inviability of mono-BFP cells. This confirmed that the *gpi16*^{R536*} allele was the recessive lethal allele responsible for this inviability.

As done previously, we evaluated the nature and frequency of the molecular mechanisms at the origin of mono-BFP cells derived from the "I-*Sce*I+TargetA+*GPI16*" strain. As we observed variability in colony size, we independently analyzed 48 large and 48 small colonies. Based on auxotrophy, SNP-typing and WGS we could conclude that large colonies had all arisen by BIR/MCO while small colonies had arisen predominantly by WCL and reduplication (85.4%). The remaining 14.6% small-colony variants had arisen by BIR/MCO. Notably, even though a cross-over between the BFP/GFP locus and the

heterozygous *HIS4*/*his4*^{G310V} could arise and allow the occurrence of mono-BFP cells with histidine auxotrophy, these were never observed.

Results presented above suggested that mono-BFP cells arose from an I-*Sce*I-dependent DNA DSB either repaired by BIR/MCO or resulting in a WCL. As we observed 46% of colonies were large and 54% were small among 1,054 FACS-sorted mono-BFP cells from strain "I-*Sce*I+TargetA+*GPI16*" upon I-*Sce*I induction, BIR/MCO and WCL seem to occur at similar frequencies. However, homozygosis of a second recessive allele located between the I-*Sce*I target site and the right telomere of Chr4B could explain the small-colony phenotype associated with Chr4B WCL (Figure 19A).

A heterozygous null mutation in the* MRF2 *gene is partially responsible for the small colony variants arising upon chromosome 4 haplotype A loss

Small colonies were predominantly observed upon Chr4A WCL in both non-induced and induced conditions (89/96 when pooled). We thus used the same approach presented above to identify a recessive allele of a non-essential gene in the region of Chr4 extending from the I-*Sce*I target site location to the right telomere responsible for the observed phenotype. Only one SNP was identified in SC5314 and is located at position 796,698 on Chr4B (796,679 on Chr4A) and results in a change from CGA (Arginine) on Chr4A to TGA (STOP codon) on Chr4B in the C4_03750C gene (*mrf2*^{R362*}) (Figure 17A p.80). The premature STOP codon resulted in a protein shorter by 34 amino acids (Figure 17B p.80). This gene is the ortholog of *S. cerevisiae MRF1* that encodes a putative mitochondrial translational release factor. Deletion of *MRF1* in *S. cerevisiae* and other organisms results in acute respiratory defects (427-429) but the consequence of inactivating the C4_03750C gene in *C. albicans* has not yet been investigated. As the name *MRF1* has been assigned to the C1_11700C gene in *C. albicans* (430), ortholog of *S. cerevisiae ETR1,* we instead refer to C4_03750C as *MRF2*.

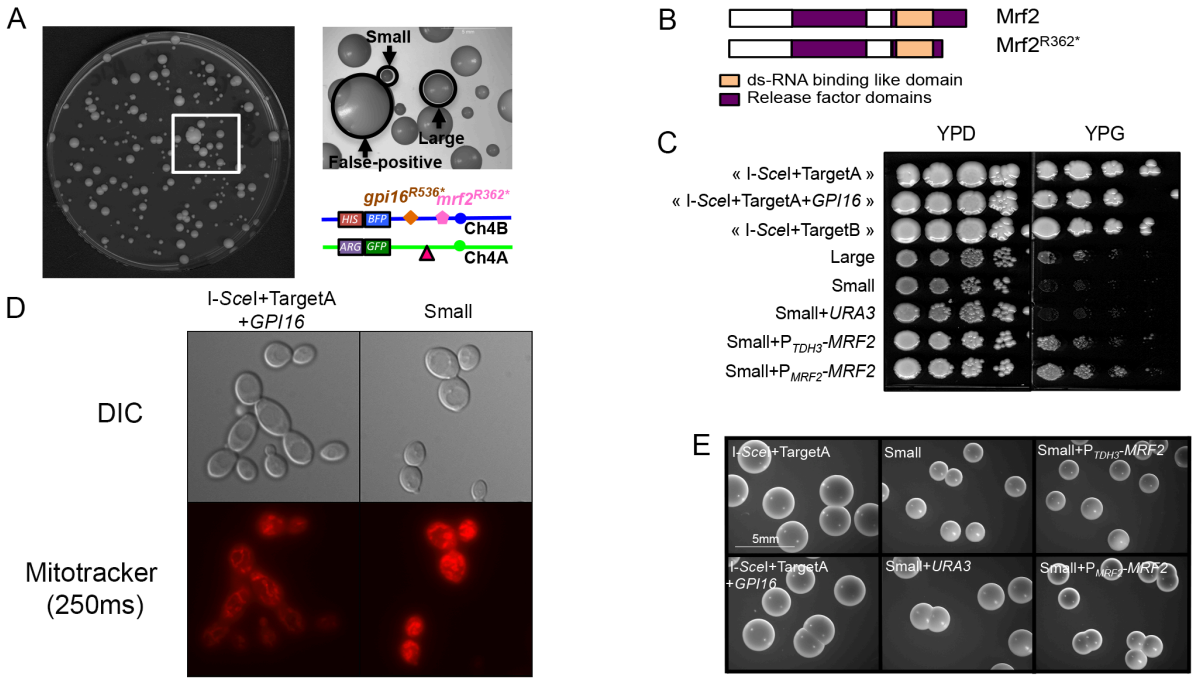


Figure 19: Homozygosis of Chr4B is associated with phenotypic heterogeneity due to an additional recessive deleterious allele.
(A) Heterogeneous colony sizes of mono-BFP cells derived from the « I-*Sce*I+TargetA+*GPI16* » strain. Small and large colonies are indicated by a black arrow. Small-colony variants are mainly resulting from WCL while large-colony variants are resulting from BIR/MCO. Very large colonies (false-positive) are doubly fluorescent, having not undergone I-*Sce*I cleavage on Chr4A. The locations of the *mrf2*^{R362*} and *gpi16*^{R536*} mutated alleles on Chr4B on both sides of the I-*Sce*I target site are shown. (B) Homozygosis of *mrf*^{R362*} allele gives rise to a truncated non-functional protein. The Mrf2 protein encoded by the *MRF2* functional allele is 396 amino acids long, but when encoded by the *mrf*^{R362*} allele, if translated, the protein would be shorter by 34 amino acids removing the C-terminal part of the release factor domain. (C) Small colonies have a respiratory defect that is restored upon complementation with *MRF2*. Cells were spotted on rich medium containing glucose or glycerol as carbon source. No growth was observed on YPG for small colonies. Complementation with P$_{TDH3}$-*MRF2* or P$_{MRF2}$-*MRF2* restored growth on YPG while complementation with *URA3* only did not. (D) Cells from the small colonies show a defect in the mitochondrial network. Cells were stained with MitoTracker for mitochondria. The panels display microscopy pictures of the control strain (I-*Sce*I+TargetA+*GPI16)* and the small colony variants-derived cells. The cells were observed at 100X in DIC and Cy3 for Mitotracker staining (250ms - red). The cells were examined under a Leica DMRXA microscope. (E) Complementation with a *MRF2* functional allele does not restore wild-type colony size in small colony variants on YPD agar medium. Pictures were taken with a stereomicroscope Leica M80 at a zoom of 7.5X.

In order to test whether *mrf2*^{R362*} was associated with respiratory defects in *C. albicans*, we first grew small-colony variants obtained from strain "I-*Sce*I+TargetA+*GPI16*" on YPD and on YP glycerol (YPG) agar plates. Because glycerol is a non-fermentable carbon source, functional mitochondria are required for its assimilation via respiration. We observed that

small-colony variants could not grow on YPG (Figure 19C), in contrast to the parental strain and the large-colony variants derived from this strain, thus suggesting that the *mrf2*^{R362*} allele is associated with mitochondrial dysfunction. Mitochondria staining of cells derived from the small-colony variants and "I-*Sce*I+TargetA+*GPI16*" parental strain reinforced our hypothesis. Indeed, while the parental strain's mitochondria appeared as an interconnected filamentous network, characteristic of healthy cells, those of the small-colony variants appeared patchier (Figure 19D). We further confirmed that the *mrf2*^{R362*} allele was responsible for the mitochondrial defect as complementation with the wild-type *MRF2* allele restored growth of the small colony variants on YPG medium (Figure 19C). Yet, the size of the colonies remained small on YPD medium (Figure 19E) suggesting the occurrence of a third recessive deleterious allele on Chr4B, though we were unable to identify it.

GC with CO are also involved in the repair of I-SceI-induced DNA DSBs

As mentioned previously, induction of I-*Sce*I expression yielded mono-fluorescent cells with an unexpected fluorescence. Mono-BFP cells were observed upon the expression of I-*Sce*I in the "I-*Sce*I+TargetB" strain while only mono-GFP cells were expected upon repair of the I-*Sce*I-induced DNA DSB (Figure 17B p.80). Similarly, mono-GFP cells were observed upon I-*Sce*I induction in the "I-*Sce*I+TargetA" (Figure 18B p.86). To understand the basis for these rare cell populations, we first collected mono-BFP cells obtained from the "I-*Sce*I+TargetB" strain after I-*Sce*I induction. Almost all mono-BFP cells were inviable on YPD upon cell sorting (~96%), leading us to hypothesize that they had undergone a LOH event that had rendered Chr4B homozygous. Among the few viable cells, 12 mono-BFP cells had likely repaired the I-*Sce*I-induced DNA DSB, as deduced from the loss of both the *URA3* gene and the I-*Sce*I target sequence. These cells were analyzed by SNP-typing and appeared to have undergone a BIR/MCO or SCL event, rendering SNP156 homozygous for haplotype B despite the presence of the I-*Sce*I target sequence on the Chr4B homolog. We further investigated the nature of these events by WGS of 5 mono-BFP clones.

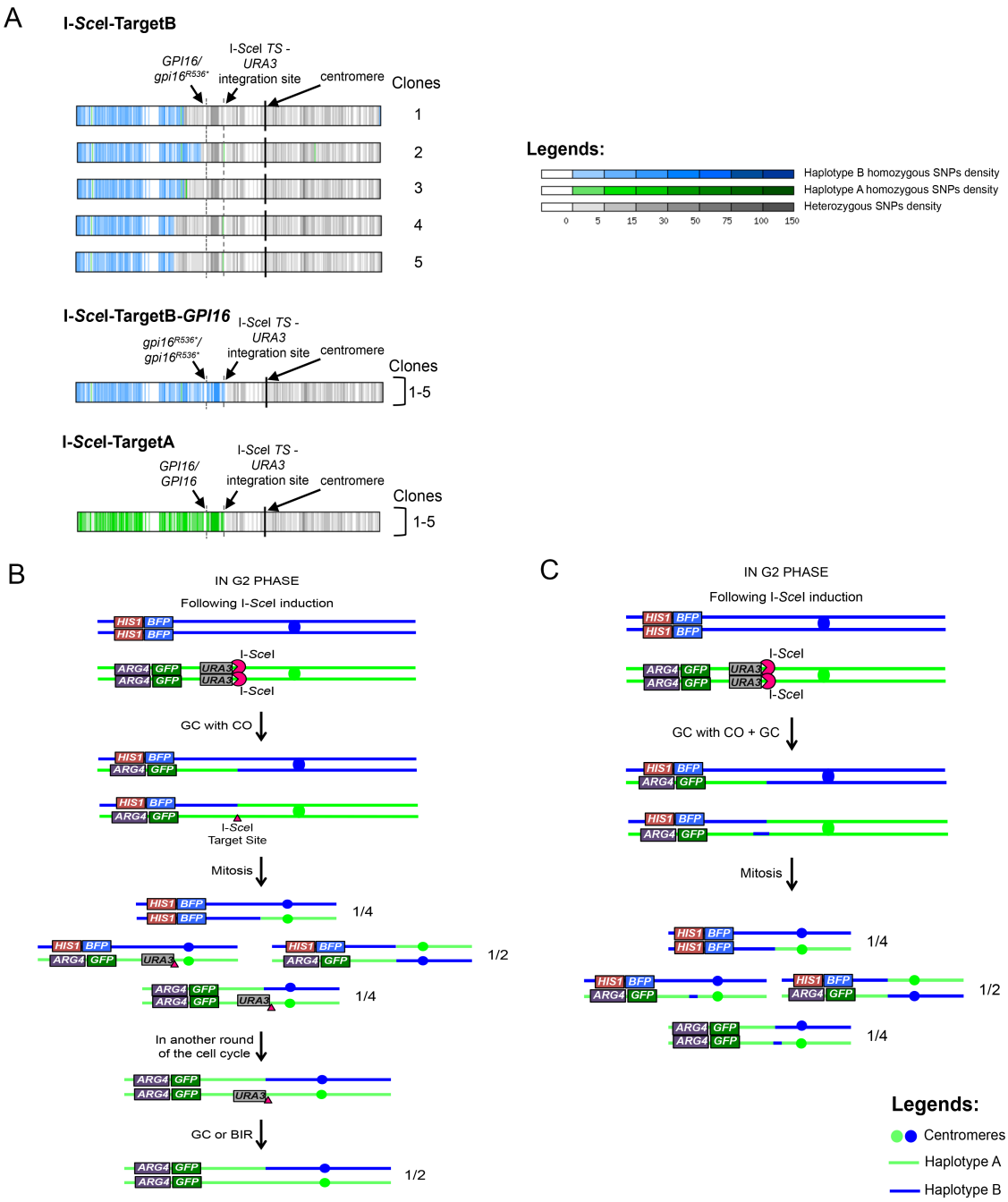


Figure 20: Unexpected LOH events result from independent BIR/MCO and GC events or GC with CO events.

(A) Representation of LOH events that occurred in sequenced unexpected mono-fluorescent cells. WGS allowed the identification of LOH events occurring in the unexpected (i) mono-BFP cells from « I-*Sce*I+TargetB » strain, (ii) mono-BFP cells from « I-*Sce*I+TargetB+*GPI16* » strain (clone 5 displays an aneuploidy on Chr5, data not shown), (iii) mono-GFP from « I-*Sce*I+TargetA » strain (clone 4 displays a truncation of Chr3, data not shown). Chr4 for each strain (or group of strains) is represented as a horizontal box with vertical bars corresponding to 1kb regions. Vertical bars are colored grey if heterozygous, green if haplotype A-homozygous and blue if haplotype B-homozygous (haplotypes A and B harbor the GFP and BFP genes, respectively). Different levels of grey, green or blue intensity indicate local changes in SNP density. White regions are homozygous in both the sequenced strain and strain SC5314 that was used to define haplotypes A and B. Centromeres are shown

as a black vertical bars. The location of the *GPI16* and I-*Sce*I target site-*URA3* loci are indicated. (B) Multiple but not simultaneous repair events could be responsible for the unexpected mono-GFP cells upon repair of an I-*Sce*I-induced DSB in the "I-*Sce*I+TargetA" strain. Upon I-*Sce*I induction, one of the two chromatids is cut and repaired using the homologous chromosome as a template by GC with CO. After a mitotic event, ¼ of the population became mono-GFP with one homolog still carrying the *URA3* marker and I-*Sce*I target sequence. Because both *URA3* marker and I-*Sce*I target sequence were found to be absent, we propose that, during the 8 hours induction, cells having inherited the GFP-bearing chromosomes with one copy of *URA3* and the target site can go through another loop of DNA DSB repair in G2 phase yielding 50% of the newly generated mono-GFP lacking the *URA3* marker or the target sequence. (C) Multiple and simultaneous repair events could alternatively be responsible for the unexpected mono-GFP cells upon repair of an I-*Sce*I-induced DNA DSB in the "I-*Sce*I+TargetA" strain. Upon I-*Sce*I induction, both sister chromatids are cut: while one chromatid is repaired using the homologous chromosome as a template by GC with CO, the second sister chromatid is repaired by GC using the homologous chromosome as a template. After mitosis, ¼ of the population has become mono-GFP. Models shown in (B) and (C) are also valid for mono-BFP cells from "I-*Sce*I+TargetB+*GPI16*".

As diagrammed in Figure 20A, all 5 clones had undergone homozygosis that had rendered Chr4 homozygous for haplotype B from the left telomere to a position <659,191 *i.e.* upstream of the *GPI16* locus instead of position 778,082, where the I-*Sce*I target site had been inserted. Hence, in these clones, heterozygosity was maintained in the region that encompasses the *GPI16* gene. We conclude that viable mono-BFP cells arising from the "I-*Sce*I+TargetB" strain are the result of two independent recombination events: an I-*Sce*I dependent GC event on Chr4B at the I-*Sce*I-*URA3* locus and a I-*Sce*I-independent BIR/MCO event on Chr4A that has led to the homozygosity of Chr4B left arm while preserving a functional *GPI16* allele.

Based on the $gpi16^{R536*}$ findings in the "I-*Sce*I+TargetA" strain we hypothesized that the inviability of most mono-BFP cells obtained from the "I-*Sce*I+TargetB" strain was the result of homozygosis of the $gpi16^{R536*}$ allele. Thus, we generated strain "I-*Sce*I+TargetB+*GPI16*" by integrating the *GPI16* wild-type allele placed under the control of the P_{TDH3} promoter at the *RPS1* locus on Chr1 in the "I-*Sce*I+TargetB" strain. When I-*Sce*I expression was induced in this strain, we observed a 58-fold increase in the appearance of mono-BFP cells (Table 3 p.83). This increase is 3 times higher than the frequency of unexpected mono-BFP cells in strain "I-*Sce*I+TargetB" but can be explained by an increased viability of the cells during the time of the experiment. As we predicted, all mono-BFP cells recovered by FACS were now viable. WGS of 5 mono-BFP clones revealed that they had undergone a LOH rendering Chr4 homozygous for haplotype B from the I-*Sce*I-*URA3* locus to the left telomere (Figure 20A). As illustrated in Figures 20B and C, we hypothesize that

these mono-BFP cells have arisen through successive or simultaneous repair events involving GC with CO followed by GC/BIR or MCO at the I-*Sce*I sites.

Similarly, we collected unexpected mono-GFP cells obtained from the "I-*Sce*I+TargetA" strain. These cells showed 100% viability and those that had repaired the I-*Sce*I-induced DNA DSB, as deduced from the loss of the *URA3* gene, appeared to have experienced BIR/MCO or SCL events. WGS of 5 mono-GFP cells confirmed that they had undergone homozygosis, rendering Chr4A homozygous from the I-*Sce*I-*URA3* locus to the left telomere (Figure 20A). Here again, we hypothesize that these mono-GFP cells have arisen through successive or simultaneous repair events involving GC with CO followed by GC/BIR or MCO at the I-*Sce*I sites (Figures 20B and C).

Taken together, these results suggested that, in addition to GC and BIR events, the repair of an I-*Sce*I-induced DNA DSB on Chr4 could involve GC with CO events.

Heterozygous mutations are also responsible for haplotype-specific LOH in clinical strains.

The identification of the recessive deleterious mutations $gpi16^{R536*}$ and $mrf2^{R362*}$ was made in *C. albicans* strain SC5314. Notably, both mutations appeared unique to this strain. Therefore, we asked whether different recessive lethal alleles might occur in other *C. albicans* isolates. To this end, we focused on Chr5, as selection for utilization of L-sorbose as the sole carbon source by *C. albicans* has been reported to trigger the loss of one Chr5 homolog and, thus, whole Chr5 homozygosis (57, 431, 432). Chr5 also carries on its left arm the mating type-like locus (*MTL*), often found to be heterozygous (*MTL***a**/α) in *C. albicans* strains, which can be used as a marker of homozygosity upon genomic rearrangement on Chr5 (Figure 21A). We hypothesized that if a *C. albicans* isolate was harboring a recessive lethal allele on Chr5, L-sorbose-utilizing (SOU^+) progeny would undergo LOH events maintaining only one of the two Chr5 haplotypes and thus, a unique mating-type. To test this hypothesis, we scanned the genomes of *C. albicans* isolates and identified strains CEC2876 and CEC3673 as harboring a heterozygous SNP on Chr5 that might generate a potentially non-functional allele of a presumably essential gene.

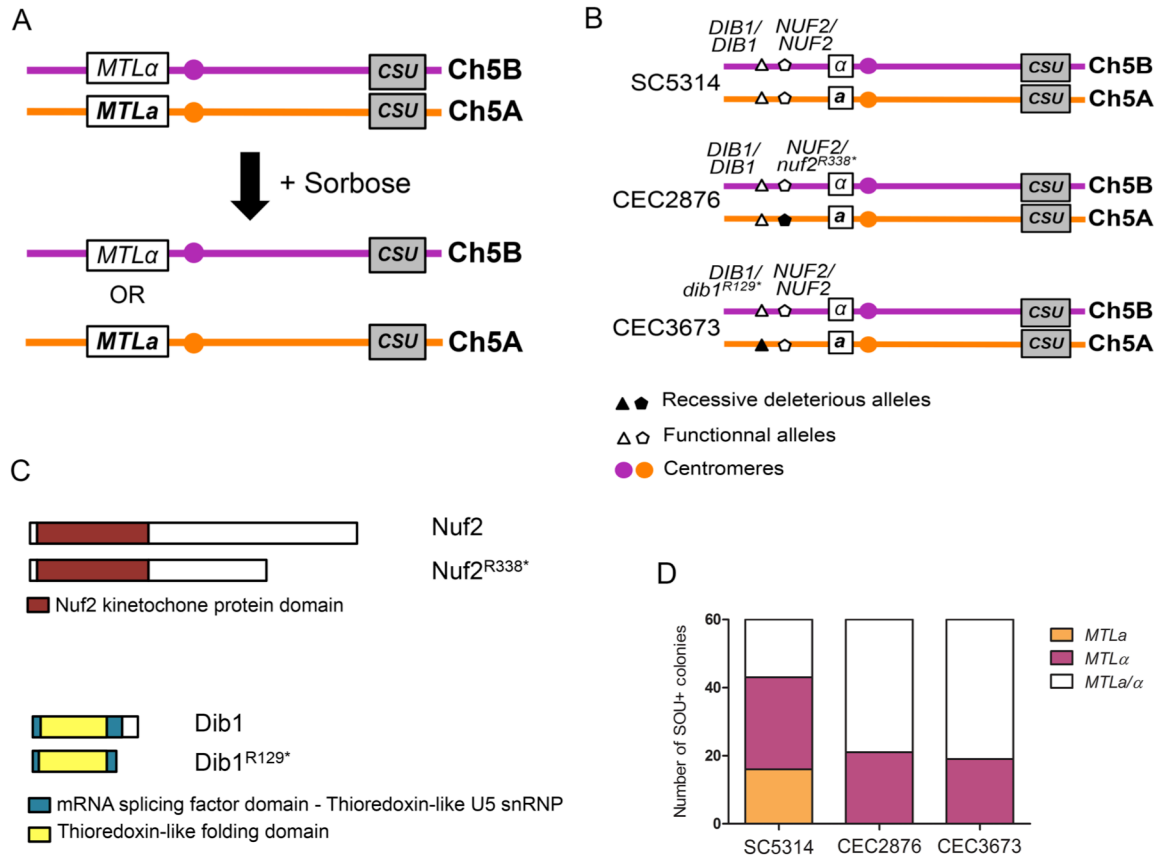


Figure 21: Heterozygous deleterious recessive alleles are also present in clinical isolates. (A) Principle of sorbose-induced WCL impacting Chr5 in *C. albicans*. *C. albicans* cells that have undergone Chr5 loss can grow on sorbose. Conveniently, Chr5 is heterozygous at the mating-type locus (*MTL**a**/α*) and the heterozygous or homozygous state of this locus can be screened by PCR. (B) Map of the Chr5 showing the localization of the mutated alleles in clinical isolates. Applying the same method as for the strains "I-*Sce*I+TargetA" and "I-*Sce*I+TargetB", we identified a mutation in *NUF2* and *DIB1* respectively found in CEC2876 and CEC3673 clinical strains. These mutations are both localized on Chr5A. (C) Homozygosis of *nuf2*^{R338*} or *dib1*^{R129*} allele gives rise to truncated non-functional proteins. The *NUF2* and *DIB1* functional alleles encode proteins of 470 and 149 amino acid long, respectively, but when encoded by the *nuf2*^{R338*} and *dib1*^{R129*} alleles, the proteins are shorter by 142 and 29 amino acids, respectively. (D) Growth on sorbose gave rise to both *MTL**a*** and *MTLα* homozygous clones for SC5314 but only *MTLα* homozygous clones for the clinical isolated. The Chr5 loss was assessed through PCR at the *MTL* locus. While both homozygous *MTL* loci can be found for SC5314 reference strain, the presence of the mutated alleles at the heterozygous state on Chr5A prevents the homozygosis of this haplotype in both clinical isolates.

In strain CEC2876, the identified SNP was located at position 289,097 on Chr5A (289,095 on Chr5B) and resulted in a change from CGA (arginine) on Chr5B to TGA (STOP) on Chr5A in the C5_01280C gene (*nuf2*^{R338*}). The premature STOP codon resulted in a protein shorter by 132 amino acids. This gene is the ortholog of *S. cerevisiae NUF2* that encodes a kinetochore component (Figures 19B and C p.90).

In strain CEC3673, the identified SNP was located at position 212,951 on Chr5A (212,941 on Chr5B) and resulted in a change from AGA (arginine) on Chr5B to TGA (STOP) on Chr5A in the C5_00920W gene (*dib1*^{R129*}). The premature STOP codon resulted in a protein shorter by 20 amino acids. This gene is the ortholog of *S. cerevisiae DIB1* that plays a role in mRNA splicing and DNA methylation regulation (Figures 19B and C p.90). Deletion of either *NUF2* or *DIB1* in *S. cerevisiae* results in lethality but the consequence of inactivating these ORFs in *C. albicans* has not yet been investigated.

After selection on sorbose containing-media, the heterozygous status of the *MTL* locus of 60 single SOU+ colonies was analyzed by PCR and SC5314 was used as a control. While isolates homozygous for both mating types were recovered for SC5314 that do not carry the *nuf2*^{R338*} and *dib1*^{R129*} alleles (16 *MTL**a*** and 27 *MTLα*) (Figure 21D), only *MTLα SOU+* derivatives were recovered for CEC2876 and CEC3673 (21 *MTLα* and 19 *MTLα*, respectively) (Figure 21D). In all three cases, the remaining progeny were found to maintain heterozygosity at the mating type-like locus suggesting that they had acquired sorbose resistance independently of an LOH event encompassing the *MTL*, *DIB1* and *NUF2* loci (Figure 21D).

Taken together, these results indicated the occurrence of recessive lethal alleles in the heterozygous state in *C. albicans* isolates, responsible for the haplotype bias observed when these isolates undergo LOH.

Discussion

Having a dynamic genome is now recognized as one of *C. albicans*' abilities that has led to its success as both a commensal and a pathogen. Yet little is known about the molecular events that lead to the genome rearrangements that are observed in *C. albicans* isolates. Our work aimed at studying the repair of DNA DSBs and facilitating the study of genome dynamics in *C. albicans*. Here, we have (i) developed a DNA DSB-inducing system that generates a DNA DSB at a defined site in the *C. albicans* genome, (ii) demonstrated that this induced DNA DSB is mainly repaired by gene conversion, (iii) explained the previously observed bias in Chr4 haplotype homozygosis by the presence of recessive lethal and deleterious alleles at the *GPI16* and *MRF2* loci in *C. albicans* strain SC5314, (iv) shown that similar haplotype biases occur in other *C. albicans* strains, due to different recessive lethal alleles, and (v) observed rare and complex molecular mechanisms involved in DNA DSB repair in *C. albicans*. This work presents for the first time a precise study of haplotype-dependent repair mechanisms in a natural heterozygous diploid organism.

***Mechanisms of double-strand break repair in* C. albicans**

In this study, we estimated the LOH frequency associated with I-*Sce*I-induced DNA DSB making use of two assays: 5-FOA counter-selection, observed upon short- and long-range LOH events, and flow cytometry, whereby loss of one fluorescence reveals long-range LOH only. Taken together, our results show that LOH events associated with the repair of a site-specific DNA DSB in *C. albicans* are mainly due to GC but also at a minor rate, BIR/MCO. Notably, the location of the I-*Sce*I target site on Chr4A or Chr4B had different outcomes in the frequency at which BIR/MCO and WCL events were observed, with BIR/MCO and WCL occurring at similar frequencies when the target site was located on Chr4A, while only BIR/MCO events were observed when the target site was located on Chr4B. We propose that different DNA conformations at the time of repair (433, 434) or the presence of heterozygous alleles on one of the homologs could explain the observed homolog-specificity of the molecular mechanisms resulting in LOH (20). Noticeably, our study did not assess Non-Homologous End Joining (NHEJ) repair events; however, NHEJ is thought to be inefficient for DNA DSB repair in *C. albicans* (171, 172).

Our work also highlighted a substantial fold increase of an unanticipated population of mono-fluorescent cells following an I-*Sce*I-induced DNA DSB. Two hypotheses may explain this cell population. First, they could result from the DNA DSB being repaired by GC with CO in the G2 phase (Figure 18B p.86). As a consequence, because I-*Sce*I expression is induced for 8 hours and the mono-fluorescent cells resulting from GC with CO would still carry the I-*Sce*I target site, it is conceivable that cells could undergo a second I-*Sce*I-mediated DNA DSB repaired by GC, BIR, MCO or SCL and associated with the loss of the *URA3* marker (Figure 18B p.86). Alternatively, as suggested by others (248, 435, 436), I-*Sce*I cleavage may occur in early S phase when DNA is the most accessible (437) and therefore happen concomitantly on both sister chromatids. One sister chromatid could be repaired by a classical GC while the other sister chromatid undergoes a GC with CO (Figure 18C p.86). Interestingly, Esposito and colleagues (438) reported that spontaneous GC events occur at a rate of 10^{-7} to 10^{-6} in diploid *S. cerevisiae*, while Haber and Hearn (439) quantified the occurrence of GC with CO events as 12 to 25% of the overall GC events in presence of large homologous regions. More recently, researchers have found that GC events with CO were increased upon generation of a DNA DSB (440), supporting the results found in *C. albicans*.

Finally, our results also suggested the occurrence of an additional rare cell population arising from multiple events on Chr4: the repair of the I-*Sce*I-induced break by GC followed by a spontaneous LOH event on Chr4A. This scenario would result in two homozygous regions:

(i) one short-tract of homozygosity surrounding the I-*Sce*I target site and the *URA3* marker; and (ii) one larger homozygous region starting upstream of the I-*Sce*I target site and extending towards the left telomere (Figure 18A p.86). We are aware that our analyses using SNP-RFLP and WGS remain descriptive and are not a proof of the molecular mechanisms at the origin of the observed LOH events.

Taken together, our results have shown that DNA DSB repair in *C. albicans* most often involves GC. GC also appears as the main repair mechanism of DNA DSBs generated by homing endonucleases (407, 441), more specifically I-*Sce*I (316), in *S. cerevisiae* and in other organisms (436, 442). Interestingly, Forche *et al.* (18) showed that stresses (H_2O_2, fluconazole, 39°C) affected the nature and/or frequency of LOH in *C. albicans*. In the absence of stress, or in presence of H_2O_2, GC and BIR were observed as the main mechanisms leading to LOH events at the *GAL1* locus on Chr1 (18). However, under oxidative stress, an increase in GC events and a decrease in BIR events were observed. Yet, these experiments could not provide information about haplotype specificity. Our approach of generating a targeted DNA DSB allows discrimination of the impact of each haplotype on LOH and the identification of both frequent and rare repair events associated with DNA DSBs. Nonetheless, our experiments have been carried out using a single locus for DNA DSB induction on Chr4 and our results could be locus specific; additional experiments should be conducted to extend our study to other chromosomes and loci. Additionally, although we show that I-*Sce*I can be used in *C. albicans* to generate targeted DNA DSBs, the *C. albicans*-optimized CRISPR-Cas9 system (355) could help extend this study, as it allows targeting haplotype-specific targets without the sophisticated genome engineering that we had to implement.

Recessive deleterious alleles in the C. albicans *genome and their impact on loss-of-heterozygosity*

In this study, we have identified two recessive alleles present in the heterozygous state on Chr4B. The $gpi16^{R356*}$ truncated allele was associated with lethality when found in the homozygous state and therefore identified to be responsible for unidirectional LOH on Chr4. We also identified a truncated allele of the non-essential *MRF2* gene. Complementation of the $mrf2^{R362*}/mrf2^{R362*}$ strains with a functional allele of *MRF2* restored mitochondrial function but did not impact colony size. This either reflects (i) the presence of a third recessive allele on haplotype B that we were not able to identify, (ii) the presence of a heterozygous or homozygous recessive allele in the *C. albicans* genome that upon homozygosis of $mrf2^{R362*}$ mutated allele leads to the small colony size phenotype or (iii) the possibility that, if

translated, the non-functional Mrf2^{R362*} protein has a dominant-negative effect. Nonetheless, the latter hypothesis does not explain why such a dominant-negative effect is not naturally observed in the heterozygous SC5314 strain, unless we take into consideration the difference in genome location which could impact the level of expression of the functional copy of *MRF2*, inserted at the *RPS1* locus, as compared to the endogenous *MRF2* locus.

Although our DNA DSB-inducing system, combined with the availability of a large panel of genome sequences for *C. albicans* isolates, allowed us to identify the mutation underlying the haplotype bias observed upon Chr4 homozygosis (10, 20, 74, 126) and a deleterious allele responsible for respiratory defects, our study was performed *in vitro* and we cannot rule out the existence of additional recessive alleles that would have deleterious effects *in vivo* when present in the homozygous state. In addition, our identification of the *gpi16*^{R356*} and *mrf2*^{R362*} alleles was unique to the reference strain SC5314. Nevertheless, we also demonstrated that haplotype bias upon LOH of Chr5 can be found in two clinical strains, suggesting that our findings may extend to the entire *C. albicans* population.

In this study, only deleterious recessive alleles located on Chr4 of strain SC5314 could be revealed. However, haplotype bias has been observed for other chromosomes in this strain (part of Chr1, Chr3, 6 and 7) (10). Similar to our observation with Chr4, it is likely that deleterious and possibly lethal recessive alleles are located on these chromosomes and responsible for these haplotype biases.

C. albicans reproduction has been shown to be predominantly clonal (443-447) and it is therefore not surprising that recessive deleterious or lethal alleles are found in the diploid genomes of different isolates. Indeed, clonal reproduction should fix such mutations more rapidly than sexual reproduction (448). Interestingly, LOH is frequent in *C. albicans* isolates and one might anticipate that clonal reproduction would progressively lead to homozygosity in this species. The occurrence of heterozygous SNPs affecting the function of genes with significant contributions to *C. albicans* fitness *in vivo* may contribute to the maintenance of heterozygosity if distributed on the two haplotypes of each chromosome. However, it is interesting to note that, in the conditions we have used, recessive alleles with a deleterious effect *in vitro* were only present on Chr4B, with Chr4A being apparently devoid of such alleles.

Material and methods

Strains and media

The *C. albicans* strains used in this study are derived from SN148 (266) and are listed in Table S1. Yeast cells were grown at 30°C in liquid media, either in YPD (1% yeast extract, 2% peptone, 2% dextrose) or SC (0.67% Yeast Nitrogen Base without amino acids, 2% dextrose supplemented with the appropriate 0.08% drop out mix of amino acids). Solid media were obtained by adding 2% of agar. Additionally, YPG (1% yeast extract, 2% peptone, 2% glycerol, 2% agar), sorbose-containing medium (0.7% Yeast Nitrogen Base without amino acid, 2% L-sorbose (Fluka Analytical), 2% agarose – agarose was used instead of agar to avoid the use of scavenger cells (personal communication from Guilhem Janbon)) and 5-FOA-containing medium (0.7% Yeast Nitrogen Base without amino acid, 0.0625% 5-Fluoro-orotic Acid (Toronto Research Chemicals), 0.01% uridine, 2% glucose, 2% agar, supplemented with leucine, arginine and histidine for the needs of the experiment) have also been used.

Plasmid and strain constructions

We constructed a series of integrative plasmids that were sequentially introduced in the *C. albicans* strain that carries the BFP/GFP system (CEC2684, see Table S1 p.134) using the Lithium Acetate/Polyethylene Glycol protocol as described (449). See Text S1 for further details. *C. albicans* transformants were checked by PCR with a primer hybridizing to the plasmid sequence and a primer hybridizing to the gDNA in the region of insertion in order to verify proper integration of the plasmid in the *C. albicans* genome (Table S2 p.140).

To facilitate reading and understanding in the following manuscript, we name CEC4012 as "I-*Sce*I + TargetB", CEC4088 as "I-*Sce*I + TargetA", CEC4045 as "I-*Sce*I only", CEC3930 as "Target only", CEC4429 as "I-*Sce*I + TargetB + *GPI16*" and CEC4430 as "I-*Sce*I + TargetA + *GPI16"*.

Induction of the Tet-On system

In order to activate the *Tet*-On promoter and achieve I-*Sce*I protein overexpression, single colonies were grown overnight in SC-His-Arg medium. After a 16h growth, the cell cultures were diluted 10 times and grown for 8 hours in YPD in presence of anhydrotetracycline at a final concentration of 3µg/mL (ATc3) (393). The cells were then allowed to recover (*i.e.* to repair the DNA DSB) overnight by diluting 130 times the 8 hours-grown cells in fresh YPD

medium. ATc is commonly used for induction experiments and does not cause major defects in cell growth, morphology and biology (402). The cells were diluted 50 times into 1X PBS. A maximum of 10^6 cells were analyzed by flow cytometry using a MACSQuant® Analyzer (Miltenyi Biotec). The results were analyzed using FlowJo 7.6 software. The gates to determine the LOH frequencies were designed arbitrarily, but remained constant for all subsequent analyses.

Cell sorting

Single colonies from YPD plates were cultivated as presented above. Each culture was filtered with BD Falcon™ Cell strainers. Cells were diluted in 1X PBS at a final concentration of at least $20x10^6$ cells/mL. The MoFlo® Astrios™ flow cytometer was used to analyze and sort the cells of interest at low pressure (25 to 40%) and using a saline solution as a buffer (NaCl 0.9% OTEC). The flow cytometer is located at the Imagopole platform of the Institut Pasteur. A minimum of 1,000 cells were recovered into 400µL YPD in 1.5mL sterile Eppendorf tubes and stored at 4°C for the time of the experiment. This step can be preceded by an enrichment step consisting in sorting and recovering at high rate a maximum of cells first detected as positive in an 1.5mL sterile tube and perform a second sorting on these enriched populations to select with a higher accuracy the truly positive cells. The sorted cells were plated immediately after cell sorting on four YPD plates and incubated at 30°C for 48h. Single colonies were counted and then cultivated overnight in 1mL of fresh YPD at 30°C in 96-well plates. Aliquots were spotted on YPD, SC-Ura, SC-Arg+Uri and SC-His+Uri using a 48 or 96-well pin replicator and incubated at 30°C for 48h. This experiment was conducted twice.

PCR reactions

Each PCR reaction was performed in an Eppendorf Mastercycler ep gradients with 2µL 10X PCR buffer; 2µL $MgCl_2$ 50mM, 1.2µL mix of dNTP 2mM; 0.5µL of each primer 10µM; 0.2µL of Taq polymerase (Invitrogen™), either 1µL of DNA or traces of cells and water to reach a volume of 20µL. The following conditions were used: initial denaturation at 94°C for 3min, 30 cycles with denaturation at 94°C for 40s, annealing at 55 at 60°C for 40s and extension at 72°C for 1min/kb, and a final extension time at 72°C for 10min. The PCR products were verified by electrophoresis on a 1% or 2% agarose gel.

SNP-RFLP

Genomic DNA was extracted from cells coming from two independent cell sorting and 5-FOA experiments with the Epicentre Kit and used as a matrix in a PCR mix with primers located upstream and downstream of the SNP(s) of interest in order to assess their heterozygous or homozygous state. We used SNP156 (*Taq*I restriction site) located on the left arm of Chr4, between the telomere and the BFP/GFP system, and SNP95 (*Alu*I restriction site), located close to the centromere on the right arm of Chr4 (365). The PCR reactions were performed as detailed above.

MitoTracker staining

Cells were grown overnight in rich medium. The cultures were then diluted to an OD_{600}=0.2 in 50mL of liquid YPD and grown for 6 hours at 30°C. Once OD_{600}=1.2 had been reached, the cells were harvested and resuspended in 10mL of YPD. The cells were stained with MitoTracker (stock solution at 200µM, diluted 1:1,000 in the culture) for 45 min at 30°C. The cells were washed with sterile water and resuspended in 10mL of liquid YPG for 15min at 30°C. The cells were fixed in 4% paraformaldehyde diluted in PBS 1X.

Microscopy

The fixed cells were observed with a DMR XA LEICA fluorescence microscope using an oil immersion objective at x100 magnification (1.4 N/A). Single bandpass filters were used for Cy3 – Filter TX2 Leica (BP560/40). Images were captured with an ORCA II-ER cooled CCD camera (Hamamatsu). Cells were exposed for 250ms for the Cy3.

5-Fluoro-orotic acid (5-FOA) counter-selection

Single colonies were cultivated as detailed above. Dilutions of the cultures were plated on 5-FOA-containing plates (268). 200,000 cells of the strains carrying both the I-*Sce*I meganuclease and its target sequence were plated on 5-FOA containing plates after growth in ATc-free medium, while only 2,000 and 20,000 cells were plated after growth in presence of ATc3. Additionally, 200,000 cells of the control strains cultured in both induced and non-induced conditions were plated on 5-FOA containing plates. The dilutions were verified by plating a volume corresponding to 100 cells on YPD plates. Plates were incubated at 30°C for 48h. This experiment was conducted twice.

Sorbose counter-selection

Single colonies from two clinical strains, CEC2876 and CEC3673, along with the reference strain, SC5314, were cultivated overnight in rich medium. The cultures were diluted and $3x10^7$ to $3x10^3$ cells were plated on sorbose-containing plates (450, 451). The dilutions were verified by plating a volume corresponding to 100 cells on YPD plates. Plates were incubated at 30°C for 10 to 12 days (451).

Single colonies were patched on sorbose plates and a multiplex PCR at the *MLTa*/*MLTα* (452) locus was performed using primers AF120 to AF123 (Table S2 p.140), which allow the amplification of a fragment of 821 bp for *MLTa* and 515 bp for *MLTα*.

Whole genome sequencing

The genomic DNA was extracted with a phenol-chloroform method. The DNA samples were prepared with the Qubit® dsDNA BR assay kit following Thermo Fisher scientific recommendations and the DNA concentrations were estimated using the Qubit® Fluorometer. Genomic DNAs were processed to prepare libraries for Illumina sequencing. DNA was randomly fragmented by sonication to an average fragment length of 500 base pairs. Illumina adapters were blunt-end ligated to the fragments: the Nextera® XT DNA preparation kit (Illumina) was used according to the manufacturer's recommendations. MiSeq and HiSeq2500 platforms were used to generate respectively 300 and 250bp paired-ends reads. The sequences were mapped to the *C. albicans* strain SC5314 reference genome assembly 22, available from CGD (78, 453) using BWA v0.5.9 (454). Single nucleotide polymorphisms (SNPs) between the sequenced genomes and the reference genome were identified using GATK v3.1 (455) at positions with a sequencing depth equal or above 18X. Heterozygous SNPs were defined as positions where 15% or more of the calls showed one allele and 85% or less of the calls showed a second allele. Homozygous SNPs were defined as positions where more than 98% of the calls differed from the reference genome. Sequencing depth and heterozygosity/homozygosity density maps were constructed as described in Loll *et al*. (19) or Abbey *et al*. (456).

Data availability

Strains are available upon request. Table S1 p.134 contains genotypes for each strain.

Funding information

AF was the recipient of a Ph.D. grant from INRA Jouy-en-Josas and Institut Pasteur. GS received support from NIH grants R01-AI077737 and R01-DE015873.

Acknowledgements

We are grateful to all members of the Fungal Biology and Pathogenicity Unit for their daily advice, help and support and particularly Dr. Sophie Bachellier-Bassi for her critical reading of the manuscript as well as Murielle Chauvel for the technical help she provided. We thank Christiane Bouchier, Laurence Ma and Magali Tichit from the Plateforme Genomique at the Institut Pasteur for Illumina sequencing. We would also like to thank Prof. James Haber, Dr. Laurent Châtre and Dr. Benoit Arcangioli for helpful discussions. We thank Dr. Guilhem Janbon and Dr. Guy-Franck Richard for helpful discussions and critical reading of the manuscript.

AF, RLK, ML and CdE designed experiments. AF and PHC performed experiments. MEB, GS, KS, NS and CdE provided materials and sequencing data. AF, ML, CM and CdE analyzed data. AF, ML and CdE wrote the manuscript. All authors read and approved the final manuscript.

3. *Additional results: High resolution analysis of repair fidelity at an I-SceI-induced DNA DSB in C. albicans*[1]

DNA repair mechanisms have been extensively studied in the budding yeast, *S. cerevisiae*. As mentioned in the general introduction to this PhD thesis, repair mechanisms ensure cell survival and the choice of the repair pathway defines the fidelity, the rapidity and the length of the repair. In contrast to NHEJ, other molecular mechanisms such as BIR, GC or monosomy lead to tracts of homozygosity of variable lengths and can have dramatic effects, such as cell death, notably via the unmasking of deleterious alleles ((300), Feri *et al., accepted for publication*, mBio). However, LOH events play a double role as, in some circumstances, they can also be associated with the homozygosis of beneficial mutations, for example mutations rendering *C. albicans* cells resistant to an antifungal agent (281, 292-294, 299). In addition to LOH events, DNA DSB repair can result in the formation of partial aneuploidy, i.e. isochromosomes that consist in two inverted copies of a single arm of a chromosome flanking a centromere. Isochromosomes have also been shown to participate in antifungal resistance in *C. albicans* (270, 281).

BIR and GC (without or with CO) (Figure 22) have been described as highly mutagenic. This is primarily due to base substitutions or the use of a low fidelity polymerase (240). Moreover, the invading strand and its template DNA form a heteroduplex during the non-CO GC events (232) and mismatches occurring in the annealed region can be repaired by mismatch repair (MMR), thus generating LOH events (232). Upon BIR resulting from DNA DSBs induced by the *S. cerevisiae* HO nuclease, a 2,800-fold increase in frameshift mutations was reported compared to normal DNA replication (215). These mutations occurred close to the DNA DSB, where the replication machinery is the least stable, but, surprisingly, they were also observed far from the break site (215). Although MMR and the proofreading domain of the PolIII DNA polymerase allow for the correction of errors incorporated during replication, Deem et *al.* (215) observed that a reduced MMR activity, a lower fidelity of PolIII and an increase in the dNTP pool were associated to BIR, thus favoring mutagenesis. Additionally, semi-conservative DNA synthesis may lead to unstable replication upon BIR (214). Besides their mutagenic nature, BIR and GC with CO can be detrimental to diploid cells as they both

[1] This work involved the contribution of Murielle Chauvel, Christophe d'Enfert and Mélanie Legrand (Unité Biologie et Pathogénicité Fongiques, Institut Pasteur), Corinne Maufrais (C3BI, Institut Pasteur), Agnès Thierry (Groupe Régulation Spatiale des Génomes, Institut Pasteur).

generate long-range LOH events and are likely to reveal deleterious alleles ((172, 300), Feri *et al., accepted for publication*, mBio).

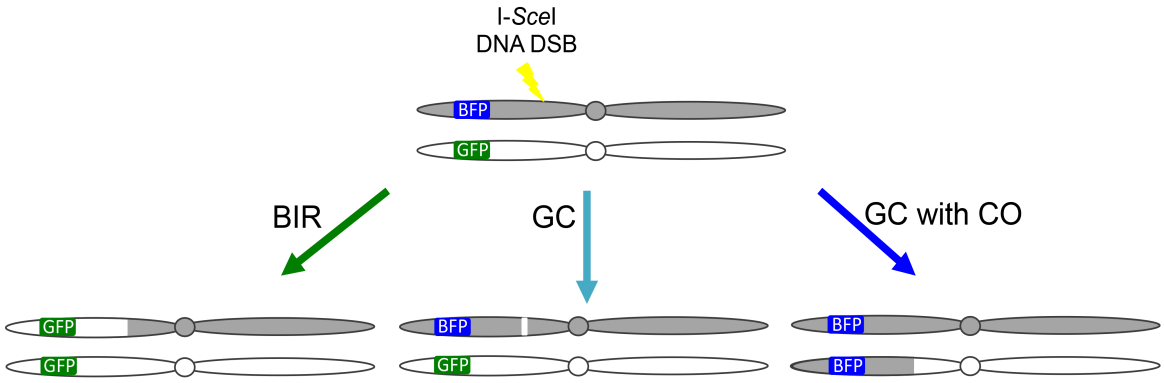


Figure 22: Distinction between homologous recombination outcomes by the BFP/GFP and I-*Sce*I systems combination
Upon I-*Sce*I-mediated DNA DSB, it is possible to determine which haplotype underwent a DNA DSB, and combined to the BFP/GFP system it is possible to distinguish between LOH resulting from BIR, GC without CO or GC with CO. In this example, if the DNA DSB occurs on the BFP-bearing chromosome, BIR would give rise to mono-GFP cells, GC without CO would remain doubly fluorescent and GC with CO would be at the origin of mono-BFP cells.

To date, fidelity of DNA DSB repair in *C. albicans* has not been investigated. This is in part because no method was available to generate a targeted DNA DSB in the *C. albicans* genome. This is also because BIR and GC with CO were indistinguishable due to the lack of suitable combinations of markers to distinguish them. As shown in the previous section (II.A.2 p.74; Feri *et al.*, *accepted for publication*), the combination of the I-*Sce*I-mediated DNA DSB-inducing system and LOH reporters has circumvented these limitations (Figure 22). Hence, we have investigated the fidelity of a DNA DSB repair upon BIR or GC with CO in *C. albicans*. To this aim, we have made use of the sequencing data obtained for strains having undergone BIR or GC with CO repair events following a DNA DSB induced by the I-*Sce*I meganuclease either on haplotype A or B of Chr4. Results presented below showed that homozygosis, interspersed with stretches of heterozygosity, spread, as anticipated, from the I-*Sce*I break site towards the telomere and, unexpectedly, extended from this break site towards the centromere. Notably, one progeny harbored a partial isochromosome that was the result of an I-*Sce*I-induced DNA DSB. Based on these observations, we discuss hypotheses regarding the functioning of repair pathways such as BIR and GC with CO in *C. albicans* that may have relevance to other organisms such as *S. cerevisiae*.

BIR is associated with events of homozygosis between the I-*Sce*I break site and the centromere

As described in the previous section (II.A.2 p.74), coupling of the BFP/GFP system (Figure 11A p.53) with the *S. cerevisiae* I-*Sce*I meganuclease to generate a DNA DSB (Figure 16 p.79) at a defined locus in the *C. albicans* genome allowed detecting, quantifying and sorting *C. albicans* cells that had undergone a long-range LOH, resulting from homologous recombination. Upon cell sorting, we could distinguish different types of events: (i) spontaneous events, (ii) I-*Sce*I-induced BIR, and (iii) I-*Sce*I-induced GC with CO events. In the same study, 5 *C. albicans* clones having undergone I-*Sce*I-dependent BIR or GC with CO were selected for both parental strains (*i.e.* carrying the I-*Sce*I target site either on Chr4A or Chr4B) and were genome sequenced. The normalized density of SNPs confirmed the molecular events predicted from FACS analysis and SNP-typing (Figure 23).

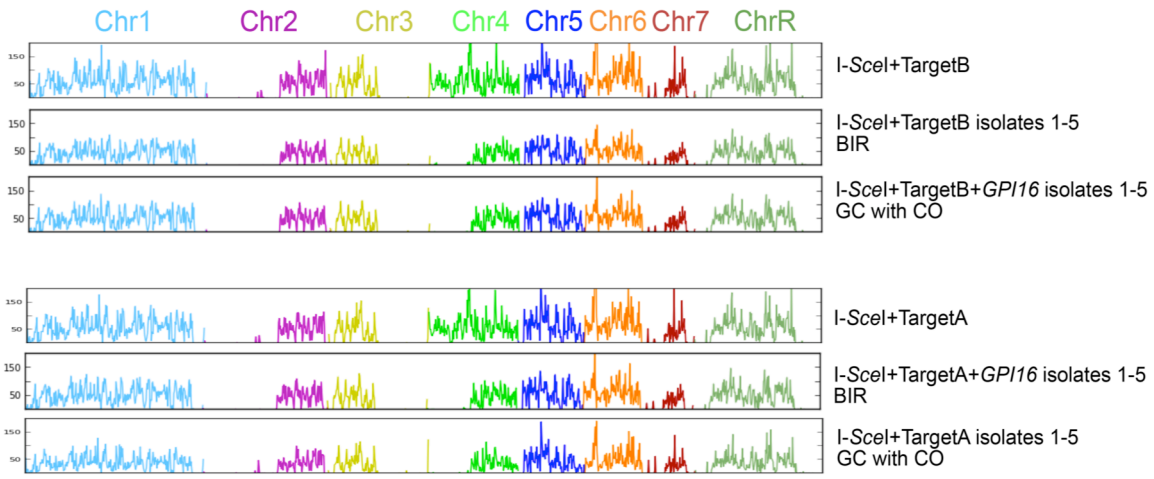


Figure 23: Normalized heterozygous SNP density of strains used in this study
The normalized heterozygous SNP density is represented by chromosome with each chromosome associated to a color. As expected the mono-fluorescent cells have undergone a LOH on Chr4 (either haplotype A or B according to the strain of interest). Note that a LOH on the left arm of Chr2 is also present in the parental strain of "I-*Sce*I+TargetA" and "I-*Sce*I+TargetB" as shown in Loll-Krippleber *et al.* Homozygosity of Chr3 right arm and Chr7 left and right arms are also observed in strain SC5314 as shown in Abbey *et al*.

In order to get a better view of the fidelity of the homology-directed repair mechanisms used in *C. albicans*, a detailed characterization of the region ranging from position 765,000 to 795,095 on Chr4A and from position 765,013 to 795,114 on Chr4B and encompassing the I-*Sce*I target site located at positions 778,062 on Chr4A and 778,081 on Chr4B was performed.
As shown in Figure 24A, only one of the five isolates (isolate 2) that had been predicted to have undergone a BIR upon I-*Sce*I-dependent cleavage of Chr4B displayed a LOH that

extended almost exactly from the I-*Sce*I site towards the telomere (from position 778,133 to the telomere). Isolates 1 and 3 showed homozygosis at the I-*Sce*I site and a LOH upstream of this site and extending towards the telomere, suggesting that they may be the result of independent GC and BIR events or had undergone template switching during BIR (see below). However, spontaneous BIR events are rare and are unlikely to have occurred here with such a high frequency. Finally and unexpectedly, isolates 4 and 5 presented an extension of the homozygous region up to 12.5 kb towards the centromere. These isolates may have occurred through BIR independently of I-*Sce*I cleavage or other mechanisms. The latter is more likely and putative mechanisms will be discussed below.

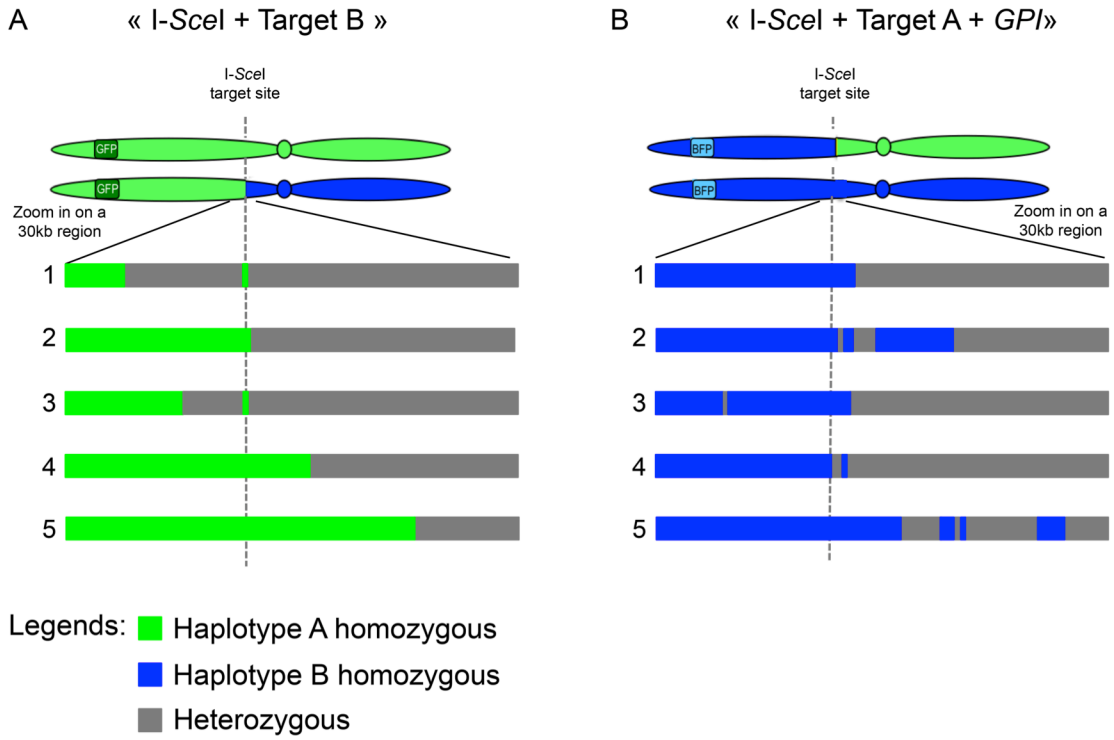


Figure 24: Results of the sequencing data analyses at the DNA DSB site in cells having undergone BIR
Representation of a 30 kb region surrounding the I-*Sce*I target site integration locus. Each of the 5 diagrams for each strain represents an independent isolate. The analysis of the cells that have undergone BIR resulting from a DNA DSB on Chr4B (« I-*Sce*I + TargetB » strain) and on Chr4A (« I-*Sce*I + TargetA + *GPI* » strain) revealed unexpected homozygosis between the target site and the centromere as well as heterozygous stretches within the homozygous region. Green: haplotype A homozygous, Blue: haplotype B homozygous, Gray: Both haplotypes A and B present (heterozygous).

As shown in Figure 24B, striking results were observed for the isolates that had undergone a BIR upon I-*Sce*I-dependent cleavage of Chr4A. Indeed, all isolates showed stretches of

homozygosity, interspersed with regions of heterozygosity, in the region extending from the I-*Sce*I site towards the centromere.

Results presented above were unexpected as they showed that among 10 isolates that appeared to have undergone BIR upon I-*Sce*I-mediated cleavage of Chr4, only one showed a LOH event that almost exactly extended from the I-*Sce*I site (Figure 24A – isolate 2). Moreover, they revealed unexpected events of LOH in the region extending from the I-*Sce*I site towards the centromere. To the best of our knowledge, no such detailed study of repair events occurring at a DNA DSB has been performed in *S. cerevisiae* or human cells, underlying the novelty of this observation. Below, we propose four mechanisms that may explain these observations.

(i) As mentioned earlier, mismatches that appear during strand invasion and annealing upon GC are subjected to repair by the MMR machinery (232). While in yeast, a minimum of 100 bp is sufficient for efficient homologous recombination (220), 500 to 1000 bp are necessary in mammalian cells (457-459). Despite the requirement for a minimum of 100 bp to perform efficient integrative transformation by homologous recombination in the *C. albicans* genome, the maximum length used for strand invasion in this species is not known.

Heterozygosity might influence the length needed for annealing and will promote repair by MMR. Consequently, LOH events of variable lengths between the I-*Sce*I site and the centromere may occur (Figure 25).

In *S. cerevisiae*, it was shown that, during GC, MMR is favored on the invading strand, leading to LOH (232). Therefore, if such mechanisms occur upon BIR in *C. albicans*, LOH events generated downstream the I-*Sce*I cutting site might reflect the repair of mismatches on the invading strand, followed by DNA synthesis using the homologous chromosome as a template and asynchronous synthesis of the lagging strand (as shown in (133)) using the newly synthesized DNA strand repaired via MMR (Figure 25).

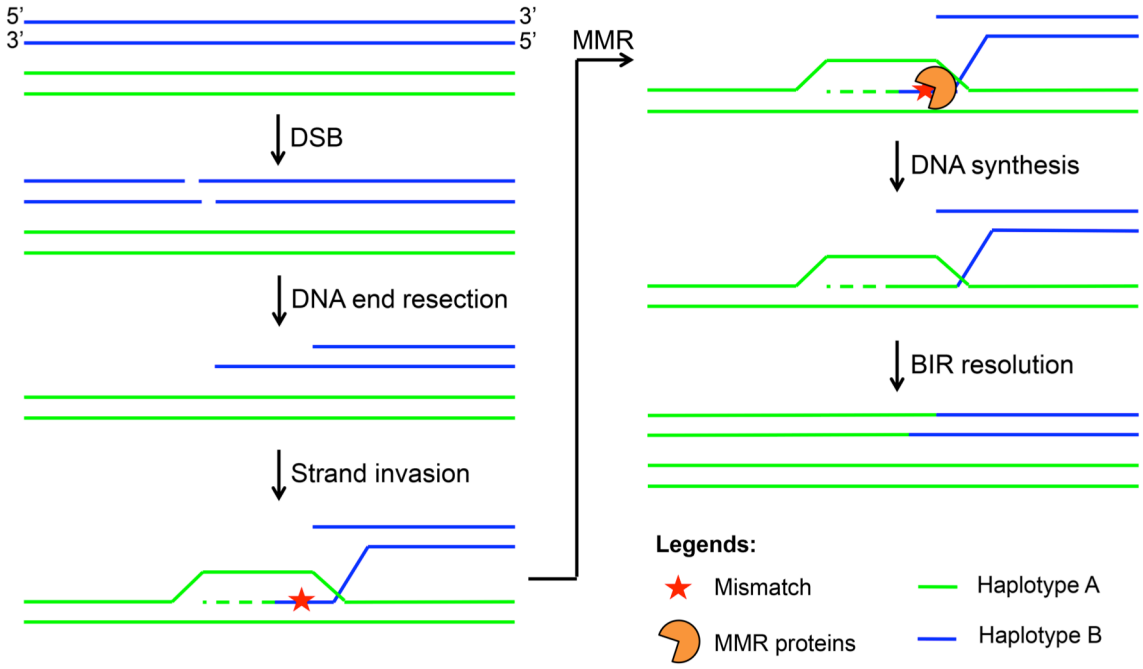


Figure 25: First mechanism – MMR heteroduplex correction during strand invasion
During BIR, the 3' end invades its template DNA. Heteroduplexes can be formed between the annealed sequences, resulting in the intervention of the MMR pathway to repair the mismatches, thus transforming a heterozygous region into a haplotype A homozygous region.

Extent of the LOH events between the I-*Sce*I target site and the centromere would depend on the length of the annealed homologous sequences and the DNA end resection of the 5' end (Figure 25). However, the occurrence of one strain carrying a LOH that extended almost exactly from the I-*Sce*I site towards the telomere cannot be explained by this hypothesis, suggesting the involvement of other mechanisms.

(ii) In all investigated isolates, generation of a DNA DSB was mediated by I-*Sce*I leaving 9 nucleotides at the 3' end of the break that are not complementary to any region on the homologous chromosome. Hence, to allow BIR, the 3' end that is not complementary to its homolog serving as template has to be removed. In *S. cerevisiae*, it is well known that the flap ends generated during repair by Alt-NHEJ and SSA are removed by the Rad1-Rad10 complex (197, 250, 251, 460, 461), a process that has been observed upon repair of an HO-induced DNA DSB at the mating type locus with 60 bp differing from the homologous chromosome (462).

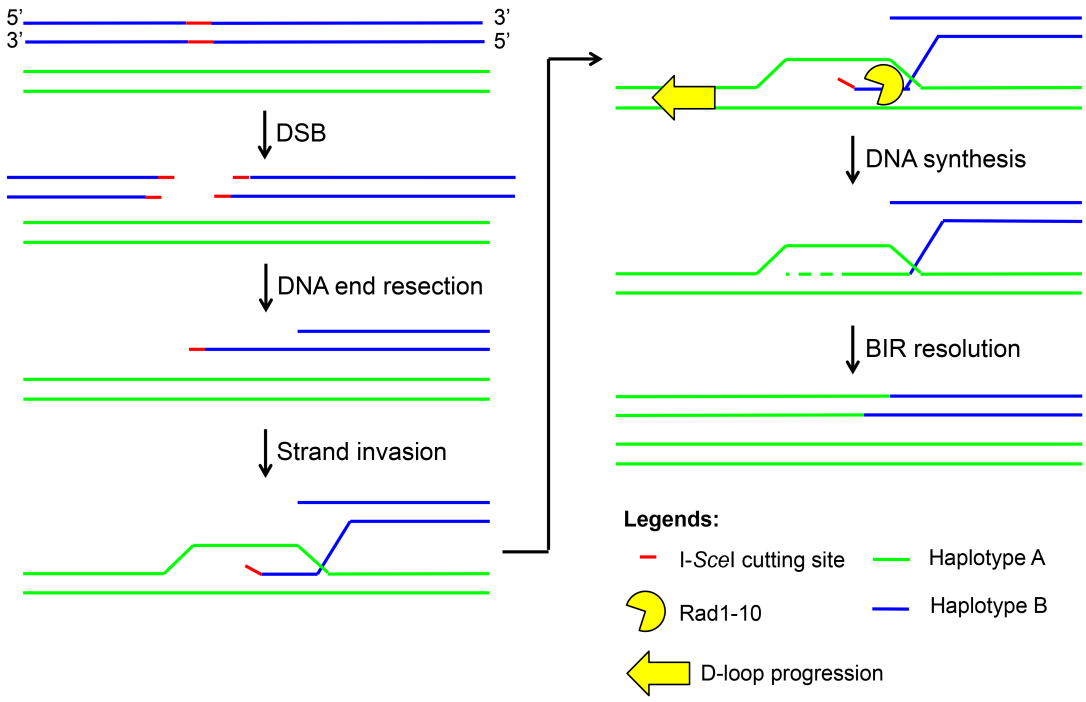


Figure 26: Second mechanism – Removal of flap ends by Rad1-Rad10
The I-*Sce*I-induced DNA DSB leaves 3' ends that are not homologous to its template DNA. Flap ends can be removed by the action of the Rad1-Rad10 proteins resulting in the occurrence of a haplotype A homozygous region.

Thus, a similar process might occur in *C. albicans* upon repair of an I-*Sce*I-mediated DNA DSB and lead to a LOH (Figure 26). Interestingly, the observed homozygous region between the break site and the centromere of *C. albicans* could reach a maximum length of 13 kb, while other strains displayed no additional homozygosis downstream the I-*Sce*I target site, thus suggesting that this proposed mechanism may not be systematically employed. Experiments are underway to test whether induction of a DNA DSB at the same locus using the CRISPR-Cas9 system (355), that does not release flap ends, leads to similar or different homozygosis profiles.

(iii) BIR has been described as a very slow process in *S. cerevisiae*. Therefore, the stability of the RPA proteins might be compromised when the repair is delayed. This would allow access of 3'-5' exonucleases to the single-stranded 3' end, degrading the 3' end of the broken strand until the homology search starts, consequently generating a longer homozygous region upon completion of DNA repair (Figure 27).

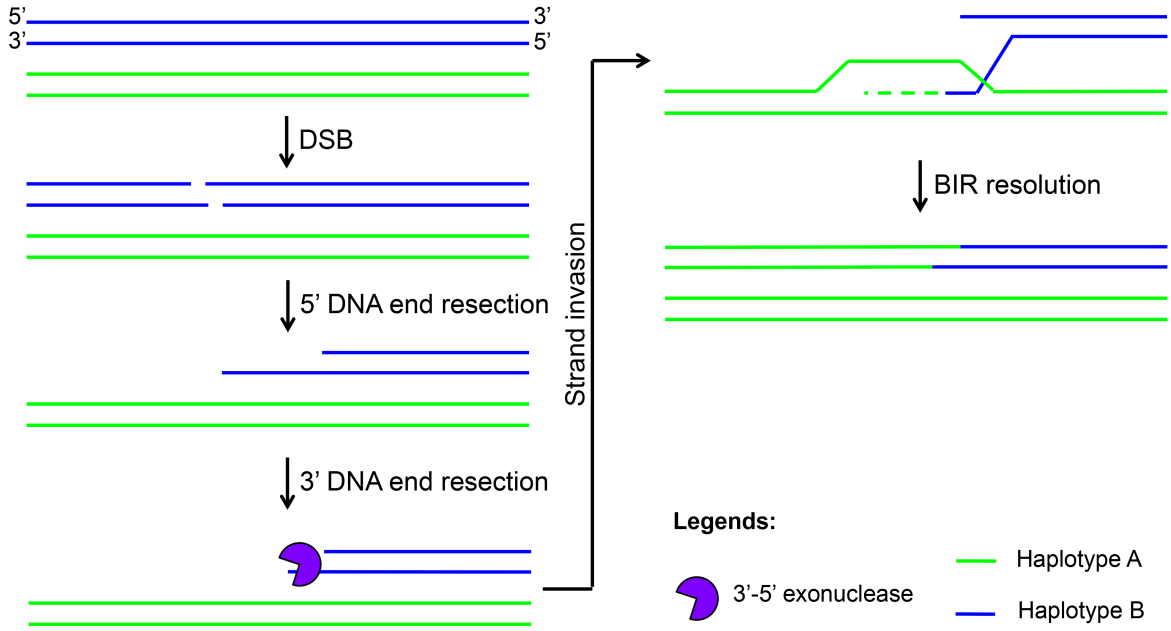


Figure 27: Third mechanism – 3' end DNA resection
BIR has been described as a very slow process allowing 3' ends to undergo non-negligible DNA end resection allowing the formation of a larger homozygous region that expected.

A 3'-5' degradation has been described in *S. cerevisiae* by Zierhut and Diffley (463). In addition to the 5'-3' end resection resulting in a single-stranded 3' end overhang, the 3' strand was also undergoing substantial resection but at a slower rate. Additionally, these authors hypothesized that the observed 3' end resection was due to the action of Mre11 belonging to the MRN complex and having both an endonuclease and 3'-5' exonuclease activity.

(iv) In *S. cerevisiae*, the mode of replication of DNA during BIR has been elucidated recently. DNA synthesis was found to be conservative but asynchronous between the leading and the lagging strand (133). In the event that DNA synthesis during BIR in *C. albicans* would be conservative but synchronous between the leading and the lagging strand, one might expect the formation of a heteroduplex at the region comprised between the break site on the 3' end and the resected 5' end (Figure 28). Subsequent generation of homozygosity tracts could involve repair by the MMR machinery of eventual mismatches between the 5' and the 3' strands of the repaired homolog (see Figure 25 p.110), or, as illustrated in Figure 28, through segregation of chromosomes upon mitosis if BIR occurs in G2.

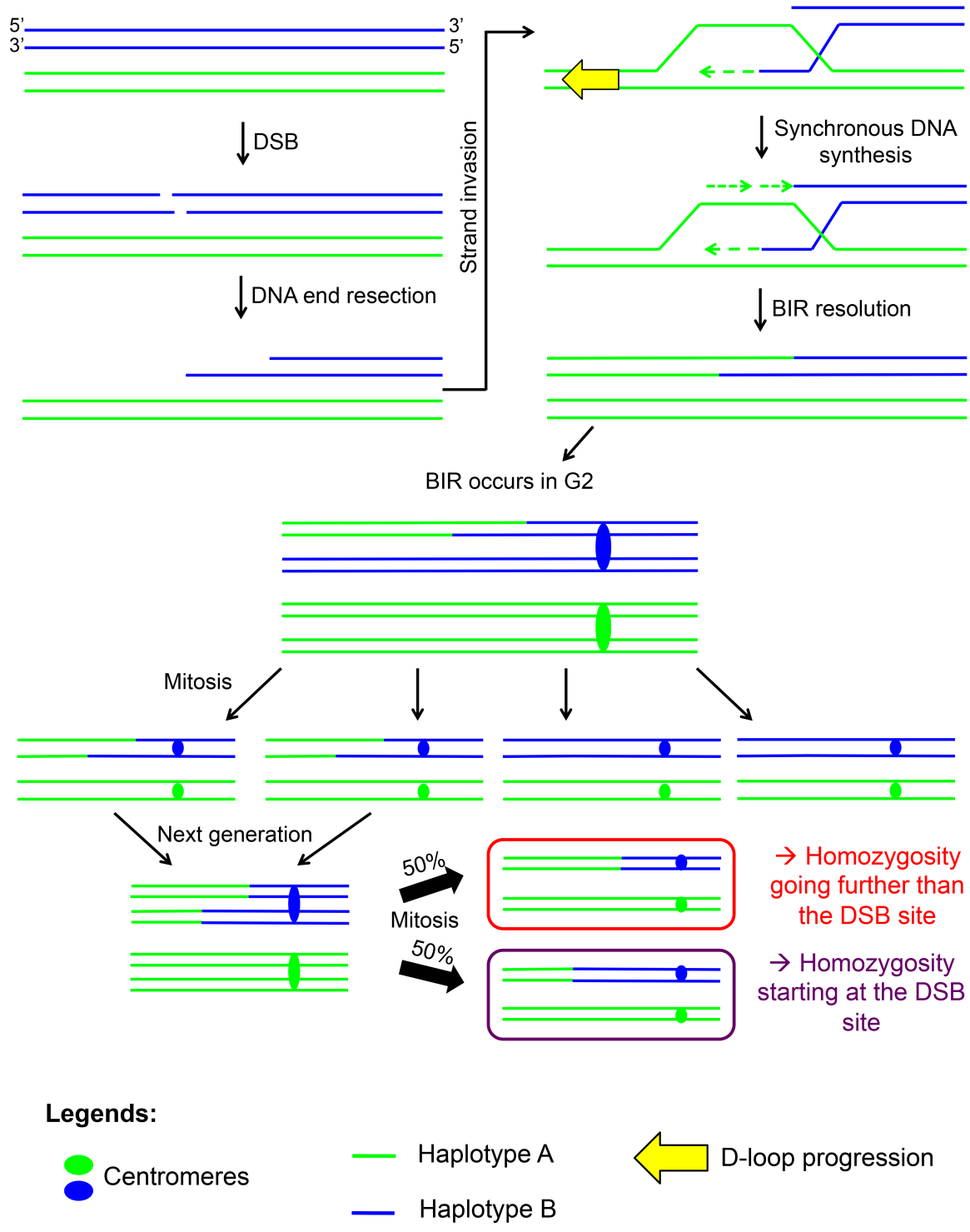


Figure 28: Fourth mechanism – DNA synthesis is conservative and synchronous between the leading and lagging strand

During strand invasion, the 3' end undergoes DNA synthesis, while the resected 5' end copies DNA from the 5' template strand involved in the D-loop. BIR repair pathway is known to be active during G2 phase, thus when the cells divide, 50% of the population harbors a BIR with a heteroduplex between the break site and the centromere. At the next mitotic division, 50% of the cells would carry a LOH resulting from a BIR starting at the break site and ending at the telomere, while the remaining 50% of the population would have undergone a BIR event going from a region in between the break site and the centromere to the proximal telomere.

Experiments are underway to test whether these proposed mechanisms are involved, through evaluation of the impact of deleting the *MSH2* gene, involved in MMR, the *MRE11* gene, involved in 3' end resection, the *SPO11* gene involved in recombination or the *POL32* gene, encoding a key player in DNA synthesis during BIR, on the homozygosis events occurring at the I-*Sce*I site upon BIR.

GC with CO: a highly mutagenic repair mechanism at the break site

Results presented in Figure 29 revealed that isolates that had undergone GC with CO displayed highly complex patterns of heterozygosity and homozygosity in the regions surrounding the I-*Sce*I cleavage site. Indeed, almost all isolates showed alternating regions of haplotype A homozygosity, haplotype B homozygosity and heterozygosity. It is unclear which mechanism(s) could explain such complex patterns.

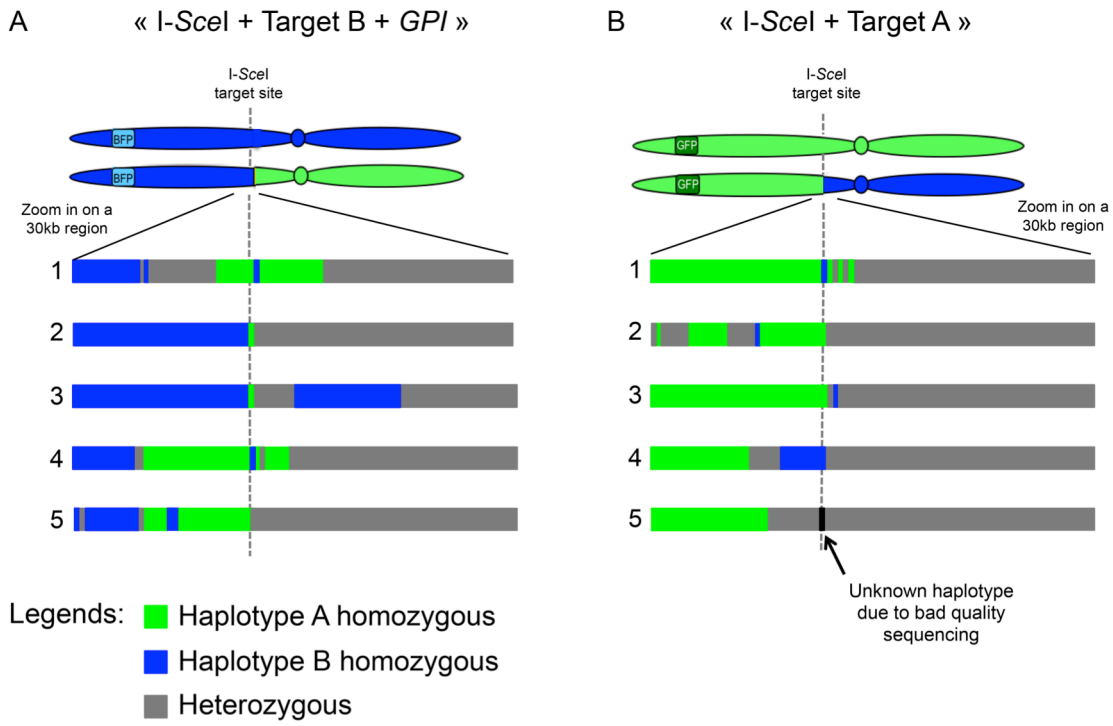


Figure 29: Results of the sequencing data analyses at the DNA DSB site in cells having undergone GC with CO

Representation of a 30 kb region surrounding the I-SceI target site integration locus. Each of the 5 diagrams for each strain represents an independent clone. The analysis of the cells that have undergone GC with CO resulting from a DNA DSB on Chr4B (« I-SceI + TargetB » strain) and on Chr4A (« I-SceI + TargetA » strain) revealed unexpected homozygosis from both haplotypes between the target site and the centromere as well as heterozygous stretches within the mainly homozygous region. Green: haplotype A homozygous, Blue: haplotype B homozygous, Gray: Both haplotypes A and B present (heterozygous).

Ongoing experiments aiming at testing the contribution of the *MSH2*, *POL32*, *SPO11, MRE11* and *MPH1* genes to repair an I-*Sce*I DNA DSB in *C. albicans* might help unravel these mechanisms.

Template switching in *Candida albicans* upon BIR and GC with CO

As mentioned above, stretches of heterozygous DNA within a mainly homozygous sequence were observed in several clones that repaired the I-*Sce*I-mediated DNA DSB by BIR (Figure 24 p.108) or GC with CO (Figure 29). Irregularities observed in the status of the newly synthesized strand have been reported in *S. cerevisiae*, and referred to as "template switching" (216, 217, 239). This process has been demonstrated in both BIR- and GC-repaired cells and has been linked to the Mph1 helicase and the instability of the PolIII polymerase for BIR-mediated template switching (217). Template switching was found to occur only within the 10,000 newly synthesized nucleotides adjacent to the break in the yeast *S. cerevisiae* (217). In *C. albicans*, the distance to which template switching was detected varied between strains and could be observed up to 9 kb from the break site in the case of a BIR (Figure 24A p.108 – isolate 1) and 13 kb upon GC with CO (Figure 29B – isolate 2), however this number is limited by the window that was selected to perform these analyses. Additionally, the accuracy offered by the genome sequencing data allowed detecting heterozygous regions whose sizes ranged from 1 nucleotide to 8.6 kb at the largest. Implication of template switching in the occurrence of these heterozygous regions is being tested using deletion mutant of the *POL32* and *MPH1* genes.

Heterozygous regions, resembling the template switching events, were also observed between the target site integration locus and the centromere. The tolerance threshold of MMR when exposed to a high number of heteroduplexes might be responsible for this observation. Indeed, upon the presence of numerous heteroduplexes, MMR could bypass successive heteroduplexes, allowing a succession of homozygous and heterozygous patches. Alternative hypotheses involve the synthesis of a lagging strand that is as instable as upon DNA synthesis of the leading strand (*i.e.* using a template other than the complementary strand).

An I-*Sce*I-mediated DNA DSB leads to isochromosome formation

In addition to the above-mentioned results, another consequence of the repair of a DNA DSB was observed in the form of an aneuploidy comprised between positions 778,000 and 996,000 of isolate 4 resulting from a BIR event following an I-*Sce*I-dependent DNA DSB on Chr4A (Table 4, Figure 24B p.108, Figures 30A and B p.117).

Table 4 – Analysis of the normalized coverage per 1 kb region for the 30 kb region of interest and for the entire Chr4

				Region 765,000-795,000 on Chr4		Entire Chr4						Region 765,000-795,000 on Chr4		Entire Chr4	
LOH	**Strains**	**Sequencing technology**	**Coverage**	**Mean of the normalized coverage per 1kb**	**Standard deviation**	**Mean of the normalized coverage per 1kb**	**Standard deviation**		**Strains**	**Sequencing technology**	**Coverage**	**Mean of the normalized coverage per 1kb**	**Standard deviation**	**Mean of the normalized coverage per 1kb**	**Standard deviation**
None	I-*Sce*I+TargetA	MiSeq 300bp paired-end	37.56	1.09	0.33	1.00	0.27		I-*Sce*I+Target B	MiSeq 300bp paired-end	47.40	1.01	0.26	0.99	0.26
BIR	I-*Sce*I+TargetA+*GPI* B+g- S1	HiSeq 250bp paired-end	197.26	0.96	0.20	0.99	0.56		I-*Sce*I+TargetB+*GPI* G+b- S1	HiSeq 250bp paired-end	279.26	1.02	0.15	0.99	0.32
BIR	I-*Sce*I+TargetA+*GPI* B+g- S2	HiSeq 250bp paired-end	112.19	0.99	0.18	0.99	0.49		I-*Sce*I+TargetB+*GPI* G+b- S2	HiSeq 250bp paired-end	306.70	1.03	0.18	1.04	0.42
BIR	I-*Sce*I+TargetA+*GPI* B+g- S3	HiSeq 250bp paired-end	85.78	1.04	0.22	1.02	0.53		I-*Sce*I+TargetB+*GPI* G+b- S3	HiSeq 250bp paired-end	294.11	1.05	0.16	0.97	0.31
BIR	I-*Sce*I+TargetA+*GPI* B+g- S4	HiSeq 250bp paired-end	712.80	1.63*	0.60	1.10	0.57		I-*Sce*I+TargetB+*GPI* G+b- S4	HiSeq 250bp paired-end	174.05	1.00	0.22	1.03	0.42
BIR	I-*Sce*I+TargetA+*GPI* B+g- S5	HiSeq 250bp paired-end	265.13	0.95	0.18	0.95	0.65		I-*Sce*I+TargetB+*GPI* G+b- S5	HiSeq 250bp paired-end	328.52	1.03	0.19	1.01	0.37
GC with CO	I-*Sce*I+TargetA+*GPI* G+b- S1	MiSeq 300bp paired-end	31.91	1.12	0.27	1.00	0.32		I-*Sce*I+TargetB+*GPI* B+g- S1	HiSeq 250bp paired-end	88.34	1.02	0.29	1.05	0.71
GC with CO	I-*Sce*I+TargetA+*GPI* G+b- S2	MiSeq 300bp paired-end	23.60	1.13	0.26	1.01	0.32		I-*Sce*I+TargetB+*GPI* B+g- S2	HiSeq 250bp paired-end	160.76	1.04	0.28	1.05	0.64
GC with CO	I-*Sce*I+TargetA+*GPI* G+b- S3	MiSeq 300bp paired-end	33.00	1.19	0.22	1.02	0.30		I-*Sce*I+TargetB+*GPI* B+g- S3	HiSeq 250bp paired-end	139.40	0.99	0.30	1.05	0.59
GC with CO	I-*Sce*I+TargetA+*GPI* G+b- S4	MiSeq 300bp paired-end	31.47	1.13	0.23	1.00	0.30		I-*Sce*I+TargetB+*GPI* B+g- S4	HiSeq 250bp paired-end	108.54	1.06	0.27	1.04	0.59
GC with CO	I-*Sce*I+TargetA+*GPI* G+b- S5	MiSeq 300bp paired-end	29.06	1.08	0.24	1.00	0.30		I-*Sce*I+TargetB+*GPI* B+g- S5	HiSeq 250bp paired-end	388.71	1.06	0.20	1.04	0.54

The normalization was calculated in log2, indicating the copy number variation. Values equal to approximately 1 correspond to 2 copies, and values equal to 2, correspond to 4 copies. *Value higher than 1.

Given that the I-*Sce*I target site is located at position 778,062 and the centromere is located between positions 992,473 and 996,110, these data suggest the formation of a partial isochromosome or an episome as observed in Thierry *et al.* (464), duplicating a 220 kb long region following a DNA DSB. However as episomes have never been observed in *C. albicans*, the occurrence of an isochromosome seems the most reasonable hypothesis. While the centromere appeared present in 3 copies, the upstream region was found in 4 copies (Figures 30A and B).

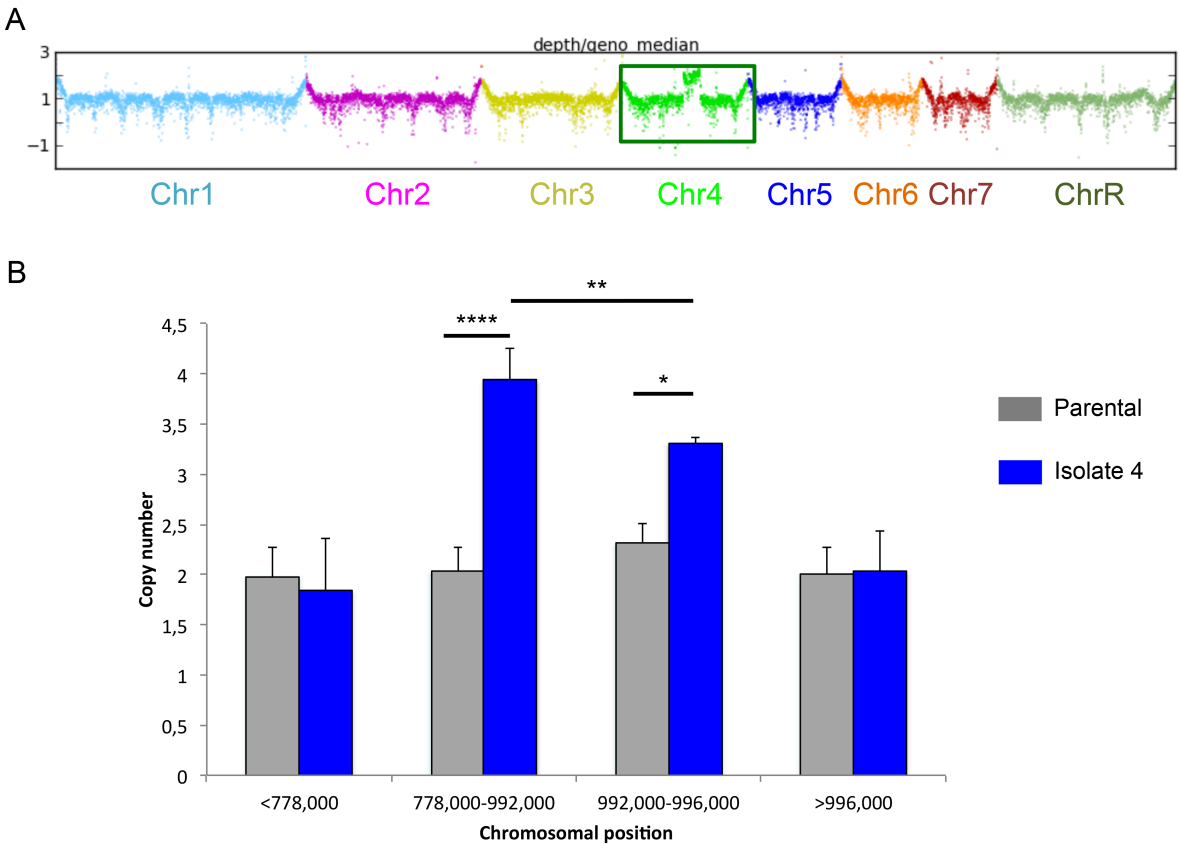


Figure 30: Sequencing depth normalized to the genome coverage median (log2 representation) per 1 kb region
(A) The normalization allows quantifying the number of gene copies. 1 corresponds to 2 copies, and 2 to 4 copies as observed for the portion of Chr4. (B) A detailed analysis of the median coverage by region of 1 kb allowed to confirm the presence of 4 copies of the region between the target site (position 778,062) and the 5' end of the centromere (992,412) and 3 copies of the centromere (992,473-994,110), while the rest of the genome is present in two copies (as expected).

The strain of interest is mono-fluorescent suggesting that it has undergone both BIR and the formation of an isochromosome. Hence, the duplication of the 220 kb region started at the break site, encompassing the centromere, at the origin of the formation of an isochromosome while the repair of the DNA DSB is made via DNA synthesis from the break site to the

proximal telomere. Isochromosomes have been found to be involved in human genetic diseases (for examples: (465-467)) and *C. albicans* antifungal resistance (270, 281). The presence of such a chromosome will be evidenced by the separation of chromosomes by a contour-clamped homogeneous electrophoretic field (CHEF) (368). Interestingly, in *S. pombe*, isochromosomes were observed upon GC failure, and identified the orthologs of Mre11-Rad50-Xrs2 complex, Rad51, Rad55 and Rad57 as repressors of break-induced isochromosomes (288). The authors also suggested that the isochromosome was the result of a BIR event involving the ortholog of Pol32 following an extensive end resection that allowed exposing single-stranded inverted repeats at the origin of isochromosome formation. The observation made in *C. albicans* is surprising and, to our knowledge, has never been described before: indeed, the isochromosome would be formed by a fragment of the left arm of Chr4 while the parental chromosome is still present and has been repaired by BIR. Once confirmed, this isochromosome will be isolated and *de novo* sequenced to better understand its origin.

In summary, our detailed characterization of *C. albicans* clones having repaired an I-*Sce*I-induced DNA DSB by BIR or GC with CO has revealed unexpected and frequent genomic rearrangements at the break site as well as the formation of a truncated isochromosome. Several hypotheses have been raised in order to explain these rearrangements and experiments using selected knock-out mutants are ongoing to assess the contribution of different repair pathways and template switching to these rearrangements. It is striking that, to our knowledge, such complex, bi-directional, rearrangements have not been reported in other species than *C. albicans*. This may reflect our use of genome sequencing that allows characterizing these events at the nucleotide scale. This may also reflect the use of a heterozygous diploid background while many of the studies that have investigated DNA DSB repair used homozygous diploid contexts except for specific markers. This may finally reflect divergence between *C. albicans* and the other species in which DNA repair mechanisms have been investigated. In any case, these data confirm that *C. albicans* is a valuable tool to study genome stability in eukaryotes.

B. A genetic screen reveals new players in the maintenance and imperilment of genome stability.

1. Introduction

In *S. cerevisiae*, several genetic screens have been performed to assess the impact of gene deletion on genome stability (468-470). Andersen *et al.* looked at the LOH rate occurring at three different loci in the genome of *S. cerevisiae* and revealed that, no matter the locus, the deletion of *HEX3*, *POL32*, *RAD27*, *RAD50*, *RTT109*, *SIC1*, *SLX8* and *XRS2* resulted in a high fold-increase in the number of LOH events. Deletion of other genes triggered locus-specific LOH rate increase, such as *CSM1* or *LOH1* that are specific to the *MET15* locus, while *ICE2* and *RAD6* deletion did not affect LOH at the *MAT* locus (381). As presented earlier, the difficulty to obtain deletion mutants in *C. albicans* encouraged the development of alternative tools to study genome biology. In particular, resources for the development of genome-wide overexpression screens are being established (471) and can be used to search for *C. albicans* genes whose overexpression impacts genome maintenance. These resources include a *C. albicans* ORFeome whereby >85% of the *C. albicans* ORFs have been cloned in the Gateway™ pDONR207 vector (M. Legrand, K. K. Lee, S. Bachellier-Bassi, Y. Chaudhari, H. Tournu, L. Arbogast, H. Boyer, M. Chauvel, V. Cabral, A. Nesseir, I. Pelinska, E. Permal, L. A. Walker, U. Zeidler, S. Znaidi, G. Larriba, P. Van Dyck, C. A. Munro, and C. d'Enfert, *in preparation*). This allows (i) the transfer using Gateway™-mediated recombinational cloning of the cloned ORFs in *C. albicans*-adapted expression vectors, yielding collections of expression plasmids, and (ii) the transformation of these plasmids in derivatives of *C. albicans* strain SC5314, yielding collections of overexpression strains (402, 472). Hence, when such expression plasmids are introduced in a *C. albicans* strain that harbors reporters of genome dynamics such as the FACS-optimized LOH reporter developed by Loll-Krippleber *et al.* (20) (appendix 1) it is possible to screen for genes whose overexpression impacts genome stability in *C. albicans.* This has been demonstrated by Loll-Krippleber *et al.* (20) who used a collection of 124 tetracycline-dependent overexpression plasmids for genes with a predicted role in DNA repair, recombination and replication and showed that overexpression of *BIM1*, *RAD53*, *RAD51* and *CDC20* increases LOH rate at the artificial BFP/GFP locus placed on Chr4.

Work by Forche *et al.* (18) has revealed that the nature of the LOH changes according to the stress the cells are exposed to. (18). In *C. albicans* genomic rearrangements such as

aneuploidies and, more importantly, LOH events are triggered by different forms of stress and some of these genomic rearrangements, in selective conditions, can lead to better adaptation to stressful environments. Thus, deciphering the pathways that mediate stress responses into genomic rearrangements in *C. albicans* may lead to (i) a better understanding of genome stability in eukaryotic species and (ii) the identification of new target genes whose drug-mediated deregulation could destabilize *C. albicans* survival in the stressful context of infections.

Thus, in the following sections, we describe the implementation of screens for genes involved in the maintenance and imperilment of genome stability in *C. albicans*. To this aim, we generated a collection of 564 overexpression plasmids carrying genes involved in signaling pathways (402) and DNA processes (20). These genes were placed under the control of a constitutive promoter and were integrated at the *RPS1* locus on *C. albicans* Chr1 in a strain carrying both the I-*Sce*I-mediated DNA DSB inducing system (see section II.A.2 p.74) and the BFP/GFP system (Figures 11 p.53 and 16 p.79) ((20) and Feri *et al.*, *accepted for publication*). The impact of gene overexpression was assessed in the absence of induction of a DNA DSB, thus allowing the identification of genes increasing LOH rate. Alternatively, as the basal LOH rate is low, the impact of gene overexpression was assessed upon induction of a DNA DSB in order to identify genes whose overexpression results in the stabilization of the *C. albicans* genome by decreasing the high LOH rate associated to I-*Sce*I activity. These two screens have led to the identification of 7 genes whose overexpression triggers genome instability and 6 genes maintaining genome integrity upon overexpression.

2. *Material and methods*

Strains and media

The *C. albicans* strains used in this study are derived from SN148 (266) and can be found in Table S1 p.134. Yeast cells were grown at 30°C in liquid media, either in YPD (1% yeast extract, 2% peptone, 2% dextrose) or SC (0.67% Yeast Nitrogen Base without amino acids, 2% dextrose supplemented with the appropriate 0.08% drop out mix of amino acids). Solid media were obtained by adding 2% of agar.

Construction of overexpression plasmids

The p*CaTDH3*-GTW-*LEU2* plasmid carries the sequence for integration at the *RPS1* locus on Chr1, the *C. maltosa* optimized *LEU2* marker and a Gateway® Cassette flanked by the *attR* sequences and under the control of the P_{TDH3} constitutive promoter. To obtain this plasmid,

the vector pFA-*CmLEU2* (473) was digested with *Pvu*II and *Spe*I, the band corresponding to the *LEU2* marker was purified and cloned into the *Ale*I-*Spe*I double-digested p*CaTDH3*-GTW-*URA3* plasmid (19). The resulting plasmid was called p*CaTDH3*-GTW-*LEU2*.

The collection of overexpression plasmids was constructed by using the Gateway™ Technology as described in Cabral *et al.* (402, 472) by LR recombination between the BP vectors for 588 ORFs, combining Chauvel and Loll-Krippleber collections (20, 402), and the destination vector pCa*TDH3*-GTW-*LEU2*. p*CaTDH3*-GTW-*LEU2* derivatives for 564 *C. albicans* ORFs were obtained.

C. albicans *strain constructions*

*Stu*I-digested or I-*Sce*I-digested p*CaTDH3*-GTW-*LEU2* derivatives were transformed into strain CEC4012 (Feri *et al., accepted for publication*, mBio) as described by Cabral *et al.* (472), using the Lithium Acetate/Polyethylene Glycol protocol as described in (449) yielding a collection of 564 overexpression strains. The control strain is CEC4012 transformed with the *Stu*I-digested p*CaTDH3*-GTW-*LEU2* vector. *C. albicans* transformants were checked by PCR with a primer hybridizing to the plasmid sequence and a primer hybridizing to the gDNA in the region of insertion in order to verify proper integration of the plasmid in the *C. albicans* genome. The list of genes that are overexpressed is presented in Table S3 p.145.

Cell preparation for flow cytometry and analysis

Single colonies from YPD plates were cultivated 16 hours into YPD medium at 30°C. Cells were then diluted 5 times and grown for 8 hours in YPD in the presence or absence of anhydrotetracycline at a final concentration of 3µg/mL (ATc3) (393). ATc3 allows activating the *Tet*-On promoter and achieving I-*Sce*I protein overexpression. Then, cells were allowed to recover for 16 hours at 30°C by diluting 66 times in fresh YPD medium. The cells were then diluted 50 times into 1X PBS. A maximum of 10^6 cells were analyzed by flow cytometry using a MACSQuant® Analyzer (Miltenyi Biotec). The results were analysed using FlowJo 7.6 software. The gates to determine the LOH frequencies were designed arbitrarily, but remained constant for all subsequent analyses. The screen was performed 3 times and each clone whose LOH rate was at least 2-fold higher or lower than the control strain for each replicate was kept as a candidate.

3. *Results and discussion*

Construction of a collection of overexpression strains for the identification of *C. albicans* genes with a role in the regulation of LOH

In this study, we aimed to identify genes whose overexpression could positively or negatively impact the rate at which LOH events occur in the *C. albicans* genome. To this aim, we first established a collection of overexpression plasmids carrying genes whose gene ontology annotations or orthologous function in *S. cerevisiae* indicated a role in DNA processes (repair, replication and recombination) or in signaling (transcription factors, protein kinases, protein phosphatases). ORFs in 588 pDONR207 derivatives described in Loll-Krippleber *et al.* (20) and Chauvel *et al*. (402) were transferred by the Gateway™ technology into p*CaTDH3*-GTW-*LEU2*, yielding a total of 564 (95.9%) p*CaTDH3*-GTW-*LEU2* derivatives. In a second step, we generated a collection of *C. albicans* overexpression strains by transforming strain CEC4012 (see section II.A.2 p.74) with the collection of 564 overexpression plasmids. In this setting, p*CaTDH3*-GTW-*LEU2* derivatives are integrated at the *RPS1* locus in *C. albicans* strain CEC4012 that carries (i) a FACS-optimized LOH reporter at the *PGA59-62* locus on Chr4 (20), (ii) the I-*Sce*I gene under the control of the tetracycline inducible promoter at the *HOG1-HOL1* locus (Feri *et al., accepted for publication,* mBio); (iii) an I-*Sce*I target site at the *CDR3-tG(GCC)2* locus on Chr4; and (iv) the pNIMX plasmid encoding the transactivator for the tetracycline inducible promoter. Overall, 564 *C. albicans* overexpression strains were obtained (see Table S3 p.145) each displaying constitutive overexpression of a selected ORF, tetracycline-dependent expression of I-*Sce*I, and a reporter of LOH on Chr4 whose LOH rate is influenced according to I-*Sce*I levels.

Genes that trigger genome instability upon overexpression under normal growth conditions

The entire overexpression collection was analyzed by flow cytometry and the LOH rates were measured by surveying 10^6 cells in triplicates for each strain. Under normal growth conditions (YPD) whereby I-*Sce*I is not induced, the occurrence of overexpression strains with an increased number of mono-fluorescent cells was anticipated.

Table 5 – Fold change of the LOH rate per replicate and candidate strain, both under normal growth conditions and in presence of ATc.

	C. albicans	*S. cerevisiae*	Replicate	B+g-		b-G+		LOH rate
				Fold-change	Mean	Fold-change	Mean	
Candidates in normal growth conditions	C1_08570C	*PCL2*	Replicate1	7.1	6.5	6.3	4.6	Increase
			Replicate 2	8.3		5.2		
			Replicate 3	4.0		2.3		
	C1_04380W	*SIT4*	Replicate1	16.7	13.7	20.0	10.8	Increase
			Replicate 2	17.6		9.1		
			Replicate 3	6.8		3.5		
	C3_04250W	-	Replicate1	23.3	11.4	17.2	9.8	Increase
			Replicate 2	3.8		3.8		
			Replicate 3	7.0		8.5		
	C3_03810W	*RAD53*	Replicate1	2.4	3.0	3.0	4.0	Increase
			Replicate 2	2.3		2.0		
			Replicate 3	4.2		6.9		
	C2_00350W	*RAD5*	Replicate1	8.0	12.0	6.1	5.0	Increase
			Replicate 2	24.7		3.6		
			Replicate 3	3.3		5.2		
	C6_01150W	*CDC20*	Replicate1	13.0	13.4	12.4	9.6	Increase
			Replicate 2	16.0		8.7		
			Replicate 3	11.3		7.8		
Candidates upon DNA breaks induction	C5_02440C	*BUD32*	Replicate1	-29.3	-23.0	-82.0	-55.0	Decrease
			Replicate 2	-14.3		-22.0		
			Replicate 3	-25.6		-60.9		
	C7_02220C	-	Replicate1	-13.9	-15.9	-34.2	-44.6	Decrease
			Replicate 2	-17.6		-55.9		
			Replicate 3	-16.4		-43.8		
	C1_02120C	*SKS1*	Replicate1	2.4	2.4	2.9	3.0	Increase
			Replicate 2	2.4		2.9		
			Replicate 3	2.5		3.1		
	C6_02160W	-	Replicate1	-3.2	-4.5	-10.6	-9.9	Decrease
			Replicate 2	-6.0		-8.3		
			Replicate 3	-4.2		-10.7		
	CR_06420W	*PPZ1*	Replicate1	-3.1	-6.1	-3.4	-6.2	Decrease
			Replicate 2	-3.7		-3.5		
			Replicate 3	-11.6		-11.8		
	C6_01190C	*VPS15*	Replicate1	-8.7	-13.7	-32.2	-24.0	Decrease
			Replicate 2	-10.2		-15.0		
			Replicate 3	-22.0		-24.7		
	CR_06110C	*RFX1*	Replicate1	-8.3	-14.6	-17.4	-28.6	Decrease
			Replicate 2	-11.1		-33.6		
			Replicate 3	-24.4		-34.8		

Overexpression strains were selected when they had at least a 2-fold increase in the LOH rate for each replicate when compared to the control strain. Thus, we identified C1_08570C (*PCL2*); C1_04380W (*SIT4*); C3_04250W (*WSC1*); C3_03810W (*RAD53*); C2_00350W (*RAD5*) and C6_01150W (*CDC20*) as genes whose overexpression generates genome instability for both cell populations, i.e. mono-BFP and mono-GFP cells (Table 5 and Figure 31A).

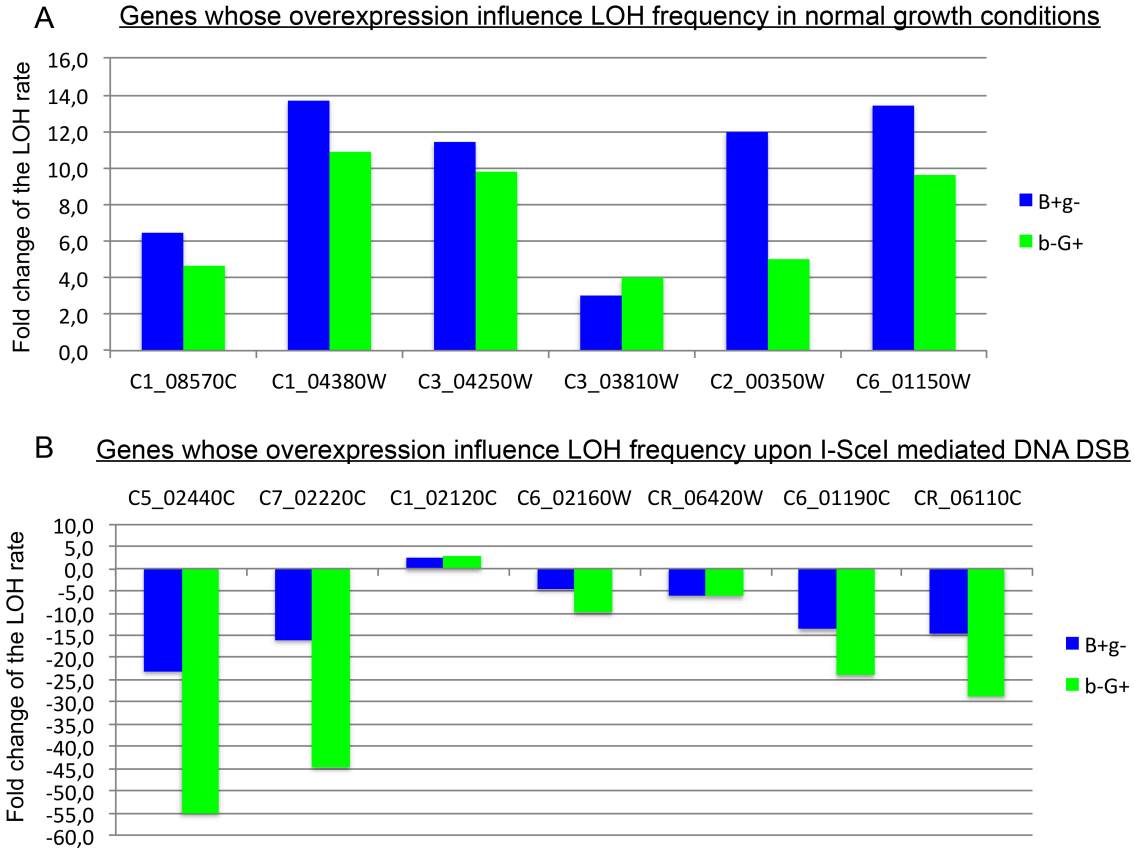


Figure 31: 13 candidate genes playing a role in the genome dynamics of *C. albicans.*
(A) Fold increase of the LOH rate upon overexpression of candidate genes in normal growth conditions. (B) Fold decrease and increase of the LOH rate upon overexpression of candidate genes in the context of I-*Sce*I induced DNA DSB.
The fold-change of the mono-BFP population is represented in blue while the fold-change of the mono-GFP is represented in green. In this graph, the mean values are represented for each strain and each population.

Interestingly, both *RAD53* and *CDC20* genes were previously identified in the screen performed by Loll-Krippleber *et al.* (See appendix 1, (20)). Indeed, *RAD53*, whose gene product is involved in the DNA damage checkpoint, and *CDC20*, which encodes an activator of the metaphase-anaphase transition, are major players in the maintenance of genome integrity (See appendix 2; (19, 474-477)). In contrast, we did not observe an increase in LOH rates upon overexpression of *BIM1*, encoding a component of microtubules and *RAD51*, encoding a protein involved in the DNA damage checkpoint, as seen by Loll-Krippleber *et al.* (20). Differences in the level of expression driven by the *TDH3* and *TET-on* promoters or in the response of cells to constitutive vs regulated expression might explain this discrepancy.
In *C. albicans*, Sit4 is a serine-threonine phosphatase catalytic subunit that plays a role in cell wall maintenance and morphogenesis; Wsc1 encodes a putative sensor-transducer of the

stress-activated Pkc1-Mpk1 signaling pathway; Pcl2 is a cyclin homolog; and Rad5, a putative DNA-dependent helicase. Among these genes, genetic interaction has been demonstrated in *S. cerevisiae* for Rad5 and Rad53. From a functional standpoint it has been shown that upon DNA DSB, leading and lagging strand DNA synthesis are uncoupled and while Rad53 may limit progression of the uncoupled fork to leave time for the repair, it also facilitates stalled fork restart via Rad5 (478).

In addition to the previously cited genes, additional overexpression strains were identified as having a validated increase in the LOH rate uniquely for one cell population, i.e. either mono-BFP or mono-GFP. We identified CR_02120C encoding a C2H2 zinc finger transcription factor; C2_07210C (*TPK2*), a protein kinase catalytic subunit involved in morphogenesis; C3_00670C (*FKH2* - *ScFKH1*), a forkhead transcription factor involved in morphogenesis; C7_03940C (*CLB4*) and C2_09370C (*KAR3*) respectively encoding a cyclin and a kinesin-like microtubule motor protein, as potential players in genome instability (Table 6). Differences in the increase of the two mono-fluorescent populations in these overexpression strains might reflect synthetic lethality between overexpression of the corresponding ORFs and recessive alleles that have been revealed upon homozygosis of one of the two haplotypes only.

It is striking that a number of the identified genes have cell wall- or morphogenesis-related functions. Overexpression of these genes might influence cell shape and consequently the aforementioned results (filaments might distort the flow cytometry outputs). Characterization of the overexpression strains for cell shape and morphogenesis remains to be undertaken.

Table 6 – Fold change of the LOH rate per replicate and putative candidate strain, both under normal growth conditions and in presence of ATc.

	C. albicans	*S. cerevisiae*	Replicate	B+g-		b-G+		Summary	
				Fold-change	Mean	Fold-change	Mean		
Candidates in normal growth conditions	CR_02120C	-	Replicate1	8.9	11.5	-2.7	-3.2	Increase	Decrease
			Replicate 2	11.3		-2.8			
			Replicate 3	14.4		-4.0			
	C2_07210C	*TPK2*	Replicate1	144.2	144.3	1.8	1.4	Increase	
			Replicate 2	187.2		0.6			
			Replicate 3	101.3		1.7			
	C3_00670C	*FKH1*	Replicate1	2.7	2.8	1.9	3.0	Increase	
			Replicate 2	2.8		4.2			
			Replicate 3	2.9		2.9			
	C7_03940C	*CLB4*	Replicate1	2.4	3.5	1.6	2.9	Increase	
			Replicate 2	3.4		1.8			
			Replicate 3	4.8		5.3			
	C2_09370C	*KAR3*	Replicate1	3.2	4.1	2.4	3.2	Increase	
			Replicate 2	7.2		4.8			
			Replicate 3	1.8		2.3			
Candidates upon DNA breaks induction	CR_02120C	-	Replicate1	2.1	2.3	-10.4	-10.8	Increase	Decrease
			Replicate 2	2.5		-9.8			
			Replicate 3	2.3		-12.2			
	C5_03310C	*PCL1*	Replicate1	1.3	1.1	-2.8	-7.1	Decrease	
			Replicate 2	1.0		-10.7			
			Replicate 3	1.0		-7.7			
	C6_01150W	*CDC20*	Replicate1	3.8	4.0	1.6	1.5	Increase	
			Replicate 2	3.5		1.1			
			Replicate 3	4.7		1.8			

Genes that trigger genome maintenance or imperilment upon overexpression in the presence of I-*Sce*I-induced DNA DSB.

Because spontaneous LOH rates are low (2.10^{-4} on average), identification of genes whose overexpression would decrease these rates is difficult. Thus, we took advantage of the presence in the collection of overexpression strains of a conditional I-*Sce*I-dependent DNA DSB-inducing system, to increase LOH rates at the BFP/GFP LOH reporter and search for genes whose overexpression would decrease these rates. The collection of 564 overexpression strains was analyzed as described above, except that cells were grown in the presence of ATc3 in order to allow production of I-*Sce*I and induction of a DNA DSB on Chr4 that harbors the BFP/GFP LOH reporter system.

Overexpression strains were selected when they had at least a 2-fold change in the LOH rate for each replicate, when compared to the control strain. This way, we identified 6 genes whose overexpression decreased the high LOH rate associated to I-*Sce*I expression. The candidate genes are C5_02440C (*ScBUD32*), C7_02220C (*ScSAT4*) and C6_02160W, that encode serine/threonine kinases; CR_06420W (*PPZ1*), encoding a serine/threonine-specific

protein phosphatase; C6_01190C (*VPS15*) whose gene product is involved in endosome-to-Golgi protein transport; and CR_06110C (*RFX1*) that encodes a transcription factor and has been shown to be involved in DNA damage response and morphogenesis. Up to now, no physical or genetic interactions have been observed between these 6 genes, and it would be interesting to go further in the understanding of the role of these genes in *C. albicans* genome stability. Interestingly, we also identified C1_02120C (*SHA3*; Sc*SKS1*) encoding a putative serine/threonine kinase involved in glucose transport, whose overexpression triggers genome instability for both mono-BFP and GFP cell populations (Table 5 p.123 and Figure 31B).
Furthermore, additional overexpressing strains were identified as having a validated impact on the LOH rate for one cell population only: CR_02120C, C5_03310C (*PCL1*) and C6_01150W (*CDC20*) (Table 6). While *PCL1* and CR_02120C, encoding respectively a cyclin and a C2H2 transcription factor, decreased the number of mono-GFP cells upon their overexpression, *CDC20* and CR_02120C increased the LOH rate towards the mono-BFP cell population.

This screen allowed identifying putative novel players in the genome dynamics of *C. albicans*. Indeed, although preliminary, this study suggests the involvement of transcription factors, phosphatases and kinases that had not been linked to genome stability yet. It should be noted that, despite the high fold-change in the LOH rate upon overexpression, no conclusive statistical analyses could be performed due to the variability between experiments. Thus additional iterations are needed to support these preliminary results. Once statistical support for the candidate genes has been obtained, further studies will be needed to understand their positive or negative role in *C. albicans* genome maintenance. This will involve (i) evaluation of the overexpression level using qPCR; (ii) impact of gene inactivation, if dispensable, on genome maintenance; (iii) impact of overexpression on the transcriptome with the aim to determine possible targets/interactors. In addition, a number of these genes have orthologues in *S. cerevisiae* that display genetic and physical interactions with other genes/gene products. Evaluating the impact of inactivating/overexpressing components of these gene networks on genome maintenance in *S. cerevisiae* will possibly allow extending the observations made in *C. albicans*. Conversely, exploring whether these gene networks are conserved in *C. albicans* might also lead to further knowledge on the mechanisms of genome maintenance in this species. Overall, further exploration of the function of these genes in *C. albicans* and other yeast species is deemed to bring new insights in the understanding of the regulation of genome stability in *C. albicans*.

III. Conclusion and perspectives

Thanks to its high tolerance to genome rearrangements, *C. albicans* is a valuable model to study genome dynamics and more particularly LOH events, which are a hallmark of cancer. My PhD work aimed at developing and validating the use of the I-*Sce*I meganuclease in order to generate a unique DNA DSB in the genome of *C. albicans*. Then, we used this tool to study the frequency at which molecular mechanisms, namely BIR, GC, MCO and chromosome loss or truncation, are used to repair a DNA DSB at one locus on Chr4 as well as the fidelity and extent of BIR and GC with CO events in the context of an I-*Sce*I-dependent DNA DSB. In addition to these investigations, we performed a screen of 564 overexpression strains under conditions favoring or not the formation of an I-*Sce*I-dependent DNA DSB and identified 13 putative new players in genome dynamics in *C. albicans*. The work that we performed using the I-*Sce*I meganuclease allowed discovering that (i) DNA DSB are mainly repaired by GC without CO; (ii) BIR, whole chromosome loss and GC with CO are also occurring upon a DNA DSB, although at a minor rate; (iii) template switching may occur during BIR and GC with CO in *C. albicans*; (iv) BIR and GC with CO generate bidirectional LOH but the underlying mechanisms are still unclear and have never been reported in other organisms; (v) GC with CO is more mutagenic than BIR; (vi) I-*Sce*I-induced DNA DSB can lead to the formation of a partial isochromosome of a 220 kb region of Chr4 left arm; (vii) while C1_08570C (*PCL2*), C1_04380W (*SIT4*), C3_04250W (*WSC1*), C3_03810W (*RAD53*), C2_00350W (*RAD5*), C6_01150W (*CDC20*) and C1_02120C (*SKS1*) have been identified as possibly triggering genome instability, C5_02440C (*BUD32*), C7_02220C, C6_02160W, CR_06420W (*PPZ1*), C6_01190C (*VPS15*) and CR_06110C (*RFX1*) were found to play the role of putative guardians of genome integrity in *C. albicans*; and (viii) recessive deleterious or lethal alleles are recurrent and can influence the outcome of a LOH within a cell population. Experiments are ongoing to confirm the candidate genes involved in genome stability, understand the mechanisms underlying the bidirectionality of homozygosity/heterozygosity observed upon BIR and GC with CO and the origin of the partial i4(L) isochromosome.

Additional uses of the I-*Sce*I meganuclease in *C. albicans* are envisioned to pursue and extend this work, and will be discussed in the following sections.

Homologous recombination-mediated repair mechanisms in C. albicans

The work that has been presented in this manuscript on genome stability and repair mechanisms in *C. albicans* allows confirming that the molecular mechanisms used to repair or respond to a DNA DSB are conserved within eukaryotes. However, NHEJ has never been demonstrated in *C. albicans*, leaving homologous recombination (HR)-dependent pathways as the unique repair mechanism available. In *C. albicans*, LOH events resulting from HR repair mechanisms are recurrent. Hence, we can expect that, as they are often used, these repair mechanisms might be highly efficient and conservative. Nevertheless, as seen in section II.A.3 p.105, HR mechanisms and more precisely, BIR and GC with CO are highly mutagenic in *C. albicans*. Thus, the fidelity of the *C. albicans* HR-mediated repair mechanisms is not higher than in *S. cerevisiae*, suggesting that low fidelity repair processes might participate to intra-species variability.

As discussed in the general introduction of this thesis, Åström *et al.* described that in *S. cerevisiae*, the heterozygous state of the *MAT* locus represses NHEJ by the formation of the Mat**a**1/α2 repressor, allowing the cells to adapt the repair pathway to the availability of the template (184). The main proteins involved in NHEJ are conserved between *S. cerevisiae* and *C. albicans*, thus, in *C. albicans*, the NHEJ pathway might be inhibited by the heterozygosity at the mating type locus. Furthermore, up to now, the attempts to demonstrate the existence of NHEJ were performed in a diploid strain (126, 171, 190). However, the detection of rare events is difficult to achieve in *C. albicans* due to its diploid state and the lack of circular DNA or minichromosome to perform such experiment. Therefore, we propose to determine the presence or absence of NHEJ in *C. albicans* haploid strains. To this aim, the I-*Sce*I system could also be used to assess the presence of the end-joining pathways in haploid isolates. In this respect, we have started a collaboration with Judith Berman (Tel Aviv University, Israel), Yue Wang and Guisheng Zeng (Institute of Molecular and Cell Biology, Singapore) to obtain unpublished relatively stable haploid strains. We are planning to place the I-*Sce*I system in the haploid strain, GZY823#6, which is auxotrophic for uracil, arginine and lysine and nourseothricin (*SAT1*) sensitive. Thus, the I-*Sce*I meganuclease under the control of the tetracycline inducible promoter, and coupled with its transactivator and the *ARG4* marker will be integrated on Chr1 at the *RPS1* locus, the BFP gene associated to the nourseothricin resistance gene will be integrated at the *PGA59-62* locus on Chr4, and a construction will be made to position two I-*Sce*I recognition sequences on both side of the *URA3* marker, a technique equivalent to what has been described in (319, 479, 480) and integrated at the *CDR3-tG(GCC)2* locus on Chr4 (Figure 32).

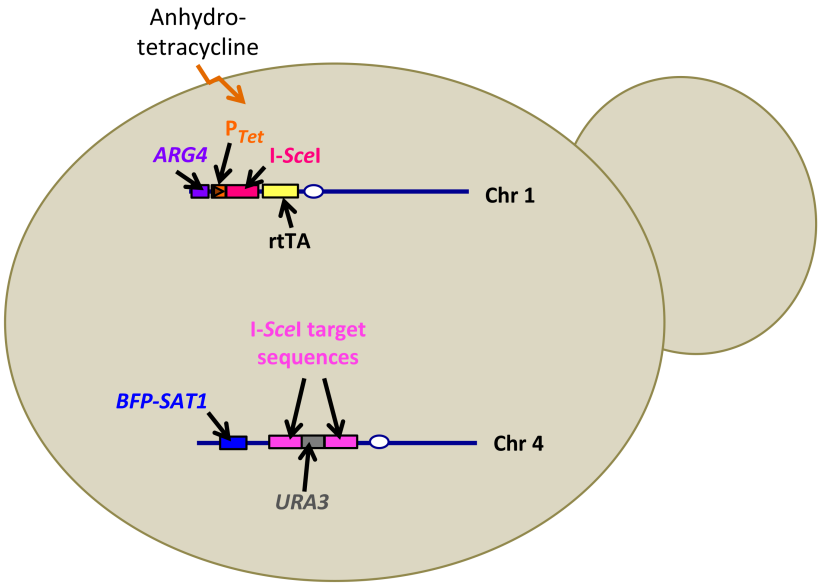


Figure 32: The DNA DSB-inducing system to demonstrate NHEJ pathway in *C. albicans*
This construction relies on: (i) the gene encoding the rare-cutting endonuclease, I-*Sce*I, placed under the control of the tetracycline inducible promoter (P_{TET}) integrated at the *XOG1-HOL1* locus on Chr1, along with the gene encoding the tetracycline-dependent *rtTA* transactivator of the P_{TET} promoter and (iii) two I-*Sce*I target sequence surrounding a *URA3* marker integrated at the *CDR3-tG(GCC)2* locus on Chr4, between the centromere and the BFP encoding gene integrated at the *PGA59-PGA62* locus. Upon DNA DSB and if end-joining pathways exist in *C. albicans*, the cells would repair using NHEJ and give rise to blue URA- progenies.

Thus, upon I-*Sce*I induction, we will search for rare cells that become auxotroph for uracil by 5-FOA counter-selection. Although we can expect a high number of dead cells as well as numerous haploid cells that would have undergone autodiploidization (easily distinguished from haploid cells by DNA content measurement), we will hopefully identify cells that have undergone repair by NHEJ.

Signaling pathways and the control of genome stability in C. albicans

The significant influence of environments or stress conditions on genome instability via signaling pathways have been suggested by the work of Forche *et al.*, showing that either azole treatment, heat shock or oxidative stress resulted in different LOH outcomes (18). Thanks to the original use of overexpression of genes that are involved in DNA repair, recombination and replication as well as signaling, our screen led to the identification of 13 genes playing a role in the maintenance or imperilment of genome integrity. In these screens, both orthologs of human genes and fungi-specific genes have been identified, underlying the occurrence of a conserved process for the maintenance of genome integrity and the possible adaptation of such mechanisms to species living in highly changing environments. Despite the

fact that the genes that we identified in these screens did not reveal specific signaling pathway(s) that could be involved in the regulation of genome stability in *C. albicans*, we evidenced several kinases, phosphatases and transcription factors that physically or genetically interact with other proteins/genes. If these genes do not belong to the present collection, it would be interesting to evaluate the effect of their overexpression on genome dynamics and thus, decipher signaling pathways involved in the control of genome rearrangements in *C. albicans*.

The identification of additional genes that have no ortholog in humans and that are involved in the positive or negative regulation of genome stability would help detecting signaling pathways involved in the control of genomic rearrangements in *C. albicans*. Thus, it would be possible to use this knowledge to identify new target genes, whose deregulation mediated by the use of drugs, would be unfavorable to *C. albicans* during infection.

Recessive lethal alleles and recessive deleterious alleles in C. albicans

As largely discussed in section II.A.2 p.74, we reported the presence of both recessive deleterious and lethal alleles in the genome of *C. albicans*. In this species, the genome is known to be heterozygous with 1 SNP every 217 bp (12), and the presence of recessive deleterious alleles is not surprising regarding the mode of reproduction of this species (*i.e.* clonal reproduction). However, as seen above, both LOH and heterozygosity play important roles in the biology of *C. albicans*. Indeed, LOH events have been often linked to better adaptation to new environments and notably, antifungal resistance (281, 292-294, 299) while the unmasking of deleterious alleles is detrimental to the cells, as shown in this study (Feri *et al., accepted for publication*, mBio). This is also supported by the fact that haploid and autodiploid strains are avirulent and exhibit growth defect (10). Additionally, from one *C. albicans* strain to another, heterozygosity is persistent but occurring at different positions in the genome. There are at least two ways to interpret this versatility. First, while heterozygosity generates intra-species variations, LOH events would allow cells to purge the genome from recessive mutations being disadvantageous when found in the homozygous state or to select for mutations that are advantageous to the cells within the host. Another way to interpret this necessity for both homozygosity and heterozygosity is supported by the fact that *C. albicans* is exposed to recurrent environmental changes inducing stresses that can only be circumvented by genome rearrangements, and more particularly, LOH events that are at the origin of genomic and phenotypic variability.

To have a better knowledge on the essential genes carried by *C. albicans* and to better characterize the SC5314 reference strain we are working with, I-*Sce*I meganuclease can be used to uncover the other recessive lethal alleles that are thought to be present on Chr1, 3, 6 and 7 (10, 74, 126). The identification of the recessive lethal allele(s) on Chr7 responsible for unidirectional LOH events is being undertaken in our laboratory by Timéa Marton (Unité Biologie et Pathogénicité Fongiques, Institut Pasteur) and points to *MTR4*, an RNA helicase whose deletion leads to inviable progeny in *S. cerevisiae*, as a potential essential gene in *C. albicans*. This gene is found in the heterozygous state in the genome of SC5314, with one recessive and possibly lethal allele on Chr7B that carries a premature STOP codon. Investigating the presence of one or several recessive possibly lethal alleles on the remaining three chromosomes would be of interest to enlarge our understanding of the *C. albicans* biology that we have.

*Additional applications of I-*SceI*: the study of centromere biology in* C. albicans

Kaustuv Sanyal, Lakshmi Sreekumar, Neha Varshney and Priya Jaitly (Jawaharlal Nehru Centre for Advanced Scientific Research, India) are working on centromere and neocentromere formation in *C. albicans*. Relying on their recent findings showing the recruitment of CENP-A (Cse4 in *C. albicans*), Rad51 and Rad52 at the neoformed centromere loci, Sanyal and collaborators suggested that the collision between the closest replication fork to the centromere and the kinetochore, during early replication of the centromere, yielded single-stranded DNA and was at the origin of centromeres upon recruitment of Rad51, Rad52 and Cse4 ((37), see Section I.A.1.i p.10). Thus, we initiated a collaboration with the Sanyal group to use the I-*Sce*I DSB-inducing system to prove or disprove the presence of single stranded DNA during neocentromere formation within a 20 kb region in the vicinity of the native centromere. To do so, we have constructed strains carrying the I-*Sce*I meganuclease under the control of the tetracycline-inducible promoter and placed the I-*Sce*I recognition sequence coupled with the *URA3* marker on both sides of the native centromere. This work is in process, and up today, our collaborators could detect an enrichment of Cse4 on and around the I-*Sce*I target site following I-*Sce*I DNA DSB induction. However, Cse4p was also detected in a non-induced condition at the target site. The construction of control strains lacking the I-*Sce*I meganuclease encoding gene is being done. This control strain will help to determine if the unexpected detection of Cse4 at the I-*Sce*I recognition site in absence of induction is due to (i) a leakage of the *TET*-promoter or (ii) the integration of the *URA3*-I-*Sce*I target site close to the centromere that would somehow undergo the recruitment of

centromere-specific proteins due to position effect. This work will hopefully bring new insights in the understanding of centromere biology in *C. albicans*.

Isochromosomes: potential applications

As mentioned in section I.A.1.i p.10, all genetic manipulations in *C. albicans* are performed by integration of DNA into the genome. Indeed, the construction of an artificial chromosome carrying a 3 kb centromeric region resulted in mitotic instability (32). Although unstable, replicative plasmids in *C. albicans* have been used and reported as not suitable for genetic studies (397, 481, 482).

The need for integration is a limitation for the potential studies that can be made in *C. albicans*. Indeed, DNA integration necessitates (i) a highly transcribed locus; (ii) a locus whose disruption by DNA integration would not alter the cell biology; (iii) a locus that is not influenced by heterochromatin (i.e. position effect observed close to the centromere or telomere); (iv) as many selection markers as the number of genetic manipulations needed, limiting the number of genetic manipulations; (v) efficient homologous recombination can issue when investigating processes such as homologous recombination; (vi) as many heat shocks or electroporations as the number of integration, thus increasing the risk for an accumulation of undesired genomic rearrangements. Although this technique is sufficiently efficient to allow the study of genes of interest, the development of a new tool such as a minichromosome (483-486) would be of great use in the *C. albicans* community.

Regarding the recent discovery of a potentially stable isochromosome of a portion of Chr4 left arm, we could isolate this isochromosome and try to reduce its size as much as possible to be able to use it as a minichromosome as it has been done in other species (483-486), avoiding the use of integrative transformations.

Altogether, this PhD work contributed to a better understanding of genome stability in *C. albicans*, notably thanks to the development of the I-*Sce*I-mediated DNA DSB inducing system coupled with the BFP/GFP system. These systems – combined or used separately – can be applied to further investigate not only genome stability but also *C. albicans* biology as a whole. In sum, this work emphasizes how intriguing this species is and paves the way for further genomic studies in this species.

Supplemental material

Table S1 – Strains used in this study

Strain Name	Genotype	Auxotrophies	Parental strain	References
SN148	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434*	arg- his- leu- uri-	-	(Noble and Johnson 2005)
CEC2684	SN148 *Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1*	ARG+ HIS+ leu- ura-	SN148	(Loll-Krippleber *et al*. 2014)
CEC3867	SN148 *Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1*	ARG+ HIS+ leu- uri- NsnR	CEC2684	This study
CEC3888	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-XOG1HOL1/ Ca21chr1_C_albicans_SC5314:625304 to 626436*	ARG+ HIS+ leu- uri- NsnR HygR	CEC3867	This study

CEC3884	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-ISceI/Ca21chr1_C_albicans_SC5314:625304 to 626436*	ARG+ HIS+ leu- uri- NsnR HygR	CEC3867	This study
CEC4021	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-NLS-ISceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436*	ARG+ HIS+ leu- uri- NsnR HygR	CEC3867	This study
CEC3927	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-ISceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::URA3-ISceI_TS/Ca21chr4_C_albicans_SC5314:775939 to 779223*	ARG+ HIS+ leu- URI+ NsnR HygR	CEC3884	This study

CEC3932	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-ISceI/Ca21chr1_C_albicans_SC5314:625304 to 626436 RPS1/rps1::CIp10*	ARG+ HIS+ leu- URI+ NsnR HygR	CEC3884	This study
CEC4012	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-ATG-NLS-ISceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::URA3-ISceI_TS/Ca21chr4_C_albicans_SC5314:775939 to 779223*	ARG+ HIS+ leu- URI+ NsnR HygR	CEC4021	This study
CEC4088	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-ATG-NLS-ISceI/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::URA3-ISceI_TS/Ca21chr4_C_albicans_SC5314:775939 to 779223*	ARG+ HIS+ leu- URI+ NsnR HygR	CEC4021	This study

CEC4045	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-ATG-NLS-ISceI/Ca21chr1_C_albicans_SC5314:625304 to 626436 RPS1/rps1::*CIp10	ARG+ HIS+ leu- URI+ NsnR HygR	CEC4021	This study
CEC3930	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:625304 to 626436Δ::CIpXH-HYGb-PTET-XOG1HOL1/ Ca21chr1_C_albicans_SC5314:625304 to 626436 Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::URA3-ISceI_TS/Ca21chr4_C_albicans_SC5314:775939 to 779223*	ARG+ HIS+ leu- URI+ NsnR HygR	CEC3888	This study
CEC4429	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1/Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::URA3-ISceI_TS/Ca21chr4_C_albicans_SC5314:775939 to 779223 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:125966 to 131098Δ::CIpXH-HYGb-PTET-ATG-NLSISceI/Ca21chr1_C_albicans_ SC5314:625304 to 626436 RPS1/rps1::PTDH3-C4_03130W_A-CdLEU2*	Prototroph NsnR HygR	CEC4012	This study

CEC4430	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1/Ca21chr4_C_albicans_SC5314:775939 to 779223Δ::URA3-ISceI_TS/Ca21chr4_C_albicans_SC5314:775939 to 779223 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:125966 to 131098Δ::CIpXH-HYGb-PTET-ATG-NLS-ISceI/Ca21chr1_C_albicans_SC5314:625304 to 626436 RPS1/rps1::PTDH3-C4_03130W_A-CdLEU2*	Prototroph NsnR HygR	CEC4088	This study
CEC4797	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:125966 to 131098Δ::CIpXH-HYGb-PTET-ATG-NLS-ISceI/Ca21chr1_C_albicans_SC5314:625304 to 626436 rps1::PTDH3-C4_03130W_A-CdLEU2/rps1::PTDH3-C4_03750C_A-CaURA3*	Prototroph NsnR HygR	CEC4430	This study
CEC4798	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:125966 to 131098Δ::CIpXH-HYGb-PTET-ATG-NLS-ISceI/Ca21chr1_C_albicans_SC5314:625304 to 626436 rps1::PTDH3-C4_03130W_A-CdLEU2/rps1::PMRF2-C4_03750C_A-CaURA3*	Prototroph NsnR HygR	CEC4430	This study

CEC4817	*arg4Δ/arg4Δ leu2Δ/leu2Δ his1Δ/his1Δ ura3::λimm434/ura3Δ::λimm434 iro1Δ::λimm434/iro1Δ::λimm434 Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-GFP-CaARG4/Ca21chr4_C_albicans_SC5314:473390 to 476401Δ::PTDH3-BFP-CdHIS1 ADH1/adh1::PTDH3-carTA::SAT1 Ca21chr1_C_albicans_SC5314:125966 to 131098Δ::CIpXH-HYGb-PTET-ATG-NLS-ISceI/Ca21chr1_C_albicans_SC5314:625304 to 626436 rps1::PTDH3-C4_03130W_A-CdLEU2/rps1::*CIp10	Prototroph NsnR HygR	CEC4430	This study
CEC2876	CLINICAL ISOLATE	Prototroph	-	(M.E.B., G.S., N.S., K.S., C.M., and C.d'E., manuscript *in preparation*)
CEC3673	CLINICAL ISOLATE	Prototroph	-	(M.E.B., G.S., N.S., K.S., C.M., and C.d'E., manuscript *in preparation*)
SC5314	Reference strain	Prototroph	-	(Fonzi and Irwin 1993)

1. Noble SM, Johnson AD: Strains and strategies for large-scale gene deletion studies of the diploid human fungal pathogen *Candida albicans*. Eukaryot Cell 2005, 4:298-309.
2. Loll-Krippleber R *et al*.: 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 2014, 91:452-471.
3. Fonzi WA, Irwin MY: Isogenic strain construction and gene mapping in *Candida albicans*. Genetics 1993, 134:717-728.

Table S2 – Primers used in this study

	Primer names	Sequence
PCR amplification for cloning	AF001	gggaaagccggcGACAATCACGAAGCCAAGTAAG
	AF002	aaaggggccggcTTACCGCTCGGCTTTGTTCC
	AF006	gggaaagttaacGAAGGAAGATGAATTGGGTTGC
	AF007	aaagggccgcggGGGAGTCTCTCGTGAAGAATATG
	AF008	gggaaactgcagattaccctgttatccctaGGGTGGTCATACTATGTGTTGTTG
	AF009	aaagggaagcttAGGAGACTACAAACTGGAAGG
	AF027	cccaaagatatcATGccaccaaaaaaaaaaagaaaagttcataaaaatattaaaaaaaatcaagtt
	AF031	aaagggatgcatttattttaaaaaagtttctgatgaaatagt
	AF049	AAACCCgatatcATGCATTTGTTACTAAGGATACC
	AF050	CCCAAAacgcgtgccggcCTATTGTGATCTATAAACATCGATC
	AF113	CTCGAGACCGGTggagagttcttaatacttga
	AF114	CAGCTGgtggtaaggtgctaatctaa
PCR MTLa/MTLα	AF120	TTGAAGCGTGAGAGGCAGGAG
	AF121	GTTTGGGTTCCTTCTTTCTCATTC
	AF122	TTCGAGTACATTCTGGTCGCG
	AF123	TGTAAACATCCTCAATTGTACCCGA
PCR amplification to check the integration	XOGHOL-verif F	GCTGTCATCTACTGGTTTGTG
	XOGHOL-verif R	GTGCGTTGGAACACCAGTTG
	Verif-CDR3-F	TGGTGGGTAAAGGGCATATTC
	Verif-TG(GCC)2-R	GTTGGTATCATGGTGATGTGTC
	CIpUR	ATTACTATTTACAATCAAAGGTGGTC
	CIpUL	ATACTACTGAAAATTTCCTGACTTTC
	CIpUL_2	AGATACTCACGCACGCCCATACTACT
	Leu2_down	GCTACTGAAGTTGGTGACGCGATTGT

	ADH1verif	ACAAGCTCATTGAGTGACGAAAAG
	PNIM1verif	tttacgggttgttaaaccttcgat
	URA3_rev_AF	gttgtcctaatccatcacct
	URA3_fwd_AF	aactcatgcctcaccagtag
SNP-RFLP	SNP-95-F#2	CATGCCCGCTTGAAACTACC
	SNP-95-R#2	GTCAGTGATTCAGTTGAAGTGG
	SNP 156-R	TGGGTTTGGACATCAGGTTCAA
	SNP-156-F#2	ACAGAACAGTAGATTCCAAC
	SNP-23-F	AGCCAACCATATTTCAGGATTGAC
	SNP-23-R	GTGCCAACTAGTAATGGTTGTCAT
	SNP-42-F	GTACTTCTATACACGCACATCTTCA
	SNP-42-R	GAAATCCACCGCATAAGAAATGGTT

Text S1 – Supplementary Material and Methods

Plasmid constructions

p*Tet*-I-*Sce*I-*HYGb-XOG1HOL1* harbors the I-*Sce*I gene under the control of a tetracycline inducible promoter (402), the hygromycin B resistance gene (487) and a sequence that allows its integration at the *XOG1-HOL1* intergenic region on chromosome 1 (Chr1). To generate this plasmid, CIpOp2, a CIp10 derivative containing the *Tet*-on promoter (402) was digested with *Eco*RV to excise the Gateway® Cassette and allowed to religate (p*Tet*-w/oGTW). The resulting plasmid was digested with *Not*I to excise the *URA3* marker and the *RPS1* sequences and recircularized, yielding p*Tet*-w/oGTW-w/o*URA3*. The *HYGb* gene was excised from pKS-*ACT1*p-*HYGb* (487) by *Xba*I and *Acc*65I, blunt-ended with the Klenow enzyme (Invitrogen™) and cloned into the *Ale*I-linearized p*Tet*-w/oGTW-w/o*URA3* plasmid, yielding p*Tet*-w/oGTW-*HYGb*. Then, a 1kb sequence from the intergenic region between *XOG1* and *HOL1* genes was amplified by PCR with oligonucleotides AF1 and AF2 (Table S2). The PCR product was digested with *Nae*I and ligated into the *Nae*I-linearized p*Tet*-w/oGTW-*HYGb* plasmid yielding p*Tet*-w/oGTW-*HYGb-XOG1HOL1*. Finally, as *C. albicans* is part of the CTG clade (488), we used http://genomes.urv.es/OPTIMIZER/ website (489) to optimize the I-*Sce*I meganuclease gene sequence to the *C. albicans* codon usage . The optimized I-*Sce*I gene flanked by *Nsi*I and *Eco*RV restriction sites was synthesized by GeneArt® Invitrogen™, and cloned into the p*Tet*-w/oGTW-*HYGb-XOG1HOL1* plasmid linearized with *Nsi*I and *Eco*RV to yield p*Tet*-I-*Sce*I-*HYGb-XOG1HOL1*.

p*Tet*-NLS-I-*Sce*I-*HYGb-XOG1HOL1* was constructed by using the Nuclear Localization Sequence (NLS) from SV40 (490) carrying a region recognized by Importin α (491) and which has already been used in *C. albicans* (492). We used the http://genomes.urv.es/OPTIMIZER/ website to optimize the codon usage for *C. albicans* as follow: ATG CCA CCA AAA AAA AAA AGA AAA GTT CAT. We fused this modified NLS sequence to I-*Sce*I by PCR using primers AF27 and AF31 (Table S2), and p*Tet*-I-*Sce*I-*HYGb-XOG1HOL1* as a template. The PCR fragment was cloned into a TOPO®-TA vector (Thermofisher), digested with *Eco*RV and *Nsi*I and cloned into p*Tet*-I-*Sce*I-*HYGb-XOG1HOL1* digested with the same enzymes, yielding p*Tet*-NLS-I-*Sce*I-*HYGb-XOG1HOL1*.

pFA-I-*Sce*I-TS-*URA3-CDR3/tG(GCC)2* carries the I-*Sce*I recognition sequence, the *URA3* marker and two sequences that allow integration in the *CDR3/tG(GCC)*$_2$ intergenic region on Chr4. To obtain this plasmid, we first amplified two sequences in the intergenic region between *CDR3* and *tG(GCC)2* genes. The first region at position 775,906-776,737, closer to *CDR3*, is

831bp long and was amplified by PCR with oligonucleotides AF6 and AF7 (Table S2). The second region at position 778,136-779,189, closer to *tG(GCC)*$_2$, is 1,055bp long and was amplified by PCR with oligonucleotides AF8 and AF9, containing the I-*Sce*I target sequence (attaccctgttatcccta, Table S2). The *Sac*II+*Hpa*I-digested AF6/AF7 PCR product was inserted at the *Sac*II and *Hpa*I sites of pFA-*CaURA3* (449), carrying the *URA3* gene. The resulting plasmid was then double-digested with *Hind*III-*Pst*I to insert the *Hind*III+*Pst*I-cut AF8/AF9 PCR product. The resulting plasmid was named pFA-I-*Sce*I-TS-*URA3-CDR3/tG(GCC)2*.

The p*CaTDH3*-GTW-*LEU2* plasmid carries the sequence for integration at the *RPS1* locus on Chr1, the *C. maltosa* optimized *LEU2* marker and a Gateway® Cassette flanked by the *attR* sequences and under the control of the P_{TDH3} constitutive promoter. To obtain this plasmid, the vector pFA-*CmLEU2* (473) was digested with *Pvu*II and *Spe*I, the band corresponding to the *LEU2* marker was purified and cloned into the *Ale*I-*Spe*I double-digested p*CaTDH3*-GTW-*URA3* plasmid (19). The resulting plasmid was called p*CaTDH3*-GTW-*LEU2*.

p*CaTDH3-GPI16-LEU2* was constructed by using the Gateway® Technology as described in Chauvel *et al*. (402) by LR recombination between the BP vector containing orf19.2677 (*GPI16*) and the destination vector pCa*TDH3*-GTW-*LEU2*.

p*CaTDH3-MRF2-URA3* was constructed by cloning orf19.1303 (*MRF2*) under the control of a constitutive promoter in p*CaTDH3*-GTW-*URA3*, a CIpOp2 derivative plasmid (19). This plasmid was digested with *Bsr*GI to excise the Gateway® Cassette and allowed to religate, resulting in p*CaTDH3-URA3*. orf19.1303 was amplified by PCR using primers AF49 and AF50 (Table S2) containing *Eco*RV and *Mlu*I-*Nae*I restriction sites at the 3' and 5' ends, respectively. The PCR amplified DNA was gel purified and cloned into a TOPO®-TA vector. Sequencing was performed to select the TOPO-TA-*MRF2* clone carrying the full-length allele. The *MRF2* gene was excised from the TOPO-TA-*MRF2* vector by *Eco*RV-*Mlu*I digestion and cloned into the p*CaTDH3-URA3* plasmid linearized with *Eco*RV-*Mlu*I, yielding p*CaTDH3-MRF2-URA3*.

p*CaMRF2-MRF2-URA3* was obtained by replacing the P_{TDH3} promoter from p*CaTDH3-MRF2-URA3* by the endogenous *MRF2* promoter. The promoter of *MRF2* was PCR amplified with AF113 and AF114 (Table S2) containing *Eco*RV and *Mlu*I-*Nae*I restriction sites at the 3' and 5' ends. The PCR product was gel purified and cloned into a TOPO®-TA vector. The *MRF2* promoter was excised from the TOPO®-TA-p*CaMRF2* vector by *Xho*I, *Pvu*II and *Dra*III digestion and cloned into the *Xho*I-*Eco*RV double-digested p*CaTDH3-MRF2-URA3* plasmid to replace the P_{TDH3} promoter yielding p*CaMRF2-MRF2-URA3*.

Strain constructions

CEC2684 was first transformed with *Sac*II-*Kpn*I-digested pNIMX (402) to introduce the transactivator for the *Tet*-on promoter (402). pNIMX (402) is a pNIM1 derivative (393) and integrates at the *ADH1* locus on Chr5. Transformants were selected on nourseothricin-containing media (300µg/mL) (493) and validated by PCR with the ADH1verif and NIM1verif primers (Table S2), giving rise to CEC3867. CEC3867 was then transformed with the *Pml*I-linearized p*Tet*-NLS-*I-SceI-HYGb-XOG1HOL1* or p*Tet*-w/oGTW-*HYGb-XOG1HOL1* plasmids, which integrate at the *XOG1-HOL1* intergenic region of Chr1. After selection on hygromycin B-containing plates (700µg/mL) (487), we obtained CEC4021 and CEC3888, respectively. Finally, CEC4021 was transformed with *Nco*I-*Hind*III-digested pFA-I-*Sce*I-TS-*URA3-CDR3/tG(GCC)2*, integrating at the *CDR3-tG(GCC)2* intergenic region, on Chr4, located 302,518Kb away from the *PGA59-PGA62* locus and 214,361Kb away from the centromere, yielding the CEC4012 ("I-*Sce*I + TargetB") or CEC4088 ("I-*Sce*I + TargetA") strains. We also constructed a first control strain that carries only the inducible *I-SceI* gene, but no I-*Sce*I recognition sequence. Because of uridine auxotrophy, this control strains was transformed with the *Stu*I-digested CIp10 plasmid (494) that integrates at the *RPS1* locus on Chr1 yielding CEC4045 ("I-*Sce*I only"). The second control strain, carrying only the I-*Sce*I target sequence but no I-*Sce*I gene, CEC3930 ("Target only"), was obtained by transforming CEC3888 with the pFA-I-*Sce*I-TS-*URA3-CDR3/tG(GCC)2* construction.

To obtain the strain overexpressing *GPI16* with or without the *MRF2* gene, we first transformed CEC4012 and CEC4088 with p*CaTDH3-GPI16-LEU2*, yielding CEC4429 and CEC4430 respectively. Subsequently, a small colony derived from CEC4430 having lost the I-*Sce*I target sequence-bearing chromosome was transformed with the p*CaTDH3-MRF2-URA3* or p*CaMRF2-MRF2-LEU2* plasmids, leading to CEC4797 and CEC4798.

Table S3 – Genes that are overexpressed

Plate coordinates	ORF19.xxxx	CX_YYYY
1A01	ORF19.6551.1	C7_01730C
1A02	ORF19.4623.3	C4_01700C
1A03	ORF19.233.1	C3_02450W
1A04	ORF19.7227	C1_14190C
1A05	ORF19.5660	C4_00320C
1A06	ORF19.5438	C3_00310C
1A07	ORF19.7644	CR_10610C
1A08	ORF19.953.1	C5_00370W
1A09	ORF19.7250	C1_14350W
1A11	ORF19.3294	C1_01060W
1A12	ORF19.4185	C4_00580W
1B01	ORF19.4427	C1_07410C
1B02	ORF19.5104	C1_08260C
1B03	ORF19.4009	C5_05160C
1B04	ORF19.2762	C4_02410C
1B05	ORF19.4796	C1_09350W
1B06	ORF19.861	C2_03590C
1B07	ORF19.3088	C4_07150W
1B08	ORF19.2363	CR_07020W
1B09	ORF19.4443	C1_07240W
1B10	mtlalpha2	C5_01785W
1B11	ORF19.2204.2	C2_07750W
1B12	ORF19.390	C1_08450C
1C01	mtlalpha1	C5_01755C
1C02	ORF19.7417	C3_06180C
1C03	ORF19.1227	C1_07670W
1C04	ORF19.3201	C5_01740C
1C05	ORF19.2843	CR_02860W
1C06	ORF19.5156	C7_03050W
1C07	ORF19.1178	C6_00180C
1C08	ORF19.3200	C5_01750C
1C09	ORF19.7206	C1_13990W
1C10	ORF19.5334	C2_10560C
1C11	ORF19.5157	C7_03040W
1C12	ORF19.927	C5_00620W
1D01	ORF19.5968	C3_05000W
1D02	ORF19.3534	C2_05030C
1D03	ORF19.2105	C2_00280C
1D04	ORF19.4086	C2_09250W

1D05	ORF19.2876	C4_06580W
1D06	ORF19.6645	CR_05670C
1D07	ORF19.3001	C1_03080C
1D08	ORF19.410.3	C1_05560W
1D09	ORF19.2393	CR_03310C
1D10	ORF19.798	C2_04220C
1D11	ORF19.1815	CR_07080W
1D12	ORF19.4767	C1_09100W
1E02	ORF19.2612	CR_02120C
1E03	ORF19.259	C3_02670W
1E04	ORF19.5501	C7_03770C
1E05	ORF19.403	C1_08570C
1E06	ORF19.5338	C2_10590W
1E07	ORF19.7025	C7_00890C
1E08	ORF19.3014	C1_03220C
1E09	ORF19.1769	C2_10120W
1E10	ORF19.4297	C5_02760W
1E11	ORF19.5326	C2_10540W
1E12	ORF19.4941	C1_13140C
1F01	ORF19.1762	C2_10190C
1F02	ORF19.4722	C1_08640W
1F03	ORF19.3957	C5_04720C
1F04	ORF19.1850	CR_06830C
1F05	ORF19.6102	C1_00080C
1F06	ORF19.5956	C3_04880W
1F07	ORF19.5975	C3_05050W
1F08	ORF19.5350	C2_10700C
1F09	ORF19.1760	C2_10210C
1F10	ORF19.2102	C2_00300C
1F11	ORF19.4252	C5_02440C
1F12	ORF19.878	C2_03430W
1G01	ORF19.6901	C7_01190W
1G02	ORF19.4378	CR_03800C
1G03	ORF19.7504	CR_00290W
1G04	ORF19.7150	C7_04230W
1G05	ORF19.5200	C1_04380W
1G06	ORF19.3149	C2_06730W
1G07	ORF19.3856	CR_06050W
1G08	ORF19.1341	C7_03340C
1G09	ORF19.4401	CR_03570C
1G10	ORF19.1358	C2_09940W
1G11	ORF19.3530	C2_04980C
1G12	ORF19.6227	C1_06820W
1H01	ORF19.6285	CR_07650W
1H03	ORF19.1729	C3_01220W

1H04	ORF19.7017	C7_00970C
1H05	ORF19.7652	CR_10660W
1H06	ORF19.793	C2_04270W
1H07	ORF19.2674	C4_03160C
1H08	ORF19.6588	C7_01410C
1H09	ORF19.6239	C1_06710W
1H10	ORF19.6514	C7_02040C
1H11	ORF19.6638	CR_05740C
1H12	ORF19.3330	C1_01420C
2A01	ORF19.5661	C4_00340W
2A02	ORF19.2087	C2_00440W
2A03	ORF19.1697	C3_01520C
2A04	ORF19.2432	C1_06130C
2A05	ORF19.6933	C3_03830W
2A06	ORF19.5867	C3_04250W
2A07	ORF19.1683	C3_01600W
2A08	ORF19.2260	C2_07040W
2A09	ORF19.7547	CR_09750C
2A10	ORF19.7164	C7_04110W
2A11	ORF19.124	C6_01170W
2A12	ORF19.6305	CR_04880W
2B01	ORF19.1007	CR_05260C
2B03	ORF19.4062	C1_05090W
2B04	ORF19.7282	CR_08870W
2B05	ORF19.460	CR_05940W
2B06	ORF19.6365	CR_08070W
2B07	ORF19.5032	C1_13940W
2B08	ORF19.2649	C5_03310C
2B09	ORF19.7208	C1_14010W
2B10	ORF19.4785	C1_09260C
2B11	ORF19.909	C2_03220C
2B12	ORF19.53	C1_04990C
2C01	ORF19.895	C2_03330C
2C02	ORF19.3407	C6_01770W
2C03	ORF19.2222	C2_08270C
2C04	ORF19.3683	C1_02260C
2C05	ORF19.217	C2_08890W
2C07	ORF19.567	C5_00690C
2C08	ORF19.59	C1_05060W
2C09	ORF19.1803	CR_04910W
2C10	ORF19.3127	C4_06820C
2C11	ORF19.5312	C4_04000W
2D01	ORF19.1028	C1_03810C
2D03	ORF19.2961	C1_02720W

2D04	ORF19.3193	C5_01810W
2D05	ORF19.3083	C4_07200C
2D07	ORF19.1985	CR_07700W
2D09	ORF19.3018	C1_03250C
2D11	ORF19.6492	C7_02220C
2D12	ORF19.791	C2_04290W
2E01	ORF19.4125	C2_04860W
2E02	ORF19.4084	C2_09230C
2E03	ORF19.5911	C3_04550C
2E04	ORF19.3774	C4_05050C
2E06	ORF19.3642	C6_00820W
2E07	ORF19.2886	C4_06480C
2E08	ORF19.6792	C3_07080W
2E09	ORF19.5247	C1_12170C
2E10	ORF19.5325	C2_10530C
2E11	ORF19.4015	C5_05220W
2F01	ORF19.5537	C6_02670C
2F02	ORF19.3476	C6_02340W
2F03	ORF19.3705	CR_07770C
2F04	ORF19.5917	C3_04580C
2F05	ORF19.1589	C2_02630W
2F06	ORF19.2324	C1_10930C
2F07	ORF19.1220	C6_04040C
2F08	ORF19.5406	C3_00570C
2F09	ORF19.1577	C2_02530W
2F10	ORF19.4127	C2_04850C
2F11	ORF19.2277	C2_07210C
2F12	ORF19.2331	C1_10860C
2G01	ORF19.5953	C3_04860W
2G02	ORF19.1874	C2_07530C
2G03	ORF19.6889	C2_05780C
2G04	ORF19.4056	C1_05140W
2G05	ORF19.5995	C3_05190C
2G06	ORF19.5343	C2_10660W
2G07	ORF19.6589	CR_09730C
2G08	ORF19.6407	CR_08410W
2G09	ORF19.1496	C2_01870C
2G10	ORF19.487	CR_04040C
2G11	ORF19.799	C2_04210W
2G12	ORF19.2014	C2_01110C
2H01	ORF19.2745	C4_02560C
2H02	ORF19.1636	C3_02100W
2H03	ORF19.3506	C6_02050W
2H04	ORF19.1960	C5_01100C
2H05	ORF19.6824	C3_06790W

2H06	ORF19.6369	CR_08100C
2H07	ORF19.1314	C4_03600C
2H08	ORF19.3049	C1_03470C
2H09	ORF19.696	CR_06610W
2H10	ORF19.6781	C3_07200C
2H11	ORF19.1150	C1_11690W
2H12	ORF19.384	C1_08390C
3A01	ORF19.1754	C2_10260C
3A02	ORF19.3281	CR_00750C
3A03	ORF19.6845	C1_04510W
3A04	ORF19.5617	C6_03290W
3A05	ORF19.7317	CR_09200C
3A06	ORF19.1759	C2_10220C
3A07	ORF19.726	CR_06420W
3A08	ORF19.4014	C5_05210W
3A09	ORF19.7186	C7_03940C
3A10	ORF19.7081	C7_00390W
3A11	ORF19.7251	C1_14360C
3A12	ORF19.5734	C6_03650C
3B01	ORF19.1446	C2_01410C
3B02	ORF19.2320	C1_10950C
3B03	ORF19.5184	C7_02780W
3B04	ORF19.1971	C5_00980W
3B05	ORF19.7001	C3_05650W
3B06	ORF19.3233	CR_01210C
3B07	ORF19.1623	C3_02220W
3B08	ORF19.7319	CR_09210W
3B09	ORF19.7385	C3_06040W
3B10	ORF19.1621	C3_02240C
3B11	ORF19.3405	C6_01790C
3B12	ORF19.5531	C6_02610C
3C01	ORF19.4534	C1_01890C
3C02	ORF19.7523	CR_00120C
3C03	ORF19.7281	CR_08860W
3C04	ORF19.6109	C1_00060W
3C05	ORF19.2458	C1_05910W
3C06	ORF19.3178	C5_01930W
3C07	ORF19.684	C6_01920C
3C08	ORF19.3300	C1_01110C
3C09	ORF19.175	CR_02500W
3C10	ORF19.2356	CR_07060C
3C11	ORF19.6817	C3_06850W
3C12	ORF19.4869	C1_10020W
3D01	ORF19.2054	C2_00720C

3D02	ORF19.1468	C2_01600C
3D03	ORF19.2315	C1_10990C
3D04	ORF19.2395	CR_03290C
3D05	ORF19.610	CR_07890W
3D06	ORF19.5389	C3_00670C
3D07	ORF19.2647	C5_03320C
3D08	ORF19.3474	C6_02320C
3D10	ORF19.5257	C1_12080W
3D11	ORF19.4192	C6_00670W
3D12	ORF19.1069	C1_04330W
3E01	ORF19.5722	C6_03550C
3E02	ORF19.505	CR_04190W
3E03	ORF19.7388	C3_06070C
3E04	ORF19.889	C2_03370W
3E05	ORF19.2379	CR_03420C
3E06	ORF19.3456	C6_02160W
3E07	ORF19.971	C5_00240W
3E08	ORF19.2423	CR_03050C
3E09	ORF19.7234	C1_14240W
3E10	ORF19.1673	C3_01710C
3E11	ORF19.6275	C1_06360W
3E12	ORF19.4450	C1_07170C
3F01	ORF19.4318	C5_02940C
3F02	ORF19.5758	C6_03830W
3F03	ORF19.2268	C2_07130C
3F04	ORF19.6850	C1_04560W
3F05	ORF19.5849	CR_05530C
3F06	ORF19.4853	C1_09870W
3F07	ORF19.6376	CR_08160W
3F08	ORF19.2399	CR_03240C
3F09	ORF19.7518	CR_00170W
3F10	ORF19.1757	C2_10230W
3F11	ORF19.6038	C1_00670C
3F12	ORF19.2538	CR_01520W
3G01	ORF19.4961	C1_13350W
3G02	ORF19.6980	C3_05420W
3G03	ORF19.1619	C3_02260C
3G04	ORF19.469	CR_03900W
3G05	ORF19.5548	C6_02750C
3G06	ORF19.3669	C1_02120C
3G07	ORF19.4866	C1_10000C
3G08	ORF19.3425	C6_01620W
3G09	ORF19.2823	CR_02640W
3G10	ORF19.2545	CR_01580C
3G11	ORF19.5676	C4_00470C

3G12	ORF19.6874	C2_05640W
3H01	ORF19.5664	C4_00370W
3H02	ORF19.794	C2_04260W
3H03	ORF19.6033	C1_00730C
3H04	ORF19.4799	C1_09370W
3H05	ORF19.4662	C4_01260W
3H06	ORF19.1565	C2_02420C
3H08	ORF19.7115	C7_00080C
3H09	ORF19.3794	C4_04850C
3H10	ORF19.1936	C5_01320W
3H11	ORF19.3876	CR_06240C
3H12	ORF19.1699	C3_01500C
4A01	ORF19.5848	CR_05540C
4A02	ORF19.1826	C1_06230C
4A03	ORF19.1576	C2_02520W
4A04	ORF19.4166	C4_00760W
4A05	ORF19.5919	C3_04610W
4A06	ORF19.3854	CR_06040W
4A07	ORF19.4645	C4_01410W
4A08	ORF19.719	CR_06470W
4A09	ORF19.681	C1_11210C
4A10	ORF19.4347	C5_03150W
4A11	ORF19.3308	C1_01200W
4A12	ORF19.6810	C3_06910C
4B01	ORF19.4972	C1_13440C
4B02	ORF19.2699	C4_02940W
4B03	ORF19.6010	C1_00950C
4B04	ORF19.5729	C6_03610W
4B05	ORF19.4433	C1_07370C
4B06	ORF19.3561	C2_05320W
4B07	ORF19.1253	C4_05680W
4B08	ORF19.7247	C1_14340C
4B09	ORF19.2528	CR_01420W
4B10	ORF19.795	C2_04250W
4B11	ORF19.4979	C1_13470W
4B12	ORF19.1275	C4_05880W
4C01	ORF19.5031	C1_13930W
4C02	ORF19.5429	C3_00380C
4C03	ORF19.4222	C5_02160W
4C04	ORF19.7583	CR_10070C
4C05	ORF19.976	C5_00210C
4C06	ORF19.3912	C5_04280C
4C07	ORF19.6915	C7_01340W
4C08	ORF19.4000	C5_05080W

4C09	ORF19.921	C5_00670C
4C10	ORF19.3077	C4_07230C
4C11	ORF19.6506	C7_02110W
4C12	ORF19.4778	C1_09200W
4D01	ORF19.2605	CR_02040W
4D03	ORF19.718	CR_06480C
4D04	ORF19.1949	C5_01210W
4D05	ORF19.3207	C5_01680C
4D06	ORF19.5408	C3_00550C
4D07	ORF19.2447	C1_05980W
4D08	ORF19.6936	C3_03810W
4D09	ORF19.3302	C1_01140C
4D11	ORF19.3449	C6_01350W
4D12	ORF19.6182	C3_07860C
4E01	ORF19.2842	CR_02850C
4E02	ORF19.399	C1_08540C
4E03	ORF19.3969	C5_04830W
4E04	ORF19.5940	C3_04770C
4E05	ORF19.5498	C2_06340W
4E06	ORF19.5651	C4_00260W
4E07	ORF19.428	C1_05370C
4E08	ORF19.1223	C2_06670C
4E09	ORF19.843	C2_03780C
4E10	ORF19.4725	C1_08670W
4E11	ORF19.7359	C3_05780C
4E12	ORF19.4909	C1_10380C
4F01	ORF19.5253	C1_12120W
4F02	ORF19.5357	C2_10750C
4F03	ORF19.3153	C3_01160W
4F04	ORF19.3305	C1_01170C
4F05	ORF19.3809	C4_04730W
4F06	ORF19.723	CR_06440C
4F07	ORF19.5908	C3_04530C
4F08	ORF19.3523	C2_04930C
4F10	ORF19.1674	C3_01700W
4F11	ORF19.5045	C4_03890W
4F12	ORF19.4752	C1_08940C
4G01	ORF19.5910	C3_04540C
4G02	ORF19.4775	C1_09170W
4G03	ORF19.7105	C7_00180W
4G06	ORF19.6124	CR_07440W
4G07	ORF19.6028	C1_00780C
4G08	ORF19.6680	C5_03680W
4G09	ORF19.829	C2_03940C
4G11	ORF19.536	CR_04450C

4G12	ORF19.7570	CR_09960C
4H01	PIKalpha	C5_01765C
4H02	ORF19.4568	C4_02220C
4H03	ORF19.2540	CR_01540W
4H04	ORF19.454	CR_05990C
4H05	ORF19.147	C2_04660C
4H06	ORF19.1135	C1_11810W
4H07	ORF19.223	C2_08860W
4H08	ORF19.5376	C3_00790W
4H09	ORF19.3415	C6_01710C
4H10	ORF19.4670	C4_01190W
4H11	ORF19.1944	C5_01250W
4H12	ORF19.844	C2_03770C
5A05	orf19.3841	C4_04450C
5A06	orf19.255	C3_02640C
5A07	orf19.2781	C1_07640C
5A08	orf19.1822	C1_06280C
5A09	orf19.3174	C5_01970C
5A10	orf19.20	C2_06420C
5A11	orf19.5855	C3_04110C
5A12	orf19.2969	C1_02810W
5B01	orf19.1187	C6_00280W
5B02	orf19.3986	C5_04970C
5B04	orf19.5380	C3_00750W
5B05	orf19.1196	C6_00350W
5B06	orf19.5160	C7_03010W
5B08	orf19.6987	C3_05520C
5B09	orf19.4776	C1_09180W
5B10	orf19.1901	C2_07350W
5B11	orf19.3252	CR_01030W
5B12	orf19.1217	C6_04060W
5C02	orf19.801	C2_04200W
5C03	orf19.191	C2_04760W
5C04	orf19.3071	C3_00800W
5C06	orf19.2678	C4_03120C
5C08	orf19.1718	C3_01330W
5C09	orf19.2280	C2_07230C
5C10	orf19.3187	C5_01850C
5C12	orf19.642	CR_05120W
5D01	orf19.3753	CR_02190C
5D02	orf19.2753	C4_02490W
5D03	orf19.5224	C1_12410C
5D05	orf19.4432	C1_07380C
5D06	orf19.2791	C1_07540C

5D07	orf19.3199	C5_01760C
5D09	orf19.4518	C2_04360W
5D10	orf19.1255	C4_05700W
5D12	orf19.35	C2_06600W
5E01	orf19.4890	C1_10210C
5E03	orf19.3188	C5_01840C
5E04	orf19.4225	C5_02180C
5E05	orf19.166	CR_02560C
5E07	orf19.3751	CR_02210W
5E08	orf19.4545	C1_01790W
5E09	orf19.5241	C1_12230W
5E11	orf19.7374	C3_05930W
5F02	orf19.6243	C1_06680W
5F03	orf19.4152	C5_01580C
5F04	orf19.3012	C1_03200C
5F05	orf19.7033	C7_00840C
5F06	orf19.2808	C3_04020C
5F09	orf19.2097	C2_00350W
5F10	orf19.4145	C5_01500C
5F12	orf19.173	CR_02510W
5G01	orf19.4766	C1_09090C
5G02	orf19.7371	C3_05910W
5G05	orf19.702	CR_06680C
5G07	orf19.7372	C3_05920W
5G10	orf19.5175	C7_02860C
5G11	orf19.2771	C4_02310W
5G12	orf19.1499	C2_01890W
5H01	orf19.2236	C2_06830C
5H02	orf19.4593	C4_02000C
5H03	orf19.5026	C1_13880C
5H06	orf19.2504	C3_01000W
5H07	orf19.7510	CR_00260W
5H09	orf19.4242	C5_02340C
5H10	orf19.940	C5_00490C
5H11	orf19.1474	C2_01640W
5H12	orf19.6193	C1_07110W
6A01	orf19.5162	C7_02990W
6A02	orf19.7451	C3_06620W
6A03	orf19.1648	C3_02000W
6A04	orf19.6926	C3_03890W
6A06	orf19.663	C1_11400C
6A07	orf19.3256	CR_01000C
6A09	orf19.1777	C2_10050W
6A11	orf19.3840	C4_04460C

6A12	orf19.1490	C2_01780W
6B01	orf19.2752	C4_02500C
6B02	orf19.4969	C1_13420C
6B03	orf19.1259	C4_05740C
6B04	orf19.4308	C5_02840C
6B05	orf19.744	C4_05140C
6B06	orf19.130	C6_01190C
6B07	orf19.1842	CR_06910W
6B08	orf19.5148	C7_03070C
6B10	orf19.2695	C4_02970C
6B11	orf19.7044	C7_00740W
6B12	orf19.1814	CR_07090W
6C03	orf19.6573	C7_01550W
6C07	orf19.1513	C2_01990C
6C08	orf19.5219	C1_12450C
6C11	orf19.139	C6_01320W
3R1A02	orf19.3947	C5_04630W
3R1A03	orf19.3551	C2_05210W
3R1A04	orf19.3263	CR_00930W
3R1A06	orf19.576	C5_00740W
3R1A07	orf19.4473	C1_03980W
3R1A08	orf19.2519	CR_01360W
3R1A09	orf19.6142	CR_07260C
3R1A11	orf19.7617	CR_10400W
3R1B01	orf19.6163	C3_00860W
3R1B03	orf19.7663	CR_10740W
3R1B04	orf19.6537	C7_01830W
3R1B05	orf19.3978	C5_04910W
3R1B06	orf19.3715	CR_07860C
3R1B07	orf19.7195	C7_03870W
3R1B08	orf19.5698	C5_00030W
3R1B09	orf19.703	CR_06690C
3R1B11	orf19.2267	C2_07120W
3R1B12	orf19.2922	C4_06150C
3R1C01	orf19.4837	C1_09730W
3R1C02	orf19.6997	C3_05620W
3R1C03	orf19.6995	C3_05600W
3R1C04	orf19.1571	C2_02470C
3R1C05	orf19.1224	C2_06680W
3R1C06	orf19.676	C1_11250W
3R1C07	orf19.6170	C3_07770C
3R1C08	orf19.4537	C1_01860W
3R1C09	orf19.115	C6_01090C
3R1C10	orf19.3589	C2_08820C

3R1C11	orf19.4105	C2_06130W
3R1C12	orf19.5485	C2_06220C
3R1D01	orf19.967	C5_00280C
3R1D02	orf19.3264	CR_00920W
3R1D03	orf19.366	C4_00030C
3R1D04	orf19.71	C1_03690W
3R1D05	orf19.3752	CR_02200C
3R1D06	orf19.7035	C7_00820W
3R1D07	orf19.3736	CR_02330C
3R1D08	orf19.2983	C1_02920W
3R1D09	orf19.3760	C1_12560C
3R1D11	orf19.5764	C6_03890C
3R1E01	orf19.5276	C1_11910W
3R1E03	orf19.7355	C3_05740C
3R1E05	orf19.2429	C1_06160W
3R1E07	orf19.5377	C3_00780W
3R1E08	orf19.4560	C4_02270C
3R1E09	orf19.6851	C1_04570C
3R1E10	orf19.3613	C2_08630C
3R1E12	orf19.245	C3_02560W
3R1F02	orf19.2119	C2_00140W
3R1F03	orf19.4438	C1_07330W
3R1F04	orf19.2174	C2_08110W
3R1F05	orf19.6885	C2_05740W
3R1F06	orf19.7308	CR_09120C
3R1F07	orf19.5551	C6_02780C
3R1F08	orf19.4284	C5_02670W
3R1F09	orf19.5954	C3_04870W
3R1F10	orf19.1476	C2_01650W
3R1G01	orf19.6034	C1_00710C
3R1G02	orf19.6437	CR_08630W
3R1G03	orf19.1574	C2_02500W
3R1G04	orf19.2728	C4_02700W
3R1G06	orf19.2084	C2_00450C
3R1G08	orf19.1542	C2_02240C
3R1G12	orf19.4206	C6_00540W
3R1H03	orf19.996	C1_10520W
3R1H04	orf19.306	C3_03090W
3R1H06	orf19.2093	C2_00380C
3R1H08	orf19.5934	C3_04740C
3R1H09	orf19.1002	C1_10560C
3R1H10	orf19.122	C6_01150W
3R1H11	orf19.7257	C1_14400C
3R1H12	orf19.7247	C1_14340C

3R2A03	orf19.6760	C3_07320W
3R2A04	orf19.564	C2_09370C
3R2A06	orf19.4549	C1_01760W
3R2A07	orf19.1434	C2_08380C
3R2A08	orf19.4162	C4_00790C
3R2B01	orf19.1251	C4_05660C
3R2B02	orf19.1956	C5_01140C
3R2B03	orf19.7473	CR_00600C
3R2B07	orf19.1792	C4_05340W
3R2B08	orf19.7310	CR_09140C
3R2B09	orf19.1943	C5_01260W
3R2B10	orf19.5367	C2_10840W
3R2C01	orf19.6861	C4_05220C
3R2C03	orf19.4251	C5_02430W
3R2C06	orf19.2713	C4_02820W
3R2C08	orf19.5265	C1_12010C
3R2C09	orf19.5004	C1_13660W
3R2D01	orf19.3865	CR_06110C
3R2D04	orf19.3093	C4_07120C
3R2D05	orf19.3231	CR_01230C
3R2E01	orf19.1605	C2_09470C
3R2E02	orf19.6610	CR_09520C
3R2E03	orf19.5798	C2_03030W
3R2E04	orf19.6222	C1_06880C
3R2E05	orf19.3608	C2_08680W
3R2E07	orf19.7221	C1_14140C
3R2E08	orf19.2417	CR_03100W
3R2E09	orf19.6568	C7_01580W
3R2F01	orf19.3623	C2_08560W
3R2F06	orf19.6763	C3_07310C
3R2F07	orf19.658	C1_11440C
3R2F08	orf19.4367	CR_03680C
3R2G01	orf19.4136	C5_01460W
3R2G03	orf19.643	CR_05100W
3R2G05	orf19.7402	C3_06300W
3R2G06	orf19.964	C5_00300C
3R2G09	orf19.6046	C1_00600W

Bibliography

1. Laehnemann D, Pena-Miller R, Rosenstiel P, Beardmore R, Jansen G, Schulenburg H. (2014). Genomics of rapid adaptation to antibiotics: convergent evolution and scalable sequence amplification. Genome Biol Evol **6:**1287-1301.
2. Shapiro RS. (2015). Antimicrobial-Induced DNA Damage and Genomic Instability in Microbial Pathogens. Plos Pathog **11**.
3. Chen GB, Bradford WD, Seidel CW, Li R. (2012). Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. Nature **482:**246-250.
4. Gerstein AC, Fu MS, Mukaremera L, Li Z, Ormerod KL, Fraser JA, Berman J, Nielsen K. (2015). Polyploid titan cells produce haploid and aneuploid progeny to promote stress adaptation. MBio **6:**e01340-01315.
5. Edmond MB, Wallace SE, McClish DK, Pfaller MA, Jones RN, Wenzel RP. (1999). Nosocomial bloodstream infections in United States hospitals: a three-year analysis. Clin Infect Dis **29:**239-244.
6. Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, *et al*. (2009). Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature **459:**657-662.
7. Dujon B. (2010). Yeast evolutionary genomics. Nat Rev Genet **11:**512-524.
8. Odds FC, Bougnoux ME, Shaw DJ, Bain JM, Davidson AD, *et al*. (2007). Molecular phylogenetics of *Candida albicans*. Eukaryot Cell **6:**1041-1052.
9. White TC, Andrews LE, Maltby D, Agabian N. (1995). The "universal" leucine codon CTG in the secreted aspartyl proteinase 1 (*SAP1*) gene of *Candida albicans* encodes a serine *in vivo*. J Bacteriol **177:**2953-2955.
10. Hickman MA, Zeng GS, Forche A, Hirakawa MP, Abbey D, *et al*. (2013). The 'obligate diploid' *Candida albicans* forms mating-competent haploids. Nature **494:**55-59.
11. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, *et al*. (2004). The diploid genome sequence of *Candida albicans*. Proc Natl Acad Sci U S A **101:**7329-7334.
12. 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.
13. Braun BR, van Het Hoog M, d'Enfert C, Martchenko M, Dungan J, *et al*. (2005). A human-curated annotation of the *Candida albicans* genome. PLoS Genet **1:**36-57.
14. Diogo D, Bouchier C, d'Enfert C, Bougnoux ME. (2009). Loss of heterozygosity in commensal isolates of the asexual diploid yeast *Candida albicans*. Fungal Genet Biol **46:**159-168.
15. Forche A, Magee PT, Selmecki A, Berman J, May G. (2009). Evolution in *Candida albicans* Populations During a Single Passage Through a Mouse Host. Genetics **182:**799-811.
16. Ford CB, Funt JM, Abbey D, Issi L, Guiducci C, *et al*. (2015). The evolution of drug resistance in clinical isolates of *Candida albicans*. Elife **4:**e00662.
17. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, *et al*. (2015). Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res **25:**413-425.
18. 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:**129-111.
19. Loll-Krippleber R, d'Enfert C, Feri A, Diogo D, Perin A, Marcet-Houben M, Bougnoux ME, Legrand M. (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.
20. 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.
21. Selmecki A, Forche A, Berman J. (2010). Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryot Cell **9:**991-1008.
22. Homann OR, Dea J, Noble SM, Johnson AD. (2009). A phenotypic profile of the *Candida albicans* regulatory network. PLoS Genet **5:**e1000783.
23. Martchenko M, Levitin A, Hogues H, Nantel A, Whiteway M. (2007). Transcriptional rewiring of fungal galactose-metabolism circuitry. Curr Biol **17:**1007-1013.
24. Stewart JA, Chaiken MF, Wang F, Price CM. (2012). Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. Mutat Res **730:**12-19.
25. Clarke L, Carbon J. (1983). Genomic substitutions of centromeres in *Saccharomyces cerevisiae*. Nature **305:**23-28.
26. Clarke L, Carbon J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature **287:**504-509.
27. McGrew J, Diehl B, Fitzgerald-Hayes M. (1986). Single base-pair mutations in centromere element III cause aberrant chromosome segregation in *Saccharomyces cerevisiae*. Mol Cell Biol **6:**530-538.
28. Morris CA, Moazed D. (2007). Centromere assembly and propagation. Cell **128:**647-650.
29. Tomkiel J, Cooke CA, Saitoh H, Bernat RL, Earnshaw WC. (1994). Cenp-C Is Required for Maintaining Proper Kinetochore Size and for a Timely Transition to Anaphase. J Cell Biol **125:**531-545.
30. Sanyal K, Baum M, Carbon J. (2004). Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. Proc Natl Acad Sci U S A **101:**11374-11379.
31. Mishra PK, Baum M, Carbon J. (2007). Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. Mol Genet Genomics **278:**455-465.
32. Baum M, Sanyal K, Mishra PK, Thaler N, Carbon J. (2006). Formation of functional centromeric chromatin is specified epigenetically in *Candida albicans*. Proc Natl Acad Sci U S A **103:**14877-14882.
33. Ketel C, Wang HSW, McClellan M, Bouchonville K, Selmecki A, Lahav T, Gerami-Nejad M, Berman J. (2009). Neocentromeres Form Efficiently at Multiple Possible Loci in *Candida albicans*. PLoS Genet **5**.
34. Jaco I, Canela A, Vera E, Blasco MA. (2008). Centromere mitotic recombination in mammalian cells. J Cell Biol **181:**885-892.
35. Symington LS, Petes TD. (1988). Meiotic Recombination within the Centromere of a Yeast Chromosome. Cell **52:**237-240.
36. Thakur J, Sanyal K. (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.
37. Mitra S, Gomez-Raja J, Larriba G, Dubey DD, Sanyal K. (2014). Rad51-Rad52 mediated maintenance of centromeric chromatin in *Candida albicans*. PLoS Genet **10:**e1004344.
38. Koren A, Tsai HJ, Tirosh I, Burrack LS, Barkai N, Berman J. (2010). Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. PLoS Genet **6:**e1001068.

39. Tsai HJ, Baller JA, Liachko I, Koren A, Burrack LS, Hickman MA, Thevandavakkam MA, Rusche LN, Berman J. (2014). Origin Replication Complex Binding, Nucleosome Depletion Patterns, and a Primary Sequence Motif Can Predict Origins of Replication in a Genome with Epigenetic Centromeres. Mbio **5**.
40. Greenfeder SA, Newlon CS. (1992). Replication Forks Pause at Yeast Centromeres. Mol Cell Biol **12:**4056-4066.
41. Forstemann K, Hoss M, Lingner J. (2000). Telomerase-dependent repeat divergence at the 3 ' ends of yeast telomeres. Nucleic Acids Res **28:**2690-2694.
42. Larrivee M, LeBel C, Wellinger RJ. (2004). The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. Genes Dev **18:**1391-1396.
43. McEachern MJ, Hicks JB. (1993). Unusually large telomeric repeats in the yeast *Candida albicans*. Mol Cell Biol **13:**551-560.
44. Gunisova S, Elboher E, Nosek J, Gorkovoy V, Brown Y, *et al*. (2009). Identification and comparative analysis of telomerase RNAs from Candida species reveal conservation of functional elements. RNA **15:**546-559.
45. Kupiec M. (2014). Biology of telomeres: lessons from budding yeast. FEMS Microbiol Rev **38:**144-171.
46. Castano I, Pan SJ, Zupancic M, Hennequin C, Dujon B, Cormack BP. (2005). Telomere length control and transcriptional regulation of subtelomeric adhesins in *Candida glabrata*. Mol Microbiol **55:**1246-1258.
47. Rosas-Hernandez LL, Juarez-Reyes A, Arroyo-Helguera OE, De Las Penas A, Pan SJ, Cormack BP, Castano I. (2008). yKu70/yKu80 and Rif1 Regulate Silencing Differentially at Telomeres in *Candida glabrata*. Eukaryot Cell **7:**2168-2178.
48. Hardy CF, Sussel L, Shore D. (1992). A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. Genes Dev **6:**801-814.
49. Chico L, Ciudad T, Hsu M, Lue NF, Larriba G. (2011). The *Candida albicans* Ku70 Modulates Telomere Length and Structure by Regulating Both Telomerase and Recombination. Plos One **6**.
50. Singh SM, Steinberg-Neifach O, Mian IS, Lue NF. (2002). Analysis of telomerase in *Candida albicans*: potential role in telomere end protection. Eukaryot Cell **1:**967-977.
51. Hsu M, Yu EY, Singh SM, Lue NF. (2007). Mutual dependence of *Candida albicans* Est1p and Est3p in telomerase assembly and activation. Eukaryot Cell **6:**1330-1338.
52. Pennock E, Buckley K, Lundblad V. (2001). Cdc13 delivers separate complexes to the telomere for end protection and replication. Cell **104:**387-396.
53. Bennett RJ. (2015). The parasexual lifestyle of *Candida albicans*. Curr Opin Microbiol **28:**10-17.
54. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, *et al*. (2004). Genome evolution in yeasts. Nature **430:**35-44.
55. Srikantha T, Daniels KJ, Pujol C, Sahni N, Yi S, Soll DR. (2012). Nonsex genes in the mating type locus of *Candida albicans* play roles in a/alpha biofilm formation, including impermeability and fluconazole resistance. PLoS Pathog **8:**e1002476.
56. Hull CM, Raisner RM, Johnson AD. (2000). Evidence for mating of the "asexual" yeast *Candida albicans* in a mammalian host. Science **289:**307-310.
57. Janbon G, Sherman F, Rustchenko E. (1998). Monosomy of a specific chromosome determines L-sorbose utilization: a novel regulatory mechanism in *Candida albicans*. Proc Natl Acad Sci U S A **95:**5150-5155.
58. Magee BB, Magee PT. (2000). Induction of mating in *Candida albicans* by construction of *MTLa* and *MTLalpha* strains. Science **289:**310-313.

59. Miller MG, Johnson AD. (2002). White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell **110:**293-302.
60. Slutsky B, Staebell M, Anderson J, Risen L, Pfaller M, Soll DR. (1987). "White-opaque transition": a second high-frequency switching system in *Candida albicans*. J Bacteriol **169:**189-197.
61. Slutsky B, Staebell M, Anderson J, Risen L, Pfaller M, Soll DR. (1987). "White-opaque transition": a second high-frequency switching system in *Candida albicans*. J Bacteriol **169:**189-197.
62. Lan CY, Newport G, Murillo LA, Jones T, Scherer S, Davis RW, Agabian N. (2002). Metabolic specialization associated with phenotypic switching in *Candida albicans*. Proc Natl Acad Sci U S A **99:**14907-14912.
63. Lohse MB, Johnson AD. (2008). Differential phagocytosis of white versus opaque *Candida albicans* by Drosophila and mouse phagocytes. PLoS One **3:**e1473.
64. Xie J, Tao L, Nobile CJ, Tong Y, Guan G, *et al*. (2013). White-opaque switching in natural *MTLa/alpha* isolates of *Candida albicans*: evolutionary implications for roles in host adaptation, pathogenesis, and sex. PLoS Biol **11:**e1001525.
65. Berman J, Hadany L. (2012). Does stress induce (para)sex? Implications for *Candida albicans* evolution. Trends Genet **28:**197-203.
66. Srikantha T, Borneman AR, Daniels KJ, Pujol C, Wu W, *et al*. (2006). TOS9 regulates white-opaque switching in *Candida albicans*. Eukaryot Cell **5:**1674-1687.
67. Lohse MB, Hernday AD, Fordyce PM, Noiman L, Sorrells TR, Hanson-Smith V, Nobile CJ, DeRisi JL, Johnson AD. (2013). Identification and characterization of a previously undescribed family of sequence-specific DNA-binding domains. Proc Natl Acad Sci U S A **110:**7660-7665.
68. Vinces MD, Kumamoto CA. (2007). The morphogenetic regulator Czf1p is a DNA-binding protein that regulates white-opaque switching in *Candida albicans*. Microbiology **153:**2877-2884.
69. Zordan RE, Miller MG, Galgoczy DJ, Tuch BB, Johnson AD. (2007). Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. PLoS Biol **5:**e256.
70. Lohse MB, Johnson AD. (2016). Identification and Characterization of Wor4, a New Transcriptional Regulator of White-Opaque Switching. G3 (Bethesda) **6:**721-729.
71. Hernday AD, Lohse MB, Nobile CJ, Noiman L, Laksana CN, Johnson AD. (2016). Ssn6 Defines a New Level of Regulation of White-Opaque Switching in *Candida albicans* and Is Required For the Stochasticity of the Switch. MBio **7**.
72. Bennett RJ, Johnson AD. (2003). Completion of a parasexual cycle in *Candida albicans* by induced chromosome loss in tetraploid strains. EMBO J **22:**2505-2515.
73. Hickman MA, Paulson C, Dudley A, Berman J. (2015). Parasexual Ploidy Reduction Drives Population Heterogeneity Through Random and Transient Aneuploidy in *Candida albicans*. Genetics **200:**781-+.
74. 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.
75. Bougnoux ME, Pujol C, Diogo D, Bouchier C, Soll DR, d'Enfert C. (2008). Mating is rare within as well as between clades of the human pathogen *Candida albicans*. Fungal Genet Biol **45:**221-231.
76. Graser Y, Volovsek M, Arrington J, Schonian G, Presber W, Mitchell TG, Vilgalys R. (1996). Molecular markers reveal that population structure of the human pathogen

Candida albicans exhibits both clonality and recombination. Proc Natl Acad Sci USA **93:**12473-12477.
77. Ibrahim AS, Magee BB, Sheppard DC, Yang M, Kauffman S, Becker J, Edwards JE, Jr., Magee PT. (2005). Effects of ploidy and mating type on virulence of *Candida albicans*. Infect Immun **73:**7366-7374.
78. van het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, *et al*. (2007). Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol **8**.
79. Chibana H, Iwaguchi SI, Homma M, Chindamporn A, Nakagawa Y, Tanaka K. (1994). Diversity of Tandemly Repetitive Sequences Due to Short Periodic Repetitions in the Chromosomes of *Candida albicans*. J Bacteriol **176:**3851-3858.
80. Iwaguchi S, Homma M, Chibana H, Tanaka K. (1992). Isolation and characterization of a repeated sequence (RPS1) of *Candida albicans*. J Gen Microbiol **138:**1893-1900.
81. 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.
82. Doi M, Chibana H, Nakagawa Y, Tanaka K. (1998). Discrimination among the clinical isolates of *Candida albicans* by amplification of the repetitive sequences, alts. Microbiol Immunol **42:**227-230.
83. Chu WS, Magee BB, Magee PT. (1993). Construction of an *Sfi*I macrorestriction map of the *Candida albicans* genome. J Bacteriol **175:**6637-6651.
84. Chindamporn A, Nakagawa Y, Mizuguchi I, Chibana H, Doi M, Tanaka K. (1998). Repetitive sequences (RPSs) in the chromosomes of *Candida albicans* are sandwiched between two novel stretches, HOK and RB2, common to each chromosome. Microbiology **144:**849-857.
85. Joly S, Pujol C, Soll DR. (2002). Microevolutionary changes and chromosomal translocations are more frequent at *RPS* loci in *Candida dubliniensis* than in *Candida albicans*. Infection Genetics and Evolution **2:**19-37.
86. Joly S, Pujol C, Rysz M, Vargas K, Soll DR. (1999). Development and characterization of complex DNA fingerprinting probes for the infectious yeast *Candida dubliniensis*. J Clin Microbiol **37:**1035-1044.
87. Padmanabhan S, Thakur J, Siddharthan R, Sanyal K. (2008). Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. Proc Natl Acad Sci U S A **105:**19797-19802.
88. Lephart PR, Chibana H, Magee PT. (2005). Effect of the major repeat sequence on chromosome loss in *Candida albicans*. Eukaryot Cell **4:**733-741.
89. Lephart PR, Magee PT. (2006). Effect of the major repeat sequence on mitotic recombination in *Candida albicans*. Genetics **174:**1737-1744.
90. Freire-Beneitez V, Price RJ, Tarrant D, Berman J, Buscaino A. (2016). *Candida albicans* repetitive elements display epigenetic diversity and plasticity. Sci Rep **6:**22989.
91. Uhl MA, Biery M, Craig N, Johnson AD. (2003). Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen *C.albicans*. EMBO J **22:**2668-2678.
92. Chibana H, Magee PT. (2009). The enigma of the major repeat sequence of *Candida albicans*. Future Microbiol **4:**171-179.
93. Petes TD. (1979). Yeast ribosomal DNA genes are located on chromosome XII. Proc Natl Acad Sci U S A **76:**410-414.
94. James SA, O'Kelly MJT, Carter DM, Davey RP, van Oudenaarden A, Roberts IN. (2009). Repetitive sequence variation and dynamics in the ribosomal DNA array of

Saccharomyces cerevisiae as revealed by whole-genome resequencing. Genome Res **19:**626-635.
95. Fritze CE, Verschueren K, Strich R, Easton Esposito R. (1997). Direct evidence for *SIR2* modulation of chromatin structure in yeast rDNA. EMBO J **16:**6495-6509.
96. Li C, Mueller JE, Bryk M. (2006). Sir2 represses endogenous polymerase II transcription units in the ribosomal DNA nontranscribed spacer. Mol Biol Cell **17:**3848-3859.
97. Straight AF, Shou W, Dowd GJ, Turck CW, Deshaies RJ, Johnson AD, Moazed D. (1999). Net1, a Sir2-associated nucleolar protein required for rDNA silencing and nucleolar integrity. Cell **97:**245-256.
98. Gottlieb S, Esposito RE. (1989). A new role for a yeast transcriptional silencer gene, *SIR2*, in regulation of recombination in ribosomal DNA. Cell **56:**771-776.
99. Sinclair DA, Guarente L. (1997). Extrachromosomal rDNA circles - A cause of aging in yeast. Cell **91:**1033-1042.
100. Huber D, Rustchenko E. (2001). Large circular and linear rDNA plasmids in *Candida albicans*. Yeast **18:**261-272.
101. Singer MF. (1982). SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell **28:**433-434.
102. Ju G, Skalka AM. (1980). Nucleotide sequence analysis of the long terminal repeat (LTR) of avian retroviruses: structural similarities with transposable elements. Cell **22:**379-386.
103. Neuveglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S. (2002). Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res **12:**930-943.
104. Zhang L, Yan L, Jiang J, Wang Y, Jiang Y, Yan T, Cao Y. (2014). The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast *Candida albicans*. Virulence **5:**655-664.
105. Matthews GD, Goodwin TJ, Butler MI, Berryman TA, Poulter RT. (1997). pCal, a highly unusual *Ty1/copia* retrotransposon from the pathogenic yeast *Candida albicans*. J Bacteriol **179:**7118-7128.
106. Horie K, Saito ES, Keng VW, Ikeda R, Ishihara H, Takeda J. (2007). Retrotransposons influence the mouse transcriptome: implication for the divergence of genetic traits. Genetics **176:**815-827.
107. Knight SA, Labbe S, Kwon LF, Kosman DJ, Thiele DJ. (1996). A widespread transposable element masks expression of a yeast copper transport gene. Genes Dev **10:**1917-1929.
108. Anaya N, Roncero MI. (1996). Stress-induced rearrangement of Fusarium retrotransposon sequences. Mol Gen Genet **253:**89-94.
109. Carreto L, Eiriz MF, Gomes AC, Pereira PM, Schuller D, Santos MA. (2008). Comparative genomics of wild type yeast strains unveils important genome diversity. BMC Genomics **9:**524.
110. Pryde FE, Gorham HC, Louis EJ. (1997). Chromosome ends: all the same under their caps. Curr Opin Genet Dev **7:**822-828.
111. Naumov GI, Naumova ES, Korhola MP. (1995). Chromosomal-Polymorphism of *MEL* Genes in Some Populations of *Saccharomyces cerevisiae*. FEMS Microbiol Lett **127:**41-45.
112. Brown CA, Murray AW, Verstrepen KJ. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr Biol **20:**895-903.
113. Wenger JW, Piotrowski J, Nagarajan S, Chiotti K, Sherlock G, Rosenzweig F. (2011). Hunger Artists: Yeast Adapted to Carbon Limitation Show Trade-Offs under Carbon Sufficiency. PLoS Genet **7**.

114. Zhang A, Petrov KO, Hyun ER, Liu Z, Gerber SA, Myers LC. (2012). The Tlo proteins are stoichiometric components of *Candida albicans* mediator anchored via the Med3 subunit. Eukaryot Cell **11:**874-884.
115. Lindsay AK, Morales DK, Liu ZL, Grahl N, Zhang A, Willger SD, Myers LC, Hogan DA. (2014). Analysis of *Candida albicans* Mutants Defective in the Cdk8 Module of Mediator Reveal Links between Metabolism and Biofilm Formation. PLoS Genet **10**.
116. Zhang A, Liu Z, Myers LC. (2013). Differential regulation of white-opaque switching by individual subunits of *Candida albicans* mediator. Eukaryot Cell **12:**1293-1304.
117. Uwamahoro N, Qu Y, Jelicic B, Lo TL, Beaurepaire C, *et al*. (2012). The functions of Mediator in *Candida albicans* support a role in shaping species-specific gene expression. PLoS Genet **8:**e1002613.
118. Tebbji F, Chen Y, Richard Albert J, Gunsalus KT, Kumamoto CA, Nantel A, Sellam A, Whiteway M. (2014). A functional portrait of Med7 and the mediator complex in *Candida albicans*. PLoS Genet **10:**e1004770.
119. Sullivan DJ, Berman J, Myers LC, Moran GP. (2015). Telomeric ORFS in *Candida albicans*: Does Mediator Tail Wag the Yeast? PLoS Pathog **11**.
120. Lasker BA, Page LS, Lott TJ, Kobayashi GS. (1992). Isolation, characterization, and sequencing of *Candida albicans* repetitive element 2. Gene **116:**51-57.
121. Thrash-Bingham C, Gorman JA. (1993). Identification, characterization and sequence of *Candida albicans* repetitive DNAs *Rel-1* and *Rel-2*. Curr Genet **23:**455-462.
122. Goodwin TJ, Poulter RT. (2000). Multiple LTR-retrotransposon families in the asexual yeast *Candida albicans*. Genome Res **10:**174-191.
123. Whelan WL, Magee PT. (1981). Natural heterozygosity in *Candida albicans*. J Bacteriol **145:**896-903.
124. Magwene PM, Kayikci O, Granek JA, Reininga JM, Scholl Z, Murray D. (2011). Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A **108:**1987-1992.
125. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, *et al*. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409:**928-933.
126. Andaluz E, Bellido A, Gomez-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.
127. Whelan WL, Partridge RM, Magee PT. (1980). Heterozygosity and segregation in *Candida albicans*. Mol Gen Genet **180:**107-113.
128. Whelan WL, Soll DR. (1982). Mitotic recombination in *Candida albicans*: recessive lethal alleles linked to a gene required for methionine biosynthesis. Mol Gen Genet **187:**477-485.
129. Gomez-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.
130. Holmes AR, Tsao S, Ong SW, Lamping E, Niimi K, *et al*. (2006). Heterozygosity and functional allelic variation in the *Candida albicans* efflux pump genes *CDR1* and *CDR2*. Mol Microbiol **62:**170-186.
131. Bauer NC, Corbett AH, Doetsch PW. (2015). The current state of eukaryotic DNA base damage and repair. Nucleic Acids Res **43:**10083-10101.
132. Kolodner RD. (2016). A personal historical view of DNA mismatch repair with an emphasis on eukaryotic DNA mismatch repair. DNA Repair (Amst) **38:**3-13.

133. Saini N, Ramakrishnan S, Elango R, Ayyar S, Zhang Y, *et al*. (2013). Migrating bubble during break-induced replication drives conservative DNA synthesis. Nature **502:**389-392.
134. Sakofsky CJ, Ayyar S, Malkova A. (2012). Break-induced replication and genome stability. Biomolecules **2:**483-504.
135. Lazzerini-Denchi E, Sfeir A. (2016). Stop pulling my strings - what telomeres taught us about the DNA damage response. Nat Rev Mol Cell Biol **17:**364-378.
136. Watt DL, Johansson E, Burgers PM, Kunkel TA. (2011). Replication of ribonucleotide-containing DNA templates by yeast replicative polymerases. DNA Repair (Amst) **10:**897-902.
137. McElhinny SAN, Watts BE, Kumar D, Watt DL, Lundstrom EB, Burgers PMJ, Johansson E, Chabes A, Kunkel TA. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. Proc Natl Acad Sci U S A **107:**4949-4954.
138. Lindahl T. (1993). Instability and decay of the primary structure of DNA. Nature **362:**709-715.
139. DeRose EF, Perera L, Murray MS, Kunkel TA, London RE. (2012). Solution structure of the Dickerson DNA dodecamer containing a single ribonucleotide. Biochemistry **51:**2407-2416.
140. Eder PS, Walder RY, Walder JA. (1993). Substrate specificity of human RNase H1 and its role in excision repair of ribose residues misincorporated in DNA. Biochimie **75:**123-126.
141. Sparks JL, Chon H, Cerritelli SM, Kunkel TA, Johansson E, Crouch RJ, Burgers PM. (2012). RNase H2-initiated ribonucleotide excision repair. Mol Cell **47:**980-986.
142. Vaisman A, Woodgate R. (2015). Redundancy in ribonucleotide excision repair: Competition, compensation, and cooperation. DNA Repair (Amst) **29:**74-82.
143. Allen-Soltero S, Martinez SL, Putnam CD, Kolodner RD. (2014). A *Saccharomyces cerevisiae* RNase H2 interaction network functions to suppress genome instability. Mol Cell Biol **34:**1521-1534.
144. Aguilera A, Klein HL. (1988). Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. Genetics **119:**779-790.
145. O'Connell K, Jinks-Robertson S, Petes TD. (2015). Elevated Genome-Wide Instability in Yeast Mutants Lacking RNase H Activity. Genetics **201:**963-975.
146. D'Autreaux B, Toledano MB. (2007). ROS as signalling molecules: mechanisms that generate specificity in ROS homeostasis. Nat Rev Mol Cell Biol **8:**813-824.
147. Cooper DN, Youssoufian H. (1988). The CpG dinucleotide and human genetic disease. Hum Genet **78:**151-155.
148. Walavalkar NM, Cramer JM, Buchwald WA, Scarsdale JN, Williams DC, Jr. (2014). Solution structure and intramolecular exchange of methyl-cytosine binding domain protein 4 (MBD4) on DNA suggests a mechanism to scan for mCpG/TpG mismatches. Nucleic Acids Res **42:**11218-11232.
149. Fu D, Calvo JA, Samson LD. (2012). Balancing repair and tolerance of DNA damage caused by alkylating agents. Nat Rev Cancer **12:**104-120.
150. Legrand M, Chan CL, Jauert PA, Kirkpatrick DT. (2008). Analysis of base excision and nucleotide excision repair in *Candida albicans*. Microbiology **154:**2446-2456.
151. Cadet J, Sage E, Douki T. (2005). Ultraviolet radiation-mediated damage to cellular DNA. Mutat Res **571:**3-17.
152. Li S. (2015). Transcription coupled nucleotide excision repair in the yeast *Saccharomyces cerevisiae*: The ambiguous role of Rad26. DNA Repair (Amst) **36:**43-48.

153. Boiteux S, Jinks-Robertson S. (2013). DNA repair mechanisms and the bypass of DNA damage in *Saccharomyces cerevisiae*. Genetics **193:**1025-1064.
154. Sung P, Prakash L, Matson SW, Prakash S. (1987). Rad3 Protein of Saccharomyces-Cerevisiae Is a DNA Helicase. Proc Natl Acad Sci U S A **84:**8951-8955.
155. Sung P, Guzder SN, Prakash L, Prakash S. (1996). Reconstitution of TFIIH and requirement of its DNA helicase subunits, Rad3 and Rad25, in the incision step of nucleotide excision repair. J Biol Chem **271:**10821-10826.
156. Eckardt F, Haynes RH. (1977). Induction of Pure and Sectored Mutant Clones in Excision-Proficient and Deficient Strains of Yeast. Mutat Res **43:**327-338.
157. Drake JW. (1999). The Distribution of Rates of Spontaneous Mutation over Viruses, Prokaryotes, and Eukaryotes. Ann N Y Acad Sci **870:**100-107.
158. Grilley M, Griffith J, Modrich P. (1993). Bidirectional excision in methyl-directed mismatch repair. J Biol Chem **268:**11830-11837.
159. Putnam CD. (2016). Evolution of the methyl directed mismatch repair system in *Escherichia coli*. DNA Repair (Amst) **38:**32-41.
160. Ghodgaonkar MM, Lazzaro F, Olivera-Pimentel M, Artola-Boran M, Cejka P, *et al*. (2013). Ribonucleotides Misincorporated into DNA Act as Strand-Discrimination Signals in Eukaryotic Mismatch Repair. Mol Cell **50:**323-332.
161. Kolodner RD. (2016). A personal historical view of DNA mismatch repair with an emphasis on eukaryotic DNA mismatch repair. DNA Repair (Amst) **38:**3-13.
162. Fishel R. (2015). Mismatch repair. J Biol Chem **290:**26395-26403.
163. Crouse GF. (2016). Non-canonical actions of mismatch repair. DNA Repair (Amst) **38:**102-109.
164. Datta A, Hendrix M, Lipsitch M, Jinks-Robertson S. (1997). Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. Proc Natl Acad Sci U S A **94:**9757-9762.
165. Sugawara N, Goldfarb T, Studamire B, Alani E, Haber JE. (2004). Heteroduplex rejection during single-strand annealing requires Sgs1 helicase and mismatch repair proteins Msh2 and Msh6 but not Pms1. Proc Natl Acad Sci U S A **101:**9315-9320.
166. Borts RH, Haber JE. (1987). Meiotic recombination in yeast: alteration by multiple heterozygosities. Science **237:**1459-1465.
167. Borts RH, Leung WY, Kramer W, Kramer B, Williamson M, Fogel S, Haber JE. (1990). Mismatch repair-induced meiotic recombination requires the *PMS1* gene product. Genetics **124:**573-584.
168. Sasaki M, Lange J, Keeney S. (2010). Genome destabilization by homologous recombination in the germ line. Nat Rev Mol Cell Biol **11:**182-195.
169. 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.
170. Healey KR, Zhao Y, Perez WB, Lockhart SR, Sobel JD, *et al*. (2016). Prevalent mutator genotype identified in fungal pathogen *Candida glabrata* promotes multi-drug resistance. Nat Commun **7:**11128.
171. Andaluz E, Calderone R, Reyes G, Larriba G. (2001). Phenotypic analysis and virulence of *Candida albicans LIG4* mutants. Infect Immun **69:**137-147.
172. Ciudad T, Andaluz E, Steinberg-Neifach O, Lue NF, Gow NA, 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.
173. Downs JA, Jackson SP. (2004). A means to a DNA end: the many roles of Ku. Nat Rev Mol Cell Biol **5:**367-378.

174. Emerson CH, Bertuch AA. (2016). Consider the workhorse: Nonhomologous end-joining in budding yeast. Biochem Cell Biol doi:10.1139/bcb-2016-0001**:**1-11.
175. Ma Y, Lu H, Tippin B, Goodman MF, Shimazaki N, Koiwai O, Hsieh CL, Schwarz K, Lieber MR. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. Mol Cell **16:**701-713.
176. Gu J, Lu H, Tippin B, Shimazaki N, Goodman MF, Lieber MR. (2007). XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. EMBO J **26:**1010-1023.
177. Tsukuda T, Fleming AB, Nickoloff JA, Osley MA. (2005). Chromatin remodelling at a DNA double-strand break site in *Saccharomyces cerevisiae*. Nature **438:**379-383.
178. Shim EY, Hong SJ, Oum JH, Yanez Y, Zhang Y, Lee SE. (2007). RSC mobilizes nucleosomes to improve accessibility of repair machinery to the damaged chromatin. Mol Cell Biol **27:**1602-1613.
179. Karathanasis E, Wilson TE. (2002). Enhancement of *Saccharomyces cerevisiae* end-joining efficiency by cell growth stage but not by impairment of recombination. Genetics **160:**1015-1027.
180. Symington LS. (2014). End Resection at Double-Strand Breaks: Mechanism and Regulation. Cold Spring Harb Perspect Biol **6**.
181. Aylon Y, Liefshitz B, Kupiec M. (2004). The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. EMBO J **23:**4868-4875.
182. Karanam K, Kafri R, Loewer A, Lahav G. (2012). Quantitative Live Cell Imaging Reveals a Gradual Shift between DNA Repair Mechanisms and a Maximal Use of HR in Mid S Phase. Mol Cell **47:**320-329.
183. Lee SE, Pâques F, Sylvan J, Haber JE. (1999). Role of yeast *SIR* genes and mating type in directing DNA double-strand breaks to homologous and non-homologous repair paths. Curr Biol **9:**767-770.
184. Astrom SU, Okamura SM, Rine J. (1999). Yeast cell-type regulation of DNA repair. Nature **397:**310.
185. Mao Z, Bozzella M, Seluanov A, Gorbunova V. (2008). Comparison of nonhomologous end joining and homologous recombination in human cells. DNA Repair (Amst) **7:**1765-1771.
186. Shibata A, Conrad S, Birraux J, Geuting V, Barton O, *et al*. (2011). Factors determining DNA double-strand break repair pathway choice in G2 phase. EMBO J **30:**1079-1092.
187. Woodbine L, Brunton H, Goodarzi AA, Shibata A, Jeggo PA. (2011). Endogenously induced DNA double strand breaks arise in heterochromatic DNA regions and require ataxia telangiectasia mutated and Artemis for their repair. Nucleic Acids Research **39:**6986-6997.
188. Andaluz E, Ciudad A, Rubio Coque J, Calderone R, Larriba G. (1999). Cell cycle regulation of a DNA ligase-encoding gene (*CaLIG4*) from *Candida albicans*. Yeast **15:**1199-1210.
189. Chauhan N, Ciudad T, Rodriguez-Alejandre A, Larriba G, Calderone R, Andaluz E. (2005). Virulence and karyotype analyses of *rad52* mutants of *Candida albicans*: regeneration of a truncated chromosome of a reintegrant strain (*rad52*/*RAD52*) in the host. Infect Immun **73:**8069-8078.
190. Andaluz E, Ciudad T, Larriba G. (2002). An evaluation of the role of in genomic instability and adaptive mutagenesis in *Candida albicans*. FEMS Yeast Res **2:**341-348.
191. Boboila C, Alt FW, Schwer B. (2012). Classical and alternative end-joining pathways for repair of lymphocyte-specific and general DNA double-strand breaks. Adv Immunol **116:**1-49.

192. Boulton SJ, Jackson SP. (1996). *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. Embo J **15:**5093-5103.
193. Yu X, Gabriel A. (2003). Ku-dependent and Ku-independent end-joining pathways lead to chromosomal rearrangements during double-strand break repair in *Saccharomyces cerevisiae*. Genetics **163:**843-856.
194. Truong LN, Li Y, Shi LZ, Hwang PY, He J, Wang H, Razavian N, Berns MW, Wu X. (2013). Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. Proc Natl Acad Sci U S A **110:**7720-7725.
195. Lee K, Lee SE. (2007). *Saccharomyces cerevisiae* Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. Genetics **176:**2003-2014.
196. Villarreal DD, Lee K, Deem A, Shim EY, Malkova A, Lee SE. (2012). Microhomology directs diverse DNA break repair pathways and chromosomal translocations. PLoS Genet **8:**e1003026.
197. Ma JL, Kim EM, Haber JE, Lee SE. (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.
198. Meyer D, Fu BX, Heyer WD. (2015). DNA polymerases delta and lambda cooperate in repairing double-strand breaks by microhomology-mediated end-joining in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A **112:**E6907-6916.
199. Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A. (2015). Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. Nature **518:**254-257.
200. Deng SK, Gibb B, de Almeida MJ, Greene EC, Symington LS. (2014). RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. Nat Struct Mol Biol **21:**405-412.
201. Mitelman F, Johansson B, Mertens F. (2007). The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer **7:**233-245.
202. Sfeir A, Symington LS. (2015). Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? Trends Biochem Sci **40:**701-714.
203. Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, *et al*. (2007). Paired-end mapping reveals extensive structural variation in the human genome. Science **318:**420-426.
204. van Schendel R, Roerink SF, Portegijs V, van den Heuvel S, Tijsterman M. (2015). Polymerase Theta is a key driver of genome evolution and of CRISPR/Cas9-mediated mutagenesis. Nat Commun **6**.
205. Symington LS. (2016). Mechanism and regulation of DNA end resection in eukaryotes. Crit Rev Biochem Mol Biol **51:**195-212.
206. Renkawitz J, Lademann CA, Jentsch S. (2014). Mechanisms and principles of homology search during recombination. Nat Rev Mol Cell Biol **15:**369-383.
207. Garcia-Prieto F, Gomez-Raja J, Andaluz E, Calderone R, Larriba G. (2010). Role of the homologous recombination genes *RAD51* and *RAD59* in the resistance of *Candida albicans* to UV light, radiomimetic and anti-tumor compounds and oxidizing agents. Fungal Genet Biol **47:**433-445.
208. Bellido A, Andaluz E, Gomez-Raja J, Alvarez-Barrientos A, Larriba G. (2015). Genetic interactions among homologous recombination mutants in *Candida albicans*. Fungal Genet Biol **74:**10-20.

209. Sugawara N, Haber JE. (1992). Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol Cell Biol **12:**563-575.
210. Jain S, Sugawara N, Lydeard J, Vaze M, Tanguy Le Gac N, Haber JE. (2009). A recombination execution checkpoint regulates the choice of homologous recombination pathway during DNA double-strand break repair. Genes Dev **23:**291-303.
211. Malkova A, Naylor ML, Yamaguchi M, Ira G, Haber JE. (2005). RAD51-dependent break-induced replication differs in kinetics and checkpoint responses from RAD51-mediated gene conversion. Mol Cell Biol **25:**933-944.
212. Wilson MA, Kwon Y, Xu Y, Chung WH, Chi P, *et al*. (2013). Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration. Nature **502:**393-396.
213. Lydeard JR, Lipkin-Moore Z, Sheu YJ, Stillman B, Burgers PM, Haber JE. (2010). Break-induced replication requires all essential DNA replication factors except those specific for pre-RC assembly. Genes Dev **24:**1133-1144.
214. Donnianni RA, Symington LS. (2013). Break-induced replication occurs by conservative DNA synthesis. Proc Natl Acad Sci U S A **110:**13475-13480.
215. Deem A, Keszthelyi A, Blackgrove T, Vayl A, Coffey B, Mathur R, Chabes A, Malkova A. (2011). Break-Induced Replication Is Highly Inaccurate. PLoS Biol **9**.
216. Smith CE, Llorente B, Symington LS. (2007). Template switching during break-induced replication. Nature **447:**102-105.
217. Stafa A, Donnianni RA, Timashev LA, Lam AF, Symington LS. (2014). Template Switching During Break-Induced Replication Is Promoted by the Mph1 Helicase in *Saccharomyces cerevisiae*. Genetics **196:**1017-+.
218. Difilippantonio MJ, Petersen S, Chen HT, Johnson R, Jasin M, Kanaar R, Ried T, Nussenzweig A. (2002). Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. J Exp Med **196:**469-480.
219. VanHulle K, Lemoine FJ, Narayanan V, Downing B, Hull K, *et al*. (2007). Inverted DNA repeats channel repair of distant double-strand breaks into chromatid fusions and chromosomal rearrangements. Mol Cell Biol **27:**2601-2614.
220. Ira G, Haber JE. (2002). Characterization of RAD51-independent break-induced replication that acts preferentially with short homologous sequences. Mol Cell Biol **22:**6384-6392.
221. Bosco G, Haber JE. (1998). Chromosome break-induced DNA replication leads to nonreciprocal translocations and telomere capture. Genetics **150:**1037-1047.
222. Signon L, Malkova A, Naylor ML, Klein H, Haber JE. (2001). Genetic requirements for RAD51- and RAD54-independent break-induced replication repair of a chromosomal double-strand break. Mol Cell Biol **21:**2048-2056.
223. Lee JA, Carvalho CM, Lupski JR. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell **131:**1235-1247.
224. Hastings PJ, Ira G, Lupski JR. (2009). A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet **5:**e1000327.
225. Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. (2009). The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nat Genet **41:**849-853.
226. Sakofsky CJ, Ayyar S, Deem AK, Chung WH, Ira G, Malkova A. (2015). Translesion Polymerases Drive Microhomology-Mediated Break-Induced Replication Leading to Complex Chromosomal Rearrangements. Mol Cell **60:**860-872.

227. McEachern MJ, Haber JE. (2006). Break-induced replication and recombinational telomere elongation in yeast. Annu Rev Biochem **75:**111-135.
228. Sun H, Treco D, Schultes NP, Szostak JW. (1989). Double-Strand Breaks at an Initiation Site for Meiotic Gene Conversion. Nature **338:**87-90.
229. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW. (1983). The double-strand-break repair model for recombination. Cell **33:**25-35.
230. Bizard AH, Hickson ID. (2014). The dissolution of double Holliday junctions. Cold Spring Harb Perspect Biol **6:**a016477.
231. Morrical SW. (2015). DNA-pairing and annealing processes in homologous recombination and homology-directed repair. Cold Spring Harb Perspect Biol **7:**a016444.
232. Haber JE, Ray BL, Kolb JM, White CI. (1993). Rapid Kinetics of Mismatch Repair of Heteroduplex DNA That Is Formed during Recombination in Yeast. Proc Natl Acad Sci U S A **90:**3363-3367.
233. Reiter LT, Hastings PJ, Nelis E, De Jonghe P, Van Broeckhoven C, Lupski JR. (1998). Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. Am J Hum Genet **62:**1023-1033.
234. Jain S, Sugawara N, Haber JE. (2016). Role of Double-Strand Break End-Tethering during Gene Conversion in *Saccharomyces cerevisiae*. PLoS Genet **12:**e1005976.
235. Hicks WM, Yamaguchi M, Haber JE. (2011). Real-time analysis of double-strand DNA break repair by homologous recombination. Proc Natl Acad Sci U S A **108:**3108-3115.
236. St Charles J, Petes TD. (2013). High-resolution mapping of spontaneous mitotic recombination hotspots on the 1.1 Mb arm of yeast chromosome IV. PLoS Genet **9:**e1003434.
237. Yim E, O'Connell KE, St Charles J, Petes TD. (2014). High-resolution mapping of two types of spontaneous mitotic gene conversion events in *Saccharomyces cerevisiae*. Genetics **198:**181-192.
238. Nickoloff JA, Sweetser DB, Clikeman JA, Khalsa GJ, Wheeler SL. (1999). Multiple heterologies increase mitotic double-strand break-induced allelic gene conversion tract lengths in yeast. Genetics **153:**665-679.
239. Tsaponina O, Haber JE. (2014). Frequent Interchromosomal Template Switches during Gene Conversion in *S. cerevisiae*. Mol Cell **55:**615-625.
240. Hicks WM, Kim M, Haber JE. (2010). Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. Science **329:**82-85.
241. Gangloff S, Zou H, Rothstein R. (1996). Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. EMBO J **15:**1715-1725.
242. Duret L, Galtier N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet **10:**285-311.
243. Pessia E, Popa A, Mousset S, Rezvoy C, Duret L, Marais GA. (2012). Evidence for widespread GC-biased gene conversion in eukaryotes. Genome Biol Evol **4:**675-682.
244. Galtier N. (2003). Gene conversion drives GC content evolution in mammalian histones. Trends Genet **19:**65-68.
245. Sugawara N, Haber JE. (1992). Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol Cell Biol **12:**563-575.
246. Mansour WY, Schumacher S, Rosskopf R, Rhein T, Schmidt-Petersen F, *et al*. (2008). Hierarchy of nonhomologous end-joining, single-strand annealing and gene conversion at site-directed DNA double-strand breaks. Nucleic Acids Res **36:**4088-4098.

247. Sugawara N, Ira G, Haber JE. (2000). DNA length dependence of the single-strand annealing pathway and the role of *Saccharomyces cerevisiae RAD59* in double-strand break repair. Mol Cell Biol **20:**5300-5309.
248. Saleh-Gohari N, Helleday T. (2004). Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. Nucleic Acids Res **32:**3683-3688.
249. Chakraborty U, George CM, Lyndaker AM, Alani E. (2016). A Delicate Balance Between Repair and Replication Factors Regulates Recombination Between Divergent DNA Sequences in *Saccharomyces cerevisiae*. Genetics **202:**525-540.
250. Mardirosian M, Nalbandyan L, Miller AD, Phan C, Kelson EP, Fischhaber PL. (2016). Saw1 localizes to repair sites but is not required for recruitment of Rad10 to repair intermediates bearing short non-homologous 3' flaps during single-strand annealing in *S. cerevisiae*. Mol Cell Biochem **412:**131-139.
251. Sugawara N, Paques F, Colaiacovo M, Haber JE. (1997). Role of *Saccharomyces cerevisiae* Msh2 and Msh3 repair proteins in double-strand break-induced recombination. Proc Natl Acad Sci U S A **94:**9214-9219.
252. Elliott B, Richardson C, Jasin M. (2005). Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. Mol Cell **17:**885-894.
253. Finn KJ, Li JJ. (2013). Single-stranded annealing induced by re-initiation of replication origins provides a novel and efficient mechanism for generating copy number expansion via non-allelic homologous recombination. PLoS Genet **9:**e1003192.
254. Bennett RJ, Forche A, Berman J. (2014). Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. Cold Spring Harb Perspect Med **4**.
255. Chibana H, Beckerman JL, Magee PT. (2000). Fine-resolution physical mapping of genomic diversity in *Candida albicans*. Genome Res **10:**1865-1877.
256. Enloe B, Diamond A, Mitchell AP. (2000). A single-transformation gene function test in diploid *Candida albicans*. J Bacteriol **182:**5730-5736.
257. Montelone BA, Hoekstra MF, Malone RE. (1988). Spontaneous mitotic recombination in yeast: the hyper-recombinational rem1 mutations are alleles of the *RAD3* gene. Genetics **119:**289-301.
258. Iwaguchi SI, Kanbe T, Tohne T, Magee PT, Suzuki T. (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.
259. Iwaguchi SI, Sato M, Magee BB, Magee PT, Makimura K, Suzuki T. (2001). Extensive chromosome translocation in a clinical isolate showing the distinctive carbohydrate assimilation profile from a candidiasis patient. Yeast **18:**1035-1046.
260. Magee BB, Sanchez MD, Saunders D, Harris D, Berriman M, Magee PT. (2008). Extensive chromosome rearrangements distinguish the karyotype of the hypovirulent species *Candida dubliniensis* from the virulent *Candida albicans*. Fungal Genet Biol **45:**338-350.
261. Polakova S, Blume C, Zarate JA, Mentel M, Jorck-Ramberg D, Stenderup J, Piskur J. (2009). Formation of new chromosomes as a virulence mechanism in yeast *Candida glabrata*. Proc Natl Acad Sci U S A **106:**2688-2693.
262. Berman J. (2016). Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. FEMS Yeast Res **16**.
263. Forche A. (2014). Large-Scale Chromosomal Changes and Associated Fitness Consequences in Pathogenic Fungi. Curr Fungal Infect Rep **8:**163-170.

264. Wilson RB, Davis D, Mitchell AP. (1999). Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. J Bacteriol **181:**1868-1874.
265. Negredo A, Monteoliva L, Gil C, Pla J, Nombela C. (1997). Cloning, analysis and one-step disruption of the *ARG5,6* gene of *Candida albicans*. Microbiology **143 (Pt 2):**297-302.
266. 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.
267. Selmecki A, Bergmann S, Berman J. (2005). Comparative genome hybridization reveals widespread aneuploidy in *Candida albicans* laboratory strains. Mol Microbiol **55:**1553-1565.
268. Fonzi WA, Irwin MY. (1993). Isogenic strain construction and gene mapping in *Candida albicans*. Genetics **134:**717-728.
269. Yoshida J, Umezu K, Maki H. (2003). Positive and negative roles of homologous recombination in the maintenance of genome stability in *Saccharomyces cerevisiae*. Genetics **164:**31-46.
270. Selmecki A, Gerami-Nejad M, Paulson C, Forche A, Berman J. (2008). An isochromosome confers drug resistance *in vivo* by amplification of two genes, *ERG11* and *TAC1*. Mol Microbiol **68:**624-641.
271. Marr KA, Lyons CN, Rustad T, Bowden RA, White TC. (1998). Rapid, transient fluconazole resistance in *Candida albicans* is associated with increased mRNA levels of *CDR*. Antimicrob Agents and Chemother **42:**2584-2589.
272. Rustchenko EP, Howard DH, Sherman F. (1994). Chromosomal alterations of *Candida albicans* are associated with the gain and loss of assimilating functions. J Bacteriol **176:**3231-3241.
273. Wu W, Pujol C, Lockhart SR, Soll DR. (2005). Chromosome loss followed by duplication is the major mechanism of spontaneous mating-type locus homozygosis in *Candida albicans*. Genetics **169:**1311-1327.
274. Bouchonville K, Forche A, Tang KE, Selmecki A, Berman J. (2009). Aneuploid chromosomes are highly unstable during DNA transformation of *Candida albicans*. Eukaryot Cell **8:**1554-1566.
275. Covo S, Puccia CM, Argueso JL, Gordenin DA, Resnick MA. (2014). The sister chromatid cohesion pathway suppresses multiple chromosome gain and chromosome amplification. Genetics **196:**373-384.
276. Chen G, Bradford WD, Seidel CW, Li R. (2012). Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. Nature **482:**246-250.
277. Harrison BD, Hashemi J, Bibi M, Pulver R, Bavli D, Nahmias Y, Wellington M, Sapiro G, Berman J. (2014). A tetraploid intermediate precedes aneuploid formation in yeasts exposed to fluconazole. PLoS Biol **12:**e1001815.
278. Chang FM, Ou TY, Cheng WN, Chou ML, Lee KC, *et al*. (2014). Short-term exposure to fluconazole induces chromosome loss in *Candida albicans*: an approach to produce haploid cells. Fungal Genet Biol **70:**68-76.
279. Diogo D, Bouchier C, d'Enfert C, Bougnoux ME. (2009). Loss of heterozygosity in commensal isolates of the asexual diploid yeast *Candida albicans*. Fungal Genet Biol **46:**159-168.
280. Magee BB, Magee PT. (1997). WO-2, a stable aneuploid derivative of *Candida albicans* strain WO-1, can switch from white to opaque and form hyphae. Microbiology **143 (Pt 2):**289-295.
281. Selmecki A, Forche A, Berman J. (2006). Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. Science **313:**367-370.

282. Brown GD, Denning DW, Gow NA, Levitz SM, Netea MG, White TC. (2012). Hidden killers: human fungal infections. Sci Transl Med **4:**165rv113.
283. Hill JA, Ammar R, Torti D, Nislow C, Cowen LE. (2013). Genetic and genomic architecture of the evolution of resistance to antifungal drug combinations. PLoS Genet **9:**e1003390.
284. Asakura K, Iwaguchi S, Homma M, Sukai T, Higashide K, Tanaka K. (1991). Electrophoretic karyotypes of clinically isolated yeasts of *Candida albicans* and *C. glabrata*. J Gen Microbiol **137:**2531-2538.
285. Chen G, Mulla WA, Kucharavy A, Tsai HJ, Rubinstein B, *et al*. (2015). Targeting the adaptability of heterogeneous aneuploids. Cell **160:**771-784.
286. 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.
287. Blaikley EJ, Tinline-Purvis H, Kasparek TR, Marguerat S, Sarkar S, *et al*. (2014). The DNA damage checkpoint pathway promotes extensive resection and nucleotide synthesis to facilitate homologous recombination repair and genome stability in fission yeast. Nucleic Acids Res **42:**5644-5656.
288. Tinline-Purvis H, Savory AP, Cullen JK, Dave A, Moss J, *et al*. (2009). Failed gene conversion leads to extensive end processing and chromosomal rearrangements in fission yeast. EMBO J **28:**3400-3412.
289. Nakamura K, Okamoto A, Katou Y, Yadani C, Shitanda T, *et al*. (2008). Rad51 suppresses gross chromosomal rearrangement at centromere in Schizosaccharomyces pombe. EMBO J **27:**3036-3046.
290. Wolff DJ, Miller AP, Van Dyke DL, Schwartz S, Willard HF. (1996). Molecular definition of breakpoints associated with human Xq isochromosomes: implications for mechanisms of formation. Am J Hum Genet **58:**154-160.
291. Murmann AE, Conrad DF, Mashek H, Curtis CA, Nicolae RI, Ober C, Schwartz S. (2009). Inverted duplications on acentric markers: mechanism of formation. Hum Mol Genet **18:**2241-2256.
292. Coste A, Turner V, Ischer F, Morschhauser J, Forche A, Selmecki A, Berman J, Bille J, Sanglard D. (2006). A mutation in Tac1p, a transcription factor regulating *CDR1* and *CDR2*, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in *Candida albicans*. Genetics **172:**2139-2156.
293. Dunkel N, Blass J, Rogers PD, Morschhauser J. (2008). Mutations in the multi-drug resistance regulator *MRR1*, followed by loss of heterozygosity, are the main cause of *MDR1* overexpression in fluconazole-resistant *Candida albicans* strains. Mol Microbiol **69:**827-840.
294. Dunkel N, Liu TT, Barker KS, Homayouni R, Morschhauser J, Rogers PD. (2008). A gain-of-function mutation in the transcription factor Upc2p causes upregulation of ergosterol biosynthesis genes and increased fluconazole resistance in a clinical *Candida albicans* isolate. Eukaryot Cell **7:**1180-1190.
295. White TC. (1997). The presence of an R467K amino acid substitution and loss of allelic variation correlate with an azole-resistant lanosterol 14alpha demethylase in *Candida albicans*. Antimicrob Agents Chemother **41:**1488-1494.
296. Heilmann CJ, Schneider S, Barker KS, Rogers PD, Morschhauser J. (2010). An A643T mutation in the transcription factor Upc2p causes constitutive *ERG11* upregulation and increased fluconazole resistance in *Candida albicans*. Antimicrob Agents Chemother **54:**353-359.
297. Manoharlal R, Gorantala J, Sharma M, Sanglard D, Prasad R. (2010). *PAP1* [poly(A) polymerase 1] homozygosity and hyperadenylation are major determinants of increased

mRNA stability of *CDR1* in azole-resistant clinical isolates of *Candida albicans*. Microbiology **156:**313-326.
298. Dodgson AR, Dodgson KJ, Pujol C, Pfaller MA, Soll DR. (2004). Clade-specific flucytosine resistance is due to a single nucleotide change in the *FUR1* gene of *Candida albicans*. Antimicrob Agents Chemother **48:**2223-2227.
299. Niimi K, Monk BC, Hirai A, Hatakenaka K, Umeyama T, *et al*. (2010). Clinically significant micafungin resistance in *Candida albicans* involves modification of a glucan synthase catalytic subunit *GSC1* (*FKS1*) allele followed by loss of heterozygosity. J Antimicrob Chemother **65:**842-852.
300. 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.
301. Anderson MZ, Wigen LJ, Burrack LS, Berman J. (2015). Real-Time Evolution of a Subtelomeric Gene Family in *Candida albicans*. Genetics **200:**907-919.
302. Chibana H, Magee BB, Grindle S, Ran Y, Scherer S, Magee PT. (1998). A physical map of chromosome 7 of *Candida albicans*. Genetics **149:**1739-1752.
303. Iwaguchi S, Homma M, Tanaka K. (1992). Clonal variation of chromosome size derived from the rDNA cluster region in *Candida albicans*. J Gen Microbiol **138:**1177-1184.
304. Rustchenko EP, Curran TM, Sherman F. (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:**7189-7199.
305. Koç A, Wheeler LJ, Mathews CK, Merrill GF. (2003). Hydroxyurea arrests DNA replication by a mechanism that preserves basal dNTP pools. J Biol Chem **279**.
306. Chen H, Ma Z, Vanderwaal RP, Feng Z, Gonzalez-Suarez I, *et al*. (2011). The mTOR inhibitor rapamycin suppresses DNA double-strand break repair. Radiat Res **175:**214-224.
307. Iqbal ZM, Kohn KW, Ewig RA, Fornace AJ, Jr. (1976). Single-strand scission and repair of DNA in mammalian cells by bleomycin. Cancer Res **36:**3834-3838.
308. Ananthaswamy HN, Eisenstark A. (1977). Repair of hydrogen peroxide-induced single-strand breaks in *Escherichia coli* deoxyribonucleic acid. J Bacteriol **130:**187-191.
309. Schwartz TR, Kmiec EB. (2005). Using methyl methanesulfonate (MMS) to stimulate targeted gene repair activity in mammalian cells. Gene Ther Mol Bio **9:**193-202.
310. Klimek M. (1965). Formation but no excision of thymine dimers in mammalian cells after UV-irradiation. Neoplasma **12:**459-460.
311. Kim H, Kim JS. (2014). A guide to genome engineering with programmable nucleases. Nat Rev Genet **15:**321-334.
312. Belfort M, Roberts J. (1997). Homing endonucleases: keeping the house in order. Nucleic Acids Res **25:**3379-3388.
313. Silva GH, Belfort M, Wende W, Pingoud A. (2006). From monomeric to homodimeric endonucleases and back: engineering novel specificity of LAGLIDADG enzymes. J Mol Biol **361:**744-754.
314. Smith J, Grizot S, Arnould S, Duclert A, Epinat JC, *et al*. (2006). A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. Nucleic Acids Res **34:**e149.
315. Daboussi F, Zaslavskiy M, Poirot L, Loperfido M, Gouble A, *et al*. (2012). Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. Nucleic Acids Res **40:**6367-6379.
316. Jacquier A, Dujon B. (1985). An Intron-Encoded Protein Is Active in a Gene Conversion Process That Spreads an Intron into a Mitochondrial Gene. Cell **41:**383-394.

317. Monteilhet C, Perrin A, Thierry A, Colleaux L, Dujon B. (1990). Purification and characterization of the in vitro activity of I-*Sce*I, a novel and highly specific endonuclease encoded by a group I intron. Nucleic Acids Res **18:**1407-1413.
318. Fairhead C, Dujon B. (1993). Consequences of unique double-stranded breaks in yeast chromosomes: death or homozygosis. Mol Gen Genet **240:**170-178.
319. Fairhead C, Thierry A, Denis F, Eck M, Dujon B. (1998). 'Mass-murder' of ORFs from three regions of chromosome XI from *Saccharomyces cerevisiae*. Gene **223:**33-46.
320. Bellaiche Y, Mogila V, Perrimon N. (1999). I-*Sce*I endonuclease, a new tool for studying DNA double-strand break repair mechanisms in *Drosophila*. Genetics **152:**1037-1044.
321. Rolloos M, Hooykaas PJ, van der Zaal BJ. (2015). Enhanced targeted integration mediated by translocated I-*Sce*I during the *Agrobacterium* mediated transformation of yeast. Sci Rep **5:**8345.
322. Kim YG, Chandrasegaran S. (1994). Chimeric restriction endonuclease. Proc Natl Acad Sci U S A **91:**883-887.
323. Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, Chandrasegaran S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. Mol Cell Biol **21:**289-297.
324. Miller JC, Holmes MC, Wang J, Guschin DY, Lee YL, *et al*. (2007). An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol **25:**778-785.
325. Bozas A, Beumer KJ, Trautman JK, Carroll D. (2009). Genetic analysis of zinc-finger nuclease-induced gene targeting in Drosophila. Genetics **182:**641-651.
326. Tebas P, Stein D, Tang WW, Frank I, Wang SQ, *et al*. (2014). Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. N Engl J Med **370:**901-910.
327. Ramirez CL, Foley JE, Wright DA, Muller-Lerch F, Rahman SH, *et al*. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. Nat Methods **5:**374-375.
328. Kay S, Hahn S, Marois E, Hause G, Bonas U. (2007). A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science **318:**648-651.
329. Li L, Atef A, Piatek A, Ali Z, Piatek M, *et al*. (2013). Characterization and DNA-binding specificities of Ralstonia TAL-like effectors. Mol Plant **6:**1318-1330.
330. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. Science **326:**1509-1512.
331. Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. (2010). Targeting DNA double-strand breaks with TAL effector nucleases. Genetics **186:**757-761.
332. Carroll D. (2014). Genome engineering with targetable nucleases. Annu Rev Biochem **83:**409-439.
333. Mussolino C, Morbitzer R, Lutge F, Dannemann N, Lahaye T, Cathomen T. (2011). A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res **39:**9283-9293.
334. Kim Y, Kweon J, Kim A, Chon JK, Yoo JY, *et al*. (2013). A library of TAL effector nucleases spanning the human genome. Nat Biotechnol **31:**251-258.
335. Mashimo T, Kaneko T, Sakuma T, Kobayashi J, Kunihiro Y, Voigt B, Yamamoto T, Serikawa T. (2013). Efficient gene targeting by TAL effector nucleases coinjected with exonucleases in zygotes. Sci Rep **3:**1253.
336. Boissel S, Jarjour J, Astrakhan A, Adey A, Gouble A, *et al*. (2014). megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. Nucleic Acids Res **42:**2591-2601.

337. Kleinstiver BP, Wang L, Wolfs JM, Kolaczyk T, McDowell B, Wang X, Schild-Poulter C, Bogdanove AJ, Edgell DR. (2014). The I-*Tev*I nuclease and linker domains contribute to the specificity of monomeric TALENs. G3 (Bethesda) **4:**1155-1165.
338. Lin J, Chen H, Luo L, Lai Y, Xie W, Kee K. (2015). Creating a monomeric endonuclease TALE-I-*Sce*I with high specificity and low genotoxicity in human cells. Nucleic Acids Res **43:**1112-1122.
339. Frederickson RM. (2015). A New Era of Innovation for CAR T-cell Therapy. Mol Ther **23:**1795-1796.
340. Jacoby E, Yang YM, Qin HY, Chien CD, Kochenderfer JN, Fry TJ. (2016). Murine allogeneic CD19 CAR T cells harbor potent antileukemic activity but have the potential to mediate lethal GVHD. Blood **127:**1361-1370.
341. Labrie SJ, Samson JE, Moineau S. (2010). Bacteriophage resistance mechanisms. Nat Rev Microbiol **8:**317-327.
342. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli,* and identification of the gene product. J Bacteriol **169:**5429-5433.
343. Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151:**2551-2561.
344. Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J Mol Evol **60:**174-182.
345. 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.
346. 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.
347. Marraffini LA, Sontheimer EJ. (2009). Invasive DNA, chopped and in the CRISPR. Structure **17:**786-788.
348. Bhaya D, Davison M, Barrangou R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet **45:**273-297.
349. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature **471:**602-607.
350. Anders C, Niewoehner O, Duerst A, Jinek M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature **513:**569-573.
351. Richard GF. (2015). Shortening trinucleotide repeats using highly specific endonucleases: a possible approach to gene therapy? Trends Genet **31:**177-186.
352. Tycko J, Myer VE, Hsu PD. (2016). Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. Mol Cell **63:**355-370.
353. Ran FA, Hsu PD, Lin CY, Gootenberg JS, Konermann S, *et al*. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell **154:**1380-1389.
354. Nelles DA, Fang MY, O'Connell MR, Xu JL, Markmiller SJ, Doudna JA, Yeo GW. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. Cell **165:**488-496.
355. Vyas VK, Barrasa MI, Fink GR. (2015). A CRISPR system permits genetic engineering of essential genes and gene families. Sci Adv **1:**e1500248.

356. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, *et al*. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell **163:**759-771.
357. Gao F, Shen XZ, Jiang F, Wu Y, Han C. (2016). DNA-guided genome editing using the *Natronobacterium gregoryi* Argonaute. Nat Biotechnol doi:10.1038/nbt.3547.
358. Makarova KS, Wolf YI, van der Oost J, Koonin EV. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct **4:**29.
359. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA. (2013). Bacterial argonaute samples the transcriptome to identify foreign DNA. Mol Cell **51:**594-605.
360. Strathern JN, Klar AJS, Hicks JB, Abraham JA, Ivy JM, Nasmyth KA, McGill C. (1982). Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the *MAT* locus. Cell **31:**183-192.
361. Yim E, O'Connell KE, St Charles J, Petes TD. (2014). High-resolution mapping of two types of spontaneous mitotic gene conversion events in *Saccharomyces cerevisiae*. Genetics **198:**181-192.
362. Lee CS, Wang RW, Chang HH, Capurso D, Segal MR, Haber JE. (2016). Chromosome position determines the success of double-strand break repair. Proc Natl Acad Sci U S A **113:**E146-154.
363. Anand RP, Tsaponina O, Greenwell PW, Lee CS, Du W, Petes TD, Haber JE. (2014). Chromosome rearrangements via template switching between diverged repeated sequences. Genes Dev **28:**2394-2406.
364. Forche A, May G, Beckerman J, Kauffman S, Becker J, Magee PT. (2003). A system for studying genetic changes in *Candida albicans* during infection. Fungal Genet Biol **39:**38-50.
365. Forche A, Steinbach M, Berman J. (2009). Efficient and Rapid identification of *C. albicans* allelic status using SNP-RFLP. FEMS Yeast Res **9:**1061-1069.
366. Haase SB, Lew DJ. (1997). Flow cytometric analysis of DNA content in budding yeast. Methods Enzymol **283:**322-332.
367. Sanséau P, Tiffoche C, Kahloun AE, Collin O, Rolland JP, Le Pennec JP. (1990). Increased resolution of large DNA fragments by a two dimensional pulse field gel electrophoresis. Anal Biochem **189:**142-148.
368. Vollrath D, Davis RW. (1987). Resolution of DNA molecules greater than 5 megabases by contour-clamped homogeneous electric fields. Nucleic Acids Res **15:**7865-7876.
369. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. (1992). Comparative Genomic Hybridization for Molecular Cytogenetic Analysis of Solid Tumors. Science **258:**818-821.
370. Pollack JR, Perou CM, Sorlie T, Alizadeh AA, Eisen MB, *et al*. (1999). Genome-wide analysis of DNA copy number variation in breast cancer using DNA microarrays. Nat Genet **23:**41-46.
371. Miyazaki T, Bressan DA, Shinohara M, Haber JE, Shinohara A. (2004). In vivo assembly and disassembly of Rad51 and Rad52 complexes during double-strand break repair. EMBO J **23:**939-949.
372. Giaever G, Chu AM, Ni L, Connelly C, Riles L, *et al*. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature **418:**387-391.
373. Kumar A, Seringhaus M, Biery MC, Sarnovsky RJ, Umansky L, *et al*. (2004). Large-scale mutagenesis of the yeast genome using a Tn7-derived multipurpose transposon. Genome Res **14:**1975-1986.

374. Li Z, Vizeacoumar FJ, Bahr S, Li J, Warringer J, *et al*. (2011). Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat Biotechnol **29:**361-367.
375. Mnaimneh S, Davierwala AP, Haynes J, Moffat J, Peng WT, *et al*. (2004). Exploration of essential gene functions via titratable promoter alleles. Cell **118:**31-44.
376. Douglas AC, Smith AM, Sharifpoor S, Yan Z, Durbic T, *et al*. (2012). Functional analysis with a barcoder yeast gene overexpression system. G3 (Bethesda) **2:**1279-1289.
377. Gelperin DM, White MA, Wilkinson ML, Kon Y, Kung LA, *et al*. (2005). Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. Genes Dev **19:**2816-2826.
378. Jones GM, Stalker J, Humphray S, West A, Cox T, Rogers J, Dunham I, Prelich G. (2008). A systematic library for comprehensive overexpression screens in *Saccharomyces cerevisiae*. Nat Methods **5:**239-241.
379. Norman KL, Kumar A. (2016). Mutant power: using mutant allele collections for yeast functional genomics. Brief Funct Genomics **15:**75-84.
380. Huang RY, Eddy M, Vujcic M, Kowalski D. (2005). Genome-wide screen identifies genes whose inactivation confer resistance to cisplatin in *Saccharomyces cerevisiae*. Cancer Res **65:**5890-5897.
381. Andersen MP, Nelson ZW, Hetrick ED, Gottschling DE. (2008). A genetic screen for increased loss of heterozygosity in *Saccharomyces cerevisiae*. Genetics **179:**1179-1195.
382. Kitagawa M, Lee SH. (2015). The chromosomal passenger complex (CPC) as a key orchestrator of orderly mitotic exit and cytokinesis. Front Cell Dev Biol **3:**14.
383. Bian Y, Kitagawa R, Bansal PK, Fujii Y, Stepanov A, Kitagawa K. (2014). Synthetic genetic array screen identifies PP2A as a therapeutic target in Mad2-overexpressing tumors. Proc Natl Acad Sci U S A **111:**1628-1633.
384. Li R, Murray AW. (1991). Feedback control of mitosis in budding yeast. Cell **66:**519-531.
385. Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JK, Markowitz SD, Kinzler KW, Vogelstein B. (1998). Mutations of mitotic checkpoint genes in human cancers. Nature **392:**300-303.
386. Sotillo R, Hernando E, Diaz-Rodriguez E, Teruya-Feldstein J, Cordon-Cardo C, Lowe SW, Benezra R. (2007). Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. Cancer Cell **11:**9-23.
387. Jones S, Wang TL, Shih Ie M, Mao TL, Nakayama K, *et al*. (2010). Frequent mutations of chromatin remodeling gene *ARID1A* in ovarian clear cell carcinoma. Science **330:**228-231.
388. Noble SM, French S, Kohn LA, Chen V, Johnson AD. (2010). Systematic screens of a *Candida albicans* homozygous deletion library decouple morphogenetic switching and pathogenicity. Nat Genet **42:**590-598.
389. Richard ML, Nobile CJ, Bruno VM, Mitchell AP. (2005). *Candida albicans* biofilm-defective mutants. Eukaryot Cell **4:**1493-1502.
390. Becker JM, Kauffman SJ, Hauser M, Huang L, Lin M, Sillaots S, Jiang B, Xu D, Roemer T. (2010). Pathway analysis of *Candida albicans* survival and virulence determinants in a murine infection model. Proc Natl Acad Sci U S A **107:**22044-22049.
391. Vandeputte P, Ischer F, Sanglard D, Coste AT. (2011). In vivo systematic analysis of *Candida albicans* Zn2-Cys6 transcription factors mutants for mice organ colonization. PLoS One **6:**e26962.
392. Finkel JS, Yudanin N, Nett JE, Andes DR, Mitchell AP. (2011). Application of the systematic "DAmP" approach to create a partially defective *C. albicans* mutant. Fungal Genet Biol **48:**1056-1061.

393. Park YN, Morschhauser J. (2005). Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. Eukaryot Cell **4:**1328-1342.
394. Roemer T, Jiang B, Davison J, Ketela T, Veillette K, *et al*. (2003). Large-scale essential gene identification in *Candida albicans* and applications to antifungal drug discovery. Mol Microbiol **50:**167-181.
395. Uhl MA, Biery M, Craig N, Johnson AD. (2003). Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen *C.albicans*. EMBO J **22:**2668-2678.
396. Bharucha N, Chabrier-Rosello Y, Xu T, Johnson C, Sobczynski S, *et al*. (2011). A large-scale complex haploinsufficiency-based genetic interaction screen in *Candida albicans*: analysis of the RAM network during morphogenesis. PLoS Genet **7:**e1002058.
397. 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.
398. Fu Y, Luo G, Spellberg BJ, Edwards JE, Jr., Ibrahim AS. (2008). Gene overexpression/suppression analysis of candidate virulence factors of *Candida albicans*. Eukaryot Cell **7:**483-492.
399. Nobile CJ, Fox EP, Nett JE, Sorrells TR, Mitrovich QM, Hernday AD, Tuch BB, Andes DR, Johnson AD. (2012). A recently evolved transcriptional network controls biofilm development in *Candida albicans*. Cell **148:**126-138.
400. Nobile CJ, Mitchell AP. (2006). Genetics and genomics of *Candida albicans* biofilm formation. Cell Microbiol **8:**1382-1391.
401. Nobile CJ, Nett JE, Hernday AD, Homann OR, Deneault JS, Nantel A, Andes DR, Johnson AD, Mitchell AP. (2009). Biofilm matrix regulation by *Candida albicans* Zap1. PLoS Biol **7:**e1000133.
402. Chauvel M, Nesseir A, Cabral V, Znaidi S, Goyard S, *et al*. (2012). A versatile overexpression strategy in the pathogenic yeast *Candida albicans*: identification of regulators of morphogenesis and fitness. PLoS One **7:**e45912.
403. Cabral V, Znaidi S, Walker LA, Martin-Yken H, Dague E, *et al*. (2014). Targeted changes of the cell wall proteome influence *Candida albicans* ability to form single- and multi-strain biofilms. PLoS Pathog **10:**e1004542.
404. Znaidi S, Nesseir A, Chauvel M, Rossignol T, d'Enfert C. (2013). A comprehensive functional portrait of two heat shock factor-type transcriptional regulators involved in *Candida albicans* morphogenesis and virulence. PLoS Pathog **9:**e1003519.
405. Zhu X, Dunn JM, Goddard AD, Squire JA, Becker A, Phillips RA, Gallie BL. (1992). Mechanisms of loss of heterozygosity in retinoblastoma. Cytogenet Cell Genet **59:**248-252.
406. Chiurazzi M, Ray A, Viret JF, Perera R, Wang XH, Lloyd AM, Signer ER. (1996). Enhancement of somatic intrachromosomal homologous recombination in Arabidopsis by the HO endonuclease. Plant Cell **8:**2057-2066.
407. Malkova A, Klein F, Leung WY, Haber JE. (2000). HO endonuclease-induced recombination in yeast meiosis resembles Spo11-induced events. Proc Natl Acad Sci U S A **97:**14500-14505.
408. Aubert M, Ryu BY, Banks L, Rawlings DJ, Scharenberg AM, Jerome KR. (2011). Successful Targeting and Disruption of an Integrated Reporter Lentivirus Using the Engineered Homing Endonuclease Y2 I-*Ani*I. PLoS One **6**.
409. Rempe D, Vangeison G, Hamilton J, Li Y, Jepson M, Federoff HJ. (2006). Synapsin I Cre transgene expression in male mice produces germline recombination in progeny. Genesis **44:**44-49.

410. Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, *et al*. (2011). A synthetic homing endonuclease-based gene drive system in the human malaria mosquito. Nature **473:**212-215.
411. Hill JA, Ammar R, Torti D, Nislow C, Cowen LE. (2013). Genetic and genomic architecture of the evolution of resistance to antifungal drug combinations. PLoS Genet **9:**e1003390.
412. Coste A, Selmecki A, Forche A, Diogo D, Bougnoux ME, d'Enfert C, Berman J, Sanglard D. (2007). Genotypic evolution of azole resistance mechanisms in sequential *Candida albicans* isolates. Eukaryot Cell **6:**1889-1904.
413. 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.
414. Lockhart SR, Pujol C, Daniels KJ, Miller MG, Johnson AD, Pfaller MA, Soll DR. (2002). In *Candida albicans*, white-opaque switchers are homozygous for mating type. Genetics **162:**737-745.
415. Brenner DJ, Ward JF. (1992). Constraints on energy deposition and target size of multiply damaged sites associated with DNA double-strand breaks. Int J Radiat Biol **61:**737-748.
416. Tscherner M, Stappler E, Hnisz D, Kuchler K. (2012). The histone acetyltransferase Hat1 facilitates DNA damage repair and morphogenesis in *Candida albicans*. Mol Microbiol **86:**1197-1214.
417. Sarachek A, Henderson LA, Eddy KB. (1991). Genetic destabilization of *Candida albicans* by hydroxyurea. Microbios **65:**39-61.
418. Whelan WL, Markie D. (1985). UV-induced instability in *Candida albicans* hybrids. Curr Genet **9:**175-177.
419. Berkovich E, Monnat RJ, Kastan MB. (2008). Assessment of protein dynamics and DNA repair following generation of DNA double-strand breaks at defined genomic sites. Nat Protoc **3:**915-922.
420. Johnson RD, Jasin M. (2001). Double-strand-break-induced homologous recombination in mammalian cells. Biochem Soc Trans **29:**196-201.
421. Lyznik LA, Djukanovic V, Yang M, Jones S. (2012). Double-strand break-induced targeted mutagenesis in plants. Methods Mol Biol **847:**399-416.
422. Munoz NM, Beard BC, Ryu BY, Luche RM, Trobridge GD, Rawlings DJ, Scharenberg AM, Kiem HP. (2012). Novel reporter systems for facile evaluation of I-*Sce*I-mediated genome editing. Nucleic Acids Res **40:**e14.
423. Kalderon D, Richardson WD, Markham AF, Smith AE. (1984). Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature **311:**33-38.
424. Collas P, Alestrom P. (1998). Nuclear localization signals enhance germline transmission of a transgene in zebrafish. Transgenic Res **7:**303-309.
425. Wang Y, Zhou XY, Xiang PY, Wang LL, Tang H, Xie F, Li L, Wei H. (2014). The meganuclease I-*Sce*I containing nuclear localization signal (NLS-I-*Sce*I) efficiently mediated mammalian germline transgenesis via embryo cytoplasmic microinjection. PLoS One **9:**e108347.
426. Nobile CJ, Mitchell AP. (2009). Large-scale gene disruption using the UAU1 cassette. Methods Mol Biol **499:**175-194.
427. Towpik J, Chacinska A, Ciesla M, Ginalski K, Boguta M. (2004). Mutations in the yeast *MRF1* gene encoding mitochondrial release factor inhibit translation on mitochondrial ribosomes. J Biol Chem **279:**14096-14103.

428. Askarian-Amiri ME, Pel HJ, Guevremont D, McCaughan KK, Poole ES, Sumpter VG, Tate WP. (2000). Functional characterization of yeast mitochondrial release factor 1. J Biol Chem **275:**17241-17248.
429. Pel HJ, Rozenfeld S, BolotinFukuhara M. (1996). The nuclear *Kluyveromyces lactis MRF1* gene encodes a mitochondrial class I peptide chain release factor that is important for cell viability. Current Genetics **30:**19-28.
430. Karababa M, Coste AT, Rognon B, Bille J, Sanglard D. (2004). Comparison of gene expression profiles of *Candida albicans* azole-resistant clinical isolates and laboratory strains exposed to drugs inducing multidrug transporters. Antimicrob Agents Chemother **48:**3064-3079.
431. Greenberg JR, Price NP, Oliver RP, Sherman F, Rustchenko E. (2005). *Candida albicans SOU1* encodes a sorbose reductase required for L-sorbose utilization. Yeast **22:**957-969.
432. Kabir MA, Ahmad A, Greenberg JR, Wang YK, Rustchenko E. (2005). Loss and gain of chromosome 5 controls growth of *Candida albicans* on sorbose due to dispersed redundant negative regulators. Proc Natl Acad Sci U S A **102:**12147-12152.
433. Lebeaupin T, Sellou H, Timinszky G, Huet S. (2015). Chromatin dynamics at DNA breaks: what, how and why? AIMS Biophysics **2:**458-475.
434. Lemaitre C, Grabarz A, Tsouroula K, Andronov L, Furst A, *et al*. (2014). Nuclear position dictates DNA repair pathway choice. Genes Dev **28:**2450-2463.
435. Hartlerode A, Odate S, Shim I, Brown J, Scully R. (2011). Cell cycle-dependent induction of homologous recombination by a tightly regulated I-*Sce*I fusion protein. PLoS One **6:**e16501.
436. Wei DS, Rong YS. (2007). A genetic screen for DNA double-strand break repair mutations in Drosophila. Genetics **177:**63-77.
437. Vu GT, Cao HX, Watanabe K, Hensel G, Blattner FR, Kumlehn J, Schubert I. (2014). Repair of Site-Specific DNA Double-Strand Breaks in Barley Occurs via Diverse Pathways Primarily Involving the Sister Chromatid. Plant Cell **26:**2156-2167.
438. Esposito MS, Ramirez RM, Bruschi CV. (1994). Nonrandomly-associated forward mutation and mitotic recombination yield yeast diploids homozygous for recessive mutations. Curr Genet **26:**302-307.
439. Haber JE, Hearn M. (1985). Rad52-independent mitotic gene conversion in *Saccharomyces cerevisiae* frequently results in chromosomal loss. Genetics **111:**7-22.
440. Ho CK, Mazon G, Lam AF, Symington LS. (2010). Mus81 and Yen1 promote reciprocal exchange during mitotic recombination to maintain genome integrity in budding yeast. Mol Cell **40:**988-1000.
441. Smith AM, Takeuchi R, Pellenz S, Davis L, Maizels N, Monnat RJ, Stoddard BL. (2009). Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-AniI LAGLIDADG homing endonuclease. Proceedings of the National Academy of Sciences of the United States of America **106:**5099-5104.
442. Wiese C, Pierce AJ, Gauny SS, Jasin M, Kronenberg A. (2002). Gene conversion is strongly induced in human cells by double-strand breaks and is modulated by the expression of BCL-x(L). Cancer Res **62:**1279-1283.
443. Pujol C, Reynes J, Renaud F, Raymond M, Tibayrenc M, Ayala FJ, Janbon F, Mallie M, Bastide JM. (1993). The Yeast *Candida albicans* Has a Clonal Mode of Reproduction in a Population of Infected Human Immunodeficiency Virus-Positive Patients. Proc Natl Acad Sci U S A **90:**9456-9459.
444. Anderson JB, Wickens C, Khan M, Cowen LE, Federspiel N, Jones T, Kohn LM. (2001). Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans*. J Bacteriol **183:**865-872.

445. Graser Y, Volovsek M, Arrington J, Schonian G, Presber W, Mitchell TG, Vilgalys R. (1996). Molecular markers reveal that population structure of the human pathogen *Candida albicans* exhibits both clonality and recombination. Proc Natl Acad Sci U S A **93:**12473-12477.
446. Lockhart SR, Fritch JJ, Meier AS, Schroppel K, Srikantha T, Galask R, Soll DR. (1995). Colonizing populations of *Candida albicans* are clonal in origin but undergo microevolution through C1 fragment reorganization as demonstrated by DNA-fingerprinting and C1 sequencing. J Clin Microbiol **33:**1501-1509.
447. Mata AL, Rosa RT, Rosa EAR, Goncalves RB, Hofling JF. (2000). Clonal variability among oral *Candida albicans* assessed by allozyme electrophoresis analysis. Oral Microbiol Immunol **15:**350-354.
448. Birky CW, Jr. (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual eukaryotes. Genetics **144:**427-437.
449. Gola S, Martin R, Walther A, Dunkler 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.
450. Andaluz E, Gomez-Raja J, Hermosa B, Ciudad T, Rustchenko E, Calderone R, Larriba G. (2007). Loss and fragmentation of chromosome 5 are major events linked to the adaptation of rad52-DeltaDelta strains of *Candida albicans* to sorbose. Fungal Genet Biol **44:**789-798.
451. Janbon G, Sherman F, Rustchenko E. (1999). Appearance and properties of L-sorbose-utilizing mutants of *Candida albicans* obtained on a selective plate. Genetics **153:**653-664.
452. Legrand M, Lephart P, Forche A, Mueller FMC, Walsh T, Magee PT, Magee BB. (2004). Homozygosity at the MTL locus in clinical strains of *Candida albicans*: karyotypic rearrangements and tetraploid formation. Mol Microbiol **52:**1451-1462.
453. Binkley J, Arnaud MB, Inglis DO, Skrzypek MS, Shah P, *et al*. (2014). The Candida Genome Database: the new homology information page highlights protein similarity and phylogeny. Nucleic Acids Res **42:**D711-716.
454. Li H, Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754-1760.
455. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, *et al*. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:**1297-1303.
456. 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:**100.
457. Mundia MM, Desai V, Magwood AC, Baker MD. (2014). Nascent DNA synthesis during homologous recombination is synergistically promoted by the Rad51 recombinase and DNA homology. Genetics **197:**107-119.
458. Shulman MJ, Nissen L, Collins C. (1990). Homologous recombination in hybridoma cells: dependence on time and fragment length. Mol Cell Biol **10:**4466-4472.
459. Hasty P, Rivera-Perez J, Bradley A. (1991). The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol **11:**5586-5591.
460. Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD. (1997). Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. EMBO J **16:**6559-6573.
461. Li F, Dong J, Eichmiller R, Holland C, Minca E, *et al*. (2013). Role of Saw1 in Rad1/Rad10 complex assembly at recombination intermediates in budding yeast. EMBO J **32:**461-472.

462. Fishman-Lobell J, Haber JE. (1992). Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene *RAD1*. Science **258:**480-484.
463. Zierhut C, Diffley JF. (2008). Break dosage, cell cycle stage and DNA replication influence DNA double strand break response. EMBO J **27:**1875-1885.
464. Thierry A, Khanna V, Creno S, Lafontaine I, Ma L, Bouchier C, Dujon B. (2015). Macrotene chromosomes provide insights to a new mechanism of high-order gene amplification in eukaryotes. Nat Commun **6:**6154.
465. Alhourani E, Rincic M, Melo JB, Carreira IM, Glaser A, Pohle B, Schlie C, Liehr T. (2015). Isochromosome 17q in Chronic Lymphocytic Leukemia. Leuk Res Treatment **2015:**489592.
466. Park Y, Lim J, Ko YH, Kim J, Kwon GC, Koo SH. (2015). A case of pentasomy 21 with two isochromosome 21s in acute megakaryoblastic leukemia associated with Down syndrome. Ann Lab Med **35:**373-375.
467. You E, Cho SY, Yang JJ, Lee HJ, Lee WI, Lee J, Cho KS, Cho EH, Park TS. (2015). A novel case of extreme thrombocytosis in acute myeloid leukemia associated with isochromosome 17q and copy neutral loss of heterozygosity. Ann Lab Med **35:**366-369.
468. Kanellis P, Gagliardi M, Banath JP, Szilard RK, Nakada S, *et al*. (2007). A screen for suppressors of gross chromosomal rearrangements identifies a conserved role for PLP in preventing DNA lesions. PLoS Genet **3:**e134.
469. Huang ME, Rio AG, Nicolas A, Kolodner RD. (2003). A genomewide screen in *Saccharomyces cerevisiae* for genes that suppress the accumulation of mutations. Proc Natl Acad Sci U S A **100:**11529-11534.
470. Yuen KW, Warren CD, Chen O, Kwok T, Hieter P, Spencer FA. (2007). Systematic genome instability screens in yeast and their potential relevance to cancer. Proc Natl Acad Sci U S A **104:**3925-3930.
471. Legrand M, Munro CA, d'Enfert C. (2011). Cool Tools 5: The *Candida albicans* ORFeome project. . *Candida and candidiasis* **2nd edition, eds Calderone RA & Clancy CJ (ASM Press, Washington DC)**.
472. Cabral V, Chauvel M, Firon A, Legrand M, Nesseir A, Bachellier-Bassi S, Chaudhari Y, Munro CA, d'Enfert C. (2012). Modular gene over-expression strategies for *Candida albicans*. Methods Mol Biol **845:**227-244.
473. Schaub Y, Dunkler A, Walther A, Wendland J. (2006). New pFA-cassettes for PCR-based gene manipulation in *Candida albicans*. J Basic Microbiol **46:**416-429.
474. Chou H, Glory A, Bachewich C. (2011). Orthologues of the anaphase-promoting complex/cyclosome coactivators Cdc20p and Cdh1p are important for mitotic progression and morphogenesis in *Candida albicans*. Eukaryot Cell **10:**696-709.
475. Stevenson LF, Kennedy BK, Harlow E. (2001). A large-scale overexpression screen in *Saccharomyces cerevisiae* identifies previously uncharacterized cell cycle genes. Proc Natl Acad Sci U S A **98:**3946-3951.
476. Shi QM, Wang YM, Zheng XD, Lee RT, Wang Y. (2007). Critical role of DNA checkpoints in mediating genotoxic-stress-induced filamentous growth in *Candida albicans*. Mol Biol Cell **18:**815-826.
477. Yu L, Pena Castillo L, Mnaimneh S, Hughes TR, Brown GW. (2006). A survey of essential gene function in the yeast cell division cycle. Mol Biol Cell **17:**4736-4747.
478. Szyjka SJ, Aparicio JG, Viggiani CJ, Knott S, Xu W, Tavare S, Aparicio OM. (2008). Rad53 regulates replication fork restart after DNA damage in *Saccharomyces cerevisiae*. Genes Dev **22:**1906-1920.

479. Bennardo N, Cheng A, Huang N, Stark JM. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genet **4:**e1000110.
480. Guirouilh-Barbat J, Huck S, Bertrand P, Pirzio L, Desmaze C, Sabatier L, Lopez BS. (2004). Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. Mol Cell **14:**611-623.
481. Cannon RD, Jenkinson HF, Shepherd MG. (1992). Cloning and expression of *Candida albicans ADE2* and proteinase genes on a replicative plasmid in *C. albicans* and in *Saccharomyces cerevisiae*. Mol Gen Genet **235:**453-457.
482. Kurtz MB, Cortelyou MW, Miller SM, Lai M, Kirsch DR. (1987). Development of autonomously replicating plasmids for *Candida albicans*. Mol Cell Biol **7:**209-217.
483. Hahnenberger KM, Baum MP, Polizzi CM, Carbon J, Clarke L. (1989). Construction of functional artificial minichromosomes in the fission yeast *Schizosaccharomyces pombe*. Proc Natl Acad Sci U S A **86:**577-581.
484. Burke DT, Carle GF, Olson MV. (1987). Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors. Science **236:**806-812.
485. Carlson SR, Rudgers GW, Zieler H, Mach JM, Luo S, Grunden E, Krol C, Copenhaver GP, Preuss D. (2007). Meiotic transmission of an in vitro-assembled autonomous maize minichromosome. PLoS Genet **3:**1965-1974.
486. Guiducci C, Ascenzioni F, Auriche C, Piccolella E, Guerrini AM, Donini P. (1999). Use of a Human Minichromosome as a Cloning and Expression Vector for Mammalian Cells. Hum Mol Genet **8:**1417-1424.
487. Basso LR, Jr., Bartiss A, Mao Y, Gast CE, Coelho PS, Snyder M, Wong B. (2010). Transformation of *Candida albicans* with a synthetic hygromycin B resistance gene. Yeast **27:**1039-1048.
488. Ohama T, Suzuki T, Mori M, Osawa S, Ueda T, Watanabe K, Nakase T. (1993). Non-universal decoding of the leucine codon CUG in several *Candida* species. Nucleic Acids Res **21:**4039-4045.
489. Puigbo P, Guzman E, Romeu A, Garcia-Vallve S. (2007). OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res **35:**W126-131.
490. Schneider J, Schindewolf C, van Zee K, Fanning E. (1988). A mutant SV40 large T antigen interferes with nuclear localization of a heterologous protein. Cell **54:**117-125.
491. Mosammaparast N, Pemberton LF. (2004). Karyopherins: from nuclear-transport mediators to nuclear-function regulators. Trends Cell Biol **14:**547-556.
492. Stynen B, Van Dijck P, Tournu H. (2010). A CUG codon adapted two-hybrid system for the pathogenic fungus *Candida albicans*. Nucleic Acids Res **38:**e184.
493. Shen J, Guo W, Kohler JR. (2005). *CaNAT1*, a heterologous dominant selectable marker for transformation of *Candida albicans* and other pathogenic *Candida* species. Infect Immun **73:**1239-1242.
494. Murad AM, Lee PR, Broadbent ID, Barelle CJ, Brown AJ. (2000). CIp10, an efficient and convenient integrating vector for *Candida albicans*. Yeast **16:**325-327.

Appendix

Appendix 1:

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.

EC
Journals.ASM.org

A FACS-Optimized Screen Identifies Regulators of Genome Stability in *Candida albicans*

Raphaël Loll-Krippleber,[a,b,c*] Adeline Feri,[a,b,c] Marie Nguyen,[d] Corinne Maufrais,[e] Jennifer Yansouni,[f] Christophe d'Enfert,[a,b] Mélanie Legrand[a,b]

Unité Biologie et Pathogénicité Fongiques, Département Mycologie, Institut Pasteur, Paris, France[a]; INRA USC2019, Paris, France[b]; Université Paris Diderot, Sorbonne Paris Cité, Paris, France[c]; Plate-Forme de Cytométrie, Imagopole, Institut Pasteur, Paris, France[d]; Centre d'Informatique en Biologie, Institut Pasteur, Paris, France[e]; BIOASTER, Genomics and Transcriptomics, Lyon, France[f]

Loss of heterozygosity (LOH) plays important roles in genome dynamics, notably, during tumorigenesis. In the fungal pathogen *Candida albicans*, LOH contributes to the acquisition of antifungal resistance. In order to investigate the mechanisms that regulate LOH in *C. albicans*, we have established a novel method combining an artificial heterozygous locus harboring the blue fluorescent protein and green fluorescent protein markers and flow cytometry to detect LOH events at the single-cell level. Using this fluorescence-based method, we have confirmed that elevated temperature, treatment with methyl methanesulfonate, and inactivation of the Mec1 DNA damage checkpoint kinase triggered an increase in the frequency of LOH. Taking advantage of this system, we have searched for *C. albicans* genes whose overexpression triggered an increase in LOH and identified four candidates, some of which are known regulators of genome dynamics with human homologues contributing to cancer progression. Hence, the approach presented here will allow the implementation of new screens to identify genes that are important for genome stability in *C. albicans* and more generally in eukaryotic cells.

Normally found as a harmless commensal organism, *Candida albicans* is also a major fungal pathogen of humans and is capable of causing serious and even life-threatening diseases when the immune system of the host is compromised (1). Although *C. albicans* is found mostly as a diploid organism, haploid and tetraploid forms have been observed in the laboratory and upon the passage of *C. albicans* in animal models of infection (2–4). The formation of tetraploids results from mating between diploids of opposite mating types that have undergone the so-called white-opaque phenotypic switch (5). Meiosis is thought to have been lost in *C. albicans*, and the formation of haploids from diploids or diploids from tetraploids results from concerted chromosome loss (2, 6). The *C. albicans* genome is highly plastic, undergoing a number of important genome rearrangements, such as loss-of-heterozygosity (LOH) events, aneuploidies, and the formation of isochromosomes (7). In particular, LOH has been shown to occur during commensal carriage (8) and during systemic infection in a mouse model (9) and to contribute to the acquisition of antifungal resistance (10, 11). Stressful conditions such as high temperature, oxidative stress, or azole antifungal treatment trigger an increase in the frequency of LOH, as well as changes in the mechanisms leading to LOH; while azole treatment and temperature cause an increase in LOH due to chromosome loss and reduplication, oxidative stress results in an increase in gene conversion events (12). Yet, little is known about the molecular mechanisms that control LOH in *C. albicans*, despite the apparent importance of LOH in the biology of this species. Overall, it has been shown that knockout mutations in genes involved in base excision repair, nucleotide excision repair, and mismatch repair had little impact on the frequency of LOH in *C. albicans* (13, 14). In contrast, null mutations in the *MRE11* and *RAD50* genes involved in homologous recombination and in the *MEC1*, *RAD53*, and *DUN1* genes involved in the DNA damage checkpoint pathway result in increased frequency of LOH (13, 15, 16).

Evaluation of the frequency of LOH in *C. albicans* has until now relied on the use of several counterselectable marker genes, such as *URA3* and *GAL1*, that confer distinctive phenotypes when in a heterozygous state or in a homozygous null state (17–19). *URA3* encodes the orotidine-5-phosphate decarboxylase; while *URA3*/*URA3* and *URA3*/*ura3* strains are prototrophic for uridine and sensitive to 5-fluoroorotic acid (5-FOA), *ura3*/*ura3* strains are auxotrophic for uridine and resistant to 5-FOA. *GAL1* encodes a galactokinase; while *GAL1*/*GAL1* and *GAL1*/*gal1* strains are sensitive to 2-deoxygalactose (2-DG), *gal1*/*gal1* strains are resistant to 2-DG. Hence, counting of the spontaneous 5-FOA-resistant or 2-DG-resistant clones that arise from a *URA3*/*ura3* or a *GAL1*/*gal1* strain, respectively, provides a measure of the frequency of LOH in these strains. Although robust, these systems are labor intensive and costly and lack the flexibility that would be required to implement a genetic screen for *C. albicans* genes regulating LOH. With the development of collections of knockout constructs or overexpression plasmids, such screens are becoming amenable to *C. albicans* (20–22) and could provide new insights into the mechanisms that control genome stability in this species and other eukaryotes.

Received 23 October 2014 **Accepted** 14 January 2015

Accepted manuscript posted online 16 January 2015

Citation 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. doi:10.1128/EC.00286-14.

Address correspondence to Mélanie Legrand, mlegrand@pasteur.fr.

* Present address: Raphaël Loll-Krippleber, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/EC.00286-14.

doi:10.1128/EC.00286-14

TABLE 1 Yeast strains used in this study

Strain	Genotype	Characteristic	Reference
SN100	*his1*Δ/*his1*Δ *URA3*/*ura3*Δ::λ*imm434 IRO1*/*iro1*Δ::λ*imm434*	WT	24
SN148	*arg4*Δ/*arg4*Δ*leu2*Δ/*leu2*Δ *his1*Δ/*his1*Δ*ura3*::λ*imm434*/*ura3*Δ::λ*imm434 iro1*Δ::λ*imm434*/*iro1*Δ::λ*imm434*	WT	24
CEC1027	SN148 *GAL1*/*gal1*::*CmLEU2*	*GAL1* het	This study
CEC3998	SN148 *Ca21chr4_C_albicans_SC5314:473390-476401*Δ::P*TDH3-GFP-CaARG4*/*Ca21chr4_C_albicans_SC5314:473390-476401*/*Ca21chr1_C_albicans_SC5314:625304-626436 RPS1*/*RPS1*::*CIp20-Leu2*	Mono-GFP prototroph	This study
CEC4007	SN148 *Ca21chr4_C_albicans_SC5314:473390-476401*Δ::P*TDH3-BFP-CdHIS*/*Ca21chr4_C_albicans_SC5314:473390-476401*/*Ca21chr1_C_albicans_SC5314:625304-626436 RPS1*/*RPS1*::*CIp30 GAL1*/*gal1*Δ::*CmLEU2*	Mono-BFP prototroph	This study
CEC2683	SN148 *Ca21chr4_C_albicans_SC5314:473390-476401*Δ::P_{TDH3}*-GFP-CaARG4*/*Ca21chr4_C_albicans_SC5314:473390-476401*Δ::P_{TDH3}*-BFP-CaSAT1*	BFP-GFP	This study
CEC2814	CEC1027 *Ca21chr4_C_albicans_SC5314:473390-476401*Δ::P_{TDH3}*-GFP-CaARG4*/*Ca21chr4_C_albicans_SC5314:473390-476401*Δ::P_{TDH3}*-BFP-CdHIS1*	*GAL1 het* BFP-GFP	This study
CEC2824	CEC2814 *ADH1*/*adh1*::P*TDH3-cartTA*::*SAT1*	*GAL1 het* BFP-GFP/pNIMX	This study
CEC3172	CEC2683 *MEC1*/*mec1*Δ::*HIS1*	*mec1*Δ auxotroph	This study
CEC3183	CEC2683 *mec1*Δ::*LEU2*/*mec1*Δ::*HIS1*	*mec1*ΔΔ auxotroph	This study
CEC3194	CEC2683 *RPS1*/*RPS1*::*CIp30*	*MEC1* prototroph	This study
CEC3203	CEC3172 *RPS1*/*RPS1*::*CIp30*	*mec1*Δ prototroph	This study
CEC3216	CEC3183 *RPS1*/*RPS1*::*CIp10*	*mec1*ΔΔ prototroph	This study
CEC3989	CEC2824 *RPS1*/*RPS1*::P*TET-GtwB*	3R collection control	This study

Here, we report the development of a novel LOH reporter system that combines fluorescent markers and flow cytometry to detect LOH in *C. albicans* and is amenable to high-throughput screening. We show that this system can report changes in the frequency of LOH triggered by different physical and chemical stresses and knockout mutations. Moreover, using a newly developed collection of overexpression plasmids for 124 genes whose orthologs in *Saccharomyces cerevisiae* are involved in DNA replication, recombination, and repair, we show that the LOH reporter system can be used to identify genes whose overexpression triggers genome instability specifically through LOH events. Overall, the characterization of new players in genome maintenance will allow a better understanding of how genomic instability may contribute to the success of *C. albicans* as a commensal and as a pathogen.

MATERIALS AND METHODS

Strains and growth conditions. *C. albicans* strains were routinely cultured at 30°C in YPD medium (1% yeast extract, 2% peptone, 2% dextrose), synthetic complete (SC) medium with the omission of the appropriate amino acids (23) or synthetic dextrose (SD) medium (0.67% yeast nitrogen base, 2% dextrose). Solid media were obtained by adding 2% agar. All of the *C. albicans* strains used in this work were derived from SN148 (24) and are listed in Table 1. For selection of transformants, nourseothricin was used at a final concentration of 300 μg/ml in YPD medium.

For the LOH assay by fluorescence-activated cell sorting (FACS), aliquots of prototrophic cells were taken from the −80°C stock and grown in 96-well plates overnight in YPD medium at 30°C with rotatory shaking for recovery. The next day, these cultures were inoculated into SC-His-Arg to eliminate cells that had undergone LOH and lost the blue fluorescent protein (BFP)-*HIS1* or green fluorescent protein (GFP)-*ARG4* cassette as a result of the freezing-thawing step. The SC-His-Arg cultures were grown overnight at 30°C with rotatory shaking. At that point, the cells were ready for possible treatments before LOH analysis by flow cytometry.

Overexpression from P_{TET} was induced by the addition of anhydrotetracycline (ATc, 3 μg/ml; Fisher Bioblock Scientific) in YPD or SD medium at 30°C. Overexpression experiments were carried out in the dark, as ATc is light sensitive.

For heat shock, *C. albicans* strain CEC3989 (Table 1) was grown in YPD medium overnight at 30, 37, or 39°C and allowed to recover overnight in YPD medium before flow cytometry analysis.

For genotoxic stress, CEC3989 was cultured for 30 min in YPD medium plus 0.01, 0.02, or 0.03% methyl methanesulfonate (MMS) at 30°C. The cells were collected, washed with fresh YPD medium, and allowed to recover in rich medium at 30°C overnight before flow cytometry analysis.

Plasmid constructions. For the oligonucleotides used in this work, see Tables S1 and S2 in the supplemental material. For the plasmids used or generated in this work, see Table S3 in the supplemental material. For descriptions of the plasmid constructions, see Text S1 in the supplemental material. They yielded six plasmids harboring (i) the gene for either GFP or BFP placed under the control of the constitutively active promoter of the *C. albicans TDH3* gene (25) and (ii) the *ARG4* (26), *HIS1* (27), *URA3* (26), or *SAT1* (27) transformation marker (see Fig. S1 in the supplemental material).

***C. albicans* strain construction. (i) Multiple heterozygous marker strains.** One allele of the *GAL1* gene was replaced with the *CmLEU2* marker gene by PCR-mediated transformation of SN148 (24). The resulting strain was named CEC1027. The P_{TDH3}*-BFP-HIS1* and P_{TDH3}*-GFP-ARG4* cassettes were PCR amplified from plasmids pECC727-BH (ECC727) and pECC596-GA (ECC596) with oligonucleotides K7 yFP-Fwd and K7 yFP-Rev carrying sequences homologous to the genomic DNA located between *PGA59* and *PGA62* on chromosome 4 (Ch4). The PCR products were sequentially transformed in CEC1027 as described by Gola et al. (26), yielding CEC2814. For activation of the P_{TET} promoter, strain CEC2814 was eventually transformed with SacII- and KpnI-digested pNIMX, which encodes a tetracycline-controlled transactivator (21), yielding CEC2824. Strain CEC2683 was obtained through successive transformation of SN148 (24) with the P_{TDH3}*-BFP-SAT1* and P_{TDH3}*-GFP-ARG4* cassettes PCR amplified from plasmids pECC729-BS (ECC729) and pECC596-GA (ECC596) with oligonucleotides K7 yFP-Fwd and K7 yFP-Rev. CEC2683 was subsequently transformed with StuI-linearized CIp30 (28) to generate CEC3194, the prototrophic parental strain.

A *mec1* null mutant was constructed by replacing both alleles with the *LEU2* and *HIS1* markers by a PCR-based method (29) in CEC2683, yielding first the heterozygous *MEC1*/*mec1* mutant CEC3172 and then the null *mec1*/*mec1* mutant CEC3183. CEC3172 and CEC3183 were then trans-

formed with StuI-linearized CIp30 (28) or CIp10 (30) to generate CEC3203 and CEC3216, the prototrophic *mec1* heterozygous and homozygous mutants.

(ii) *C. albicans* overexpression strains. StuI-digested or I-*Sce*I-digested CIp10-P_{TET}-GTW derivatives were transformed into strain CEC2824 as described by Chauvel et al. (21), resulting in 124 overexpression strains (see Table S2 in the supplemental material) that, together with the CEC3989 control strain derived from CEC2824 transformed with StuI-linearized CIp10-P_{TET}-GTW, will be referred to as the 3R overexpression collection.

All *C. albicans* transformants were verified by PCR with primers inside the transformed plasmid/cassette and primers in the genomic DNA regions of insertion in order to assess the proper integration of the plasmid/cassette in the *C. albicans* genome.

Flow cytometry analysis. Cells from overnight cultures were diluted 1:1,000 in 1× phosphate-buffered saline (PBS) in BD Falcon tubes (product code 352054), and 10^6 cells were analyzed by flow cytometry in a 96-well plate format with a MACSQuant (Miltenyi) flow cytometer. We used the 405- and 488-nm lasers to excite the BFP and GFP proteins and the 425/475 and 500/550 filters to detect the BFP and GFP emission signals.

LOH screen with the BFP-GFP system. (i) Primary screen. Strains in the 3R overexpression collection were screened for the frequency of LOH by flow cytometry following overnight growth at 30°C in YPD medium containing ATc (3 μg/ml). To evaluate the frequency of LOH at the GFP or BFP loci, we first analyzed the flow cytometry output for 14 independent cultures of the control strain (CEC3989) with FlowJo 7.6. All profiles were very similar. We created gates to define the Bfp^+ Gfp^- (mono-BFP) and Bfp^- Gfp^+ (mono-GFP) populations in one of the control samples and applied these gates to the rest of the data set for the control cultures. The percentages of Bfp^+ Gfp^- and Bfp^- Gfp^+ cells were exported in Excel format. The mean and standard deviation were calculated. The same gates were applied to all of the mutants, resulting in percentages of Bfp^+ Gfp^- and Bfp^- Gfp^+ cells for each mutant strain. We calculated a Z score for each mutant and selected mutants that showed Z scores of >5 for the Bfp^- Gfp^+ population (corresponding to 5 standard deviations from the mean value of the control strain).

(ii) Secondary screen. Because the overexpression-induced morphology changes observed in some of the candidate overexpression mutants (data not shown) could distort the flow cytometry results, a secondary screen was carried out to confirm the candidate genes. All of the mutants identified in the primary screen were grown in triplicate in YPD medium and in YPD-ATc (3 μg/ml) for 8 h and reinoculated into YPD medium alone to recover overnight at 30°C. The next day, 10^6 cells in the yeast form, as confirmed by microscopy, were analyzed by flow cytometry and a *t* test was performed to validate the candidates.

Sorting of cells having undergone a LOH event. The control strain and the candidate mutants were grown in the presence of ATc (3 μg/ml) for 8 h and allowed to recover overnight in YPD medium alone. FACS of these cultures into 1.5-ml Eppendorf tubes with a FACSAria III cell sorter (BD Biosciences) was performed at a rate of 10,000 events/s. A thousand cells from the Bfp^- Gfp^+ and Bfp^+ Gfp^- populations were sorted, recovered in 400 μl of YPD medium, and immediately plated onto YPD medium plates. The plates were incubated at 30°C for 2 days. For the Bfp^- Gfp^+ population, 48 colonies were inoculated into 800 μl of YPD medium in 96-well plates and grown overnight at 30°C. We observed a colony size phenotype of the Bfp^+ Gfp^- population, and therefore, 24 small and 24 large colonies were inoculated into 800 μl of YPD medium in 96-well plates and grown overnight at 30°C. Cultures were then analyzed by flow cytometry to validate monofluorescence and spotted onto YPD medium, SC-His-Arg, SC-His, and SC-Arg to confirm the LOH events by checking the loss of the auxotrophic marker associated with the BFP or GFP marker. We next investigated the molecular mechanisms giving rise to these LOH events by single-nucleotide polymorphism (SNP) typing of 16 isolates, when possible. Genomic DNA was extracted from the sorted clones, and PCR-restriction fragment length polymorphism (RFLP) was performed to assess the extent of the LOH. We used SNPs 95 and 156, which are part of AluI and TaqI restriction sites, respectively (31). The SNP typing analysis was carried out for both the Bfp^+ Gfp^- (small and large colonies) and Bfp^- Gfp^+ populations and was performed as described previously (31).

Whole-genome sequencing and sequence analysis. Genomic DNAs isolated from the *C. albicans* isolates were processed to prepare libraries for Illumina sequencing. The TruSeq Nano DNA Sample Prep kit (Illumina) was used according to the manufacturer's recommendations. DNA was randomly fragmented by sonication to an average fragment length of 550 bp, and Illumina adapters were blunt-end ligated to the fragments. The final libraries were amplified by PCR and then sequenced on a MiSeq platform by using v3 chemistry. Three-hundred-nucleotide paired-end reads were aligned with *C. albicans* strain SC5314 reference genome assembly 21, which is available from the Candida Genome Database (32, 33), by Burrows-Wheeler Aligner (34). Single-nucleotide variants (SNVs) between the sequenced genomes and the reference genome were identified with the Genome Analysis Toolkit (35) at positions with a sequencing depth of ≥18×. Heterozygous SNVs were defined as positions where 15% or more of the calls showed one allele and 85% or less of the calls showed a second allele. Homozygous SNVs were defined as positions where more than 90% of the calls differed from the reference genome. Sequencing depth and heterozygous SNP density maps were constructed as described by Loll-Krippleber et al. (16). Homozygous SNP density maps were constructed by determining the number of homozygous positions per 10-kb region and plotting each value.

RESULTS

A new reporter system for LOH. As an alternative to the *GAL1* and *URA3* systems routinely used for studies of LOH in *C. albicans* (12, 18), we developed a new system combining the use of fluorescent markers and flow cytometry that is more suitable for high-throughput studies. This system relies on the construction of an artificial locus on Ch4 in the intergenic region between the *PGA59* and *PGA62* open reading frames (ORFs) (36). This 9-kb region has been used previously as a platform for the integration of *C. albicans* two-hybrid plasmids (37), and its modification is thought to be neutral to *C. albicans* biology, as its deletion did not result in any phenotype (36). As shown in Fig. 1A, one of the Ch4 homologs was engineered to carry the gene for GFP (38) under the control of the promoter of the constitutively highly expressed *TDH3* gene (25) and linked to an auxotrophic marker (*ARG4*, *HIS1*, *URA3*) or the nourseothricin resistance marker (*SAT1*). The other homolog was engineered to carry the gene for BFP (Evrogen) placed under the control of the *TDH3* promoter and linked to another auxotrophic marker or the nourseothricin resistance marker. A series of cassettes have been developed that associate the *BFP* or *GFP* gene to different selection markers (*SAT1*, *HIS1*, *URA3*, and *ARG4*), allowing their use in various genetic backgrounds (see Fig. S1 in the supplemental material). In this setup, LOH events can be detected by flow cytometry since cells that undergo LOH at the *PGA59-62* locus will produce only one of the two fluorescent proteins. As shown in Fig. 1B, FACS analysis of strain CEC2683, which carries the *BFP-SAT1/GFP-ARG4* system, showed that the majority (>99%) of the cells coexpressed the two fluorescent proteins BFP and GFP. In addition, mono-GFP and mono-BFP populations were observed at frequencies of 0.020 and 0.033%, respectively. In order to verify that the cells in these populations had undergone genuine LOH events, we recovered cells from each population by cell sorting and characterized 76 and 52 cells from the mono-GFP and mono-BPF populations, respectively. Of the cells recovered from both populations, 95% (72/76)

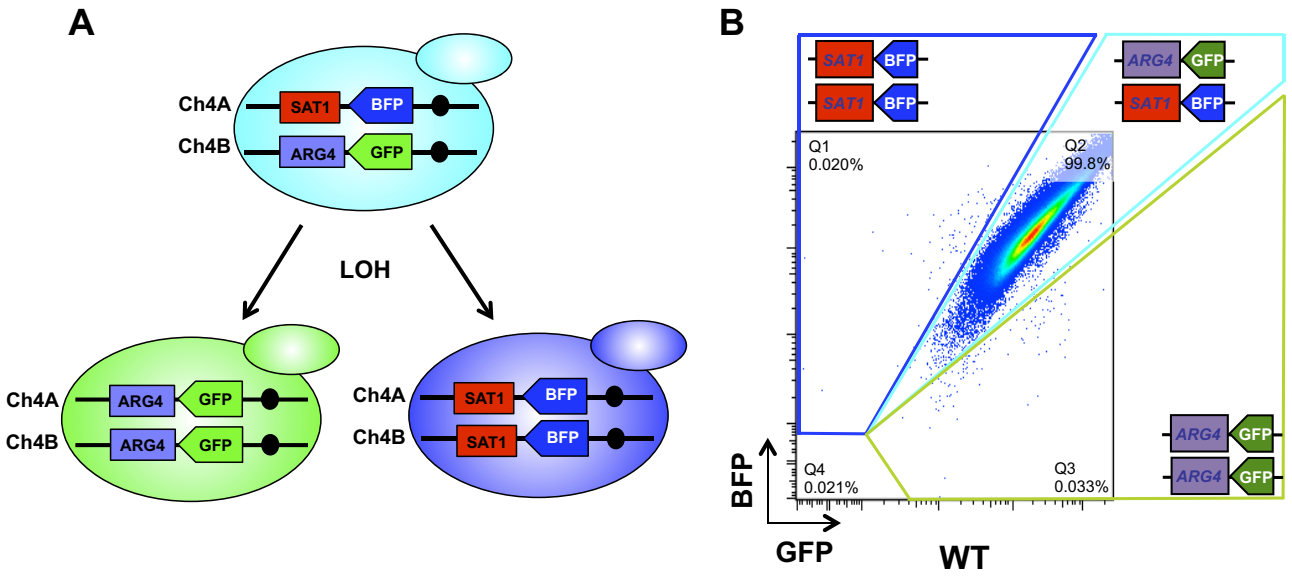


FIG 1 The BFP-GFP system. (A) Artificial heterozygous locus on Ch4. BFP and GFP expression cassettes were integrated into the gene-free *PGA59-PGA62* intergenic region of Ch4. One homolog carries the gene coding for BFP, while the other carries the gene coding for GFP. Both the *BFP* and *GFP* genes are under the control of the promoter of the constitutively highly expressed *TDH3* gene and linked to a selection marker. When LOH takes place at this locus, the cells express only one of the fluorescent proteins and LOH events can be detected by flow cytometry. (B) Flow cytometry analysis of a WT BFP-GFP strain (CEC2683). Cells were grown overnight in rich medium, diluted in 1× PBS, and analyzed on a MACSQuant cytometer (Miltenyi Biotech). One million events are displayed.

and 94% (49/52) were monofluorescent (data not shown). Moreover, spotting assays on appropriate media indicated that all of the monofluorescent cells displayed the phenotypes expected from a concomitant loss of the *BFP* and *SAT1* genes (mono-GFP cells) or the *GFP* and *ARG4* genes (mono-BFP cells), i.e., sensitivity to nourseothricin or arginine auxotrophy, respectively (data not shown). This initial analysis of a BFP-GFP strain revealed a spontaneous LOH frequency of 2×10^{-4} to 3×10^{-4} at this locus. Interestingly, the cells recovered from the Bfp^{-} Gfp^{-} population did not grow, indicating that these are nonviable cells. Fluorescence plots of mono-BFP (CEC4007), mono-GFP (CEC3998), unlabeled (SN100), and BFP-GFP-labeled strains are presented in Fig. S2A in the supplemental material.

Increased LOH at the *PGA59-62* locus in *mec1* mutants. The Mec1 kinase is a central regulator of the DNA damage checkpoint in all eukaryotes (39, 40). A previous study has shown that inactivation of the *C. albicans MEC1* gene results in an increase in LOH at the *GAL1* locus (15). In order to test whether an increase in LOH could be detected with the BFP-GFP system, we inactivated the *MEC1* gene in a strain carrying this system. Results presented in Fig. 2A and Table 2 show a statistically significantly greater proportion of mono-GFP or mono-BFP cells in the *mec1* null mutant than in the wild-type (WT) strain. Interestingly, we also observed an increase in the proportion of Bfp^{-} Gfp^{-} cells, suggesting that inactivation of *MEC1* increases cell death (Fig. 2A). Cell death in the Bfp^{-} Gfp^{-} population was confirmed by propidium iodide staining and microscopic analysis of these cells after cell sorting (data not shown). Hence, the BFP-GFP LOH reporter system could detect an increase in the frequency of LOH by flow cytometry upon inactivation of the *MEC1* gene.

Increased LOH at the *PGA59-62* locus in response to different stresses. Forche et al. (12) have shown increased LOH in response to different stresses that are reminiscent of the conditions encountered by fungi during human infection, e.g., elevated temperatures, oxidative stress, and the presence of azole antifungals. We evaluated the frequency of LOH in a control strain carrying the BFP-GFP system when grown overnight at 30, 37, or 39°C. Although cells were filamenting after 7 h of growth at 37 or 39°C, microscopic analysis of cells after overnight growth at 37 or 39°C revealed that the majority of the cells reverted to the yeast form and were therefore suitable for flow cytometry analysis (data not shown). Still, we allowed the cells to recover overnight in YPD medium before flow cytometry analysis. Similar to the data obtained with the *GAL1* system (12), the BFP-GFP system revealed that elevated temperatures promote LOH in *C. albicans* (Fig. 2B and Table 2). The fluorescence plots presented in Fig. S2B in the supplemental material confirmed that the distribution of mono-labeled cells remains the same in the absence or presence of heat stress.

We also treated the control strain carrying the BFP-GFP system with the genotoxic agent MMS. After a 30-min treatment with different concentrations of MMS, cells were allowed to recover in rich medium overnight. Flow cytometry analysis revealed a frequency of LOH that increased together with MMS concentrations (Fig. 2C and Table 2).

Taken together, these data indicated that the BFP-GFP LOH reporter system was suitable for the detection of increases in the frequency of LOH resulting from physical or chemical stress.

Construction of the 3R overexpression collection. We reasoned that the BFP-GFP LOH reporter system could provide an efficient means for the identification of genes that control LOH in *C. albicans*. Therefore, we implemented an overexpression screen aimed at identifying *C. albicans* genes whose overexpression would alter the WT frequency of LOH. To do this, we generated a collection of *C. albicans* overexpression strains focused on genes that, on the basis of their Gene Ontology database annotations or

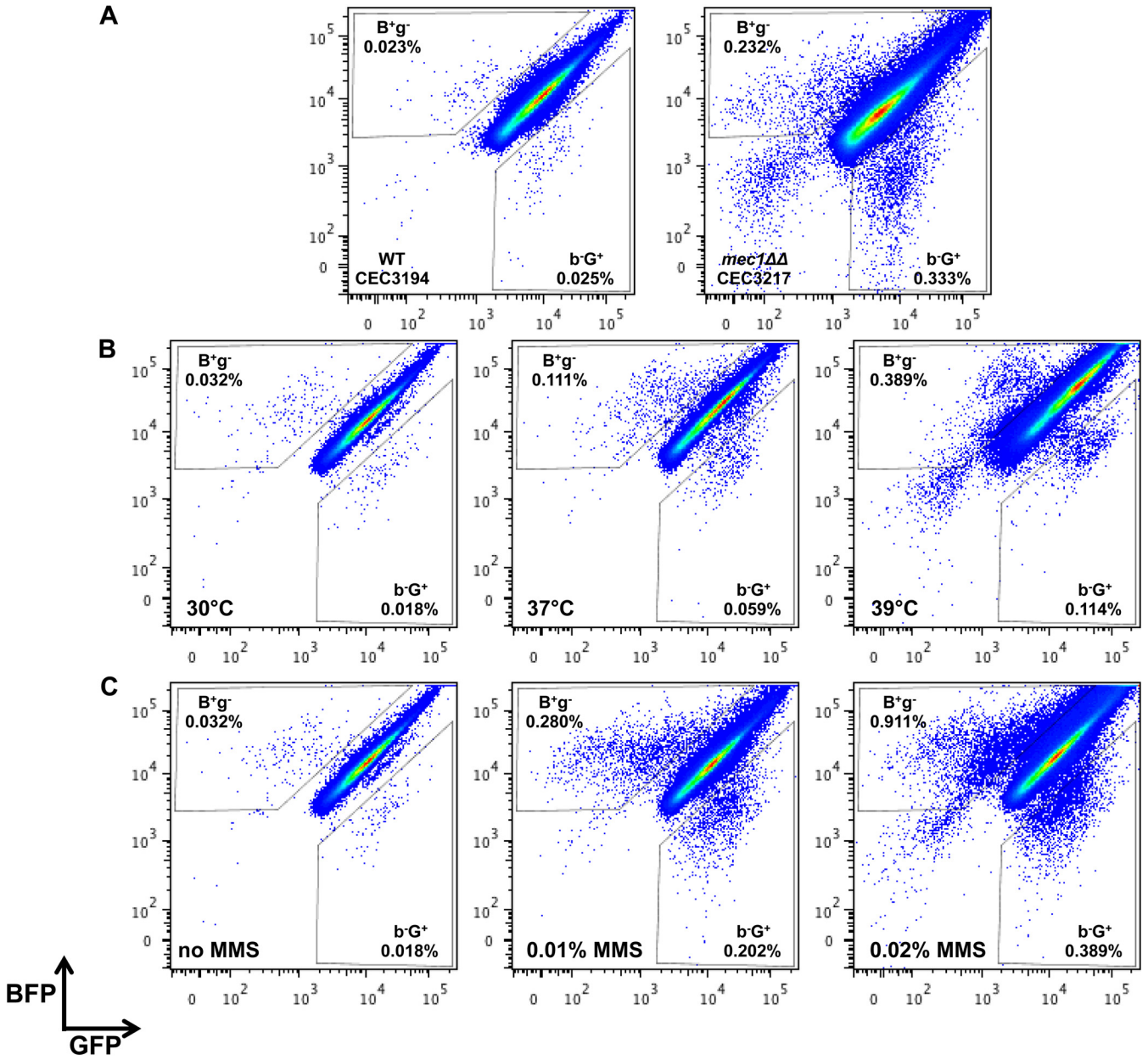


FIG 2 Flow cytometry analysis reveals an increase in LOH in *mec1* null mutants and in response to physical and chemical stresses. (A) Flow cytometry analysis of homozygous *mec1* deletion strains. WT BFP-GFP strain CEC3194 and the *mec1*ΔΔ mutant were grown overnight in YPD medium at 30°C and analyzed by flow cytometry. (B) Flow cytometry analysis upon heat treatment. WT BFP-GFP strain CEC3989 was grown overnight in YPD medium at 30, 37, or 39°C and allowed to recover overnight in YPD medium before flow cytometry analysis. (C) Flow cytometry analysis in the presence of the DNA-damaging agent MMS. WT BFP-GFP strain CEC3989 was treated with increasing concentrations of MMS for 30 min. Cells were then allowed to recover overnight in fresh YPD medium and analyzed by flow cytometry.

the function of their orthologs in *S. cerevisiae*, were likely to be involved in different aspects of genome maintenance such as DNA repair, DNA replication, recombination, chromosome segregation, the cell cycle, and telomere maintenance. PCR products extending from the start codon to the penultimate codon of 204 selected genes (see Table S2 in the supplemental material) were cloned into the pDONR207 donor vector. Following Sanger and Illumina/Solexa sequence validation, a total of 151 (74%) derivatives of pDONR207 were obtained. ORFs cloned into pDONR207 were subsequently transferred into uniquely barcoded CIp10-P_{TET}-GTW plasmids (20, 21). A total of 147 (72%) CIp10-P_{TET}-GTW derivatives were obtained and subsequently introduced at the *RPS1* locus into *C. albicans* strain CEC2824, which harbors a *GFP-HIS1*/*BFP-ARG4* system at the *PGA59-62* locus and the pNIMX plasmid encoding a tetracycline-controlled transactivator (21). Eventually, 124 *C. albicans* P_{TET}-driven overexpression strains were obtained, indicating a 60.8% overall success rate, slightly below that reported previously by Chauvel et al. (21) for a nonoverlapping set of overexpression plasmids, possibly because of the larger size of the cloned ORFs in our study (27% were

TABLE 2 LOH increase in the *mec1* null mutant and in response to physical and chemical stresses

Genotype, lab ID, and growth conditions	LOH quantification on Ch 4 (BFP loss)		
	Mean frequency of LOH ± SEM ($\times 10^{-4}$)[a]	Fold change	*P* value[d]
WT, CEC3194, YPD, 30°C	2.6 ± 0.2		
*mec1*Δ, CEC3206, YPD, 30°C	2.2 ± 0.1	1[b]	≤0.005
*mec1*ΔΔ, CEC3216, YPD, 30°C	31.0 ± 1.0	12[b]	≤0.005
WT, CEC3989			
YPD, 30°C	0.8 ± 0.1		
YPD, 37°C	1.2 ± 0.1	1.4[c]	≤0.05
YPD, 39°C	1.3 ± 0.1	1.5[c]	≤0.01
YPD + 0.01% MMS, 30°C	19.4 ± 2.6	12[c]	≤0.005
YPD + 0.02% MMS, 30°C	45.3 ± 1.8	28[c]	≤0.005
YPD + 0.03% MMS, 30°C	59.0 ± 1.8	37[c]	≤0.005

[a] Six independent cultures were tested per *mec1* mutant and per physical and/or chemical stress condition. Standard errors of the means were determined with GraphPad Prism software.
[b] Ratio of the LOH frequency of a given mutant to that of the parental strain.
[c] Ratio of the LOH frequency in the presence of the stress to that in the absence of the stress.
[d] A Mann-Whitney test was performed to determine if the BFP loss frequency in the different mutants was statistically significantly different from the rate observed in the control strains or under the control conditions.

>2,500 bp versus none in the study of Chauvel et al.). Together with the CEC3989 control strain harboring the empty CIp10-P_{TET}-GTW plasmid, these 124 overexpression strains are referred to here as the 3R overexpression collection.

FACS-optimized LOH screens upon gene overexpression. Prior to analyzing the overexpression mutants for the frequency of LOH under inducing conditions, we evaluated the frequency of LOH in control strain CEC3989. Fourteen independent YPD-ATc (3 μg/ml) cultures of the control strain were analyzed by flow cytometry. For each analysis, 1 million cells were tested and the mean frequency of LOH and the standard deviation were calculated. We found that LOH occurred at an average frequency of $2.2 \times 10^{-4} \pm 0.4 \times 10^{-4}$. Subsequently, we submitted the entire 3R overexpression collection to a primary screen whereby 1 million cells of each overexpression strain were analyzed by flow cytometry after overnight induction. Thirty-three overexpression strains were selected that had a Z score (for the frequency of Bfp^- Gfp^+ cells) of >5 (see Fig. S3 in the supplemental material). In the secondary screen, we allowed the cells to recover in rich medium after 8 h of induction in YPD-ATc3 medium and reanalyzed the 33 overexpression strains identified in the primary screen in triplicate. While we saw no difference between Bfp^+ Gfp^- and Bfp^- Gfp^+ frequencies of the control strain and the overexpression mutants in the absence of ATc treatment, we confirmed the increase in LOH at the *PGA59-62* locus after ATc treatment for four of these genes, namely, *BIM1* (10-fold for Bfp^+ Gfp^- cells and 13-fold for Bfp^- Gfp^+ cells), *CDC20* (32- and 38-fold, respectively), *RAD51* (7- and 8-fold, respectively), and *RAD53* (13-fold) (Table 3). Notably, we could not identify any overexpression strain that had a significantly reduced frequency of LOH, possibly owing to the relatively low frequency of LOH in the control strain.

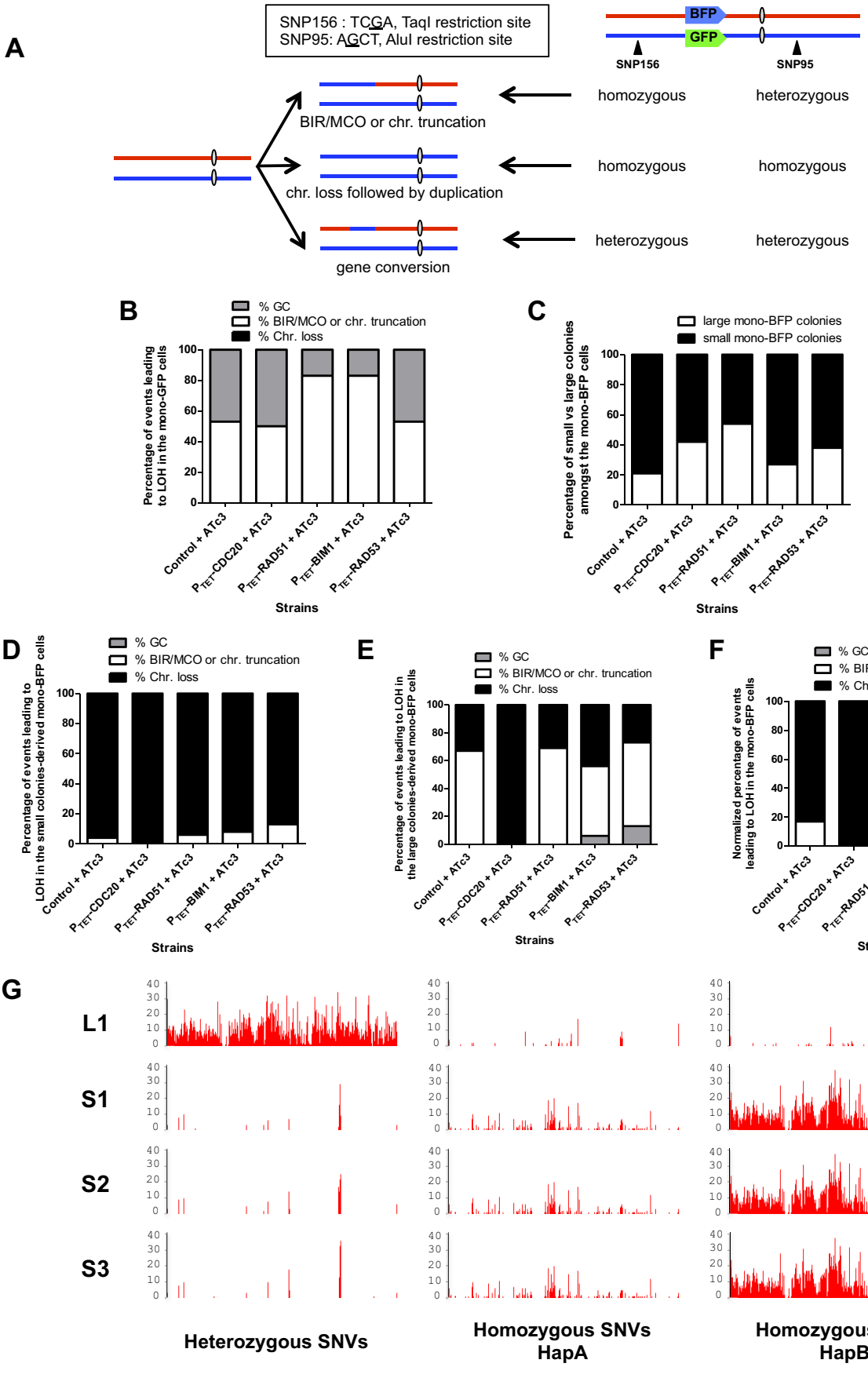
Characterization of the molecular mechanisms leading to LOH. LOH at the BFP-GFP locus could result from gene conversion, mitotic crossing over (MCO), break-induced replication (BIR), or chromosome truncation or loss. Further, to identifying genes whose overexpression increased the frequency of LOH, we tested whether the overexpression of these genes could also alter the balance between these different types of events, as observed in the control strain.

FACS was used in order to recover individual monofluorescent cells and subsequently analyze these cells by SNP-RFLP typing, which can reveal the molecular mechanism—gene conversion, MCO/BIR, or chromosome truncation or loss—underlying the LOH event responsible for loss of the BFP or GFP marker on Ch4 (Fig. 3A) (31). Cells sorted from the Bfp^- Gfp^+ control and mutant populations and plated on YPD medium gave rise to colonies homogeneous in size. For the control strain, SNP-RFLP typing showed that LOH in the cells that had yielded these colonies had

TABLE 3 LOH quantification in the four candidate overexpression mutants validated in the secondary LOH screen

Strain[a]	LOH quantification on Ch4					
	Bfp^+ Gfp^- cells			Bfp^- Gfp^+ cells		
	Mean frequency of LOH ± SEM ($\times 10^{-4}$)	Fold change[b]	*P* value[c]	Mean frequency of LOH ± SEM ($\times 10^{-4}$)	Fold change[b]	*P* value[c]
CEC3989 (control)	1.3 ± 0.5			0.7 ± 0.3		
P_{TET}-*CDC20*	42.0 ± 11.0	32	≤0.01	26.8 ± 7.9	38	≤0.01
P_{TET}-*RAD53*	17.0 ± 6.0	13	≤0.01	9.1 ± 1.5	13	≤0.01
P_{TET}-*BIM1*	12.5 ± 1.7	10	≤0.01	9.0 ± 2.9	13	≤0.01
P_{TET}-*RAD51*	8.7 ± 1.4	7	≤0.01	5.3 ± 1.3	8	≤0.01

[a] All strains were grown on YPD + Atc.
[b] Ratio of the LOH frequency in a given mutant to that in control strain CEC3989.
[c] Mann-Whitney test.



arisen at somewhat equal frequencies through MCO/BIR or chromosome truncation, these two types of events being undistinguishable by SNP-RFLP typing, and gene conversion (Fig. 3B). Complete loss of the homolog carrying the BFP marker was never observed. A similar pattern was observed upon the overexpression of *CDC20* and *RAD53*. In contrast, we detected an increase in the proportion of MCO/BIR or chromosome truncation events versus gene conversions resulting in loss of the BFP marker in the *RAD51* and *BIM1* overexpression mutants (Fig. 3B).

Notably, colonies of different sizes were observed when cells sorted from the Bfp$^+$ Gfp$^-$ populations were plated on YPD medium (data not shown). This included colonies of a size similar to that of colonies obtained from the Bfp$^-$ Gfp$^+$ populations and smaller colonies. This small-colony phenotype was observed for the control strain and all of the overexpression mutants tested. In the control strain, 80 and 20% of the mono-BFP cells were derived from small and large colonies, respectively (Fig. 3C). We observed an increase in the proportion of the large-colony-derived mono-BFP cells in the *CDC20*, *RAD51*, and *RAD53* overexpression mutants (Fig. 3C). SNP-RFLP typing of 16 mono-BFP cells from small and large colonies revealed that, unlike in the Bfp$^-$ Gfp$^+$ population, chromosome loss was the major mechanism responsible for loss of the GFP marker in the small-colony-derived cells of both the control strain and the overexpression mutants, other LOH events resulting from MCO/BIR, or chromosome truncation (Fig. 3D). On the other hand, SNP-RFLP typing revealed that LOH was often the result of recombination-mediated events or chromosome truncations in the larger colonies of the control strain and the *RAD51*, *BIM1*, and *RAD53* overexpression mutants (Fig. 3E). In contrast, chromosome loss was the only mechanism leading to LOH in the Bfp$^+$ Gfp$^-$ cells of the *CDC20* overexpression mutants (Fig. 3D and E). SNP-RFLP typing data for both types of colonies were normalized to the proportion of small versus large colonies and combined to represent the percentage of each molecular mechanism (gene conversion, MCO/BIR, or chromosome truncation or loss) in the overall Bfp$^+$ Gfp$^-$ population (Fig. 3F).

Although the homozygous status of both SNPs 156 and 95 is routinely recognized as a mark of LOH chromosome loss on Ch4 (31), multiple gene conversion events could also be responsible for the concomitant homozygous status of these SNPs. To determine the extent of the LOH, we performed whole-genome sequencing of three small Bfp$^+$ Gfp$^-$ colonies obtained from an overexpression mutant that did not show any change in the frequency of LOH relative to the control and in which LOH was the result of chromosome loss, as demonstrated by SNP-RFLP typing. Heterozygous and homozygous SNP density maps against the *C. albicans* SC5314 HapA and HapB reference sequences (41) revealed that cells from the three small colonies had lost all of Ch4B carrying the GFP marker (Fig. 3G). The observation that Ch4 had remained disomic (data not shown) indicates that the other homolog, Ch4A, has been reduplicated. Homozygosis of Ch4A also accounts for the histidine auxotrophy displayed by Bfp$^+$ Gfp$^-$ cells, despite the maintenance of the *BFP-HIS1* cassette at the *PGA59-62* locus (data not shown). Indeed, the inactive *HIS4* allele reported by Gómez-Raja et al. (42) is carried by Ch4A.

Overall, our SNP-RFLP analysis revealed several interesting features. (i) The molecular mechanisms causing the LOH differed according to the Ch4 homolog undergoing the LOH event, with chromosome loss being observed only for Ch4B. (ii) This was reflected by the occurrence of small-colony clones when all or part of Ch4B was lost. (iii) When chromosome loss occurred, overexpression of *CDC20* favored this mechanism. (iv) *RAD51* overexpression favored MCO/BIR events. (v) *BIM1* overexpression favored MCO/BIR events unless chromosome loss was possible. (vi) *RAD53* overexpression favored MCO/BIR events when chromosome loss was possible.

DISCUSSION

Here, we present a FACS-optimized genetic system for the detection of LOH in *C. albicans* and its application to the identification of genes whose overexpression results in an increase in the frequency of LOH and changes in the frequencies of the molecular events that are at the origin of LOH.

The new BFP-GFP system displays features more appealing than those of other genetic systems currently used to study LOH in *C. albicans*. Indeed, we provide a powerful and robust tool that allows rare-event analysis and high-throughput LOH detection since up to 96 samples can be processed in a single run. This system is robust and reproducible because LOH is studied on a cellular scale and a large number of cells are analyzed. Unlike the systems based on the *GAL1* or *URA3* marker, which require large amounts of expensive drugs, the BFP-GFP system does not require any costly consumables. The use of drugs such as 2-DG or 5-FOA also raises the issue of exposing cells to selective pressure that might distort the evaluation of the frequency of LOH. In contrast, the BFP-GFP system allows measurement of the frequency of spontaneous LOH since cells are not exposed to any

FIG 3 SNP typing reveals a shift in the molecular mechanisms leading to LOH upon the overexpression of some of the candidate genes. (A) Map of Ch4 and localization of the BFP-GFP system. The SNPs used for RFLP characterization are indicated by black triangles. Telomere-proximal SNP 156 is part of a TaqI restriction site. One allele contains the TaqI site, while the other does not. Centromere-proximal SNP 95 is located in the middle of an AluI restriction site. One allele contains the AluI site, while the other does not. The heterozygosity status of these SNPs provides information on the molecular mechanisms that give rise to LOH. chr., chromosome. (B) SNP-RFLP typing of Bfp$^-$ Gfp$^+$ cells. The histogram shows the proportion of MCO/BIR or chromosome truncation versus chromosome loss in the Bfp$^-$ Gfp$^+$ population. MCO/BIR or chromosomal truncation events correspond to isolates that have maintained a heterozygous SNP 95 and displayed a homozygous SNP 156. Chromosome loss events correspond to isolates in which both SNPs 95 and 156 became homozygous. Gene conversion events correspond to isolates in which both SNPs 95 and 156 remained heterozygous. (C) Proportion of small versus large colonies giving rise to true mono-BFP cells. (D) SNP-RFLP typing of small-colony-derived Bfp$^+$ Gfp$^-$ cells. (E) SNP-RFLP typing of large-colony-derived Bfp$^+$ Gfp$^-$ cells. (F) SNP-RFLP typing of Bfp$^+$ Gfp$^-$ cells. The normalized percentage of each molecular mechanism, for instance, gene conversion, was calculated as follows: normalized % GC = (% small colonies × % GC in small colonies) + (% large colonies × % GC in large colonies). (G) Heterozygous and homozygous SNV density maps of one large- and three small-colony-derived cells that have undergone LOH at the BFP-GFP locus of Ch4. The number of heterozygous SNVs in each 10-kb region along Ch4 was computed and is shown, revealing LOH by gene conversion in the large-colony-derived cells and by chromosome loss and reduplication on Ch4 in the small-colony-derived cells. Sequencing reads were mapped on the HapA or HapB reference sequence and yielded similar heterozygous SNP density maps. Only the result of mapping on HapB is shown (left). The number of homozygous SNVs in each 10-kb region along HapA (middle) or HapB (right) was computed and is shown. The high density of homozygous SNVs upon mapping on HapB indicates that chromosome loss in small-colony-derived cells affected Ch4B.

stress. Like the other LOH reporter systems, the BFP-GFP system is, in theory, flexible since the markers can be inserted anywhere in the genome, thus enabling determination of the frequency of LOH at different genomic locations and therefore the study of site-specific LOH events.

The BFP-GFP system was used to screen a collection of 124 *C. albicans* overexpression mutants to identify genes whose overexpression would result in an increase in the frequency of LOH. Our screen identified four candidates as regulators of genome stability in *C. albicans*. Null or conditional deletion mutants have been obtained and studied for three of these genes, *CDC20* (43), *RAD51* (44), and *RAD53* (45), yet the role of these genes in genome integrity was never addressed, except for *RAD53*, whose deletion increases the frequency of LOH (16). Reports of *C. albicans* mutants overexpressing the four candidate genes could not be found in the literature. In contrast, *S. cerevisiae* mutants null for the four candidate genes, containing conditional deletions of them, or overexpressing them have been described. Although genome stability was investigated in the deletion mutants, the question of genome integrity in the overexpression mutants was not addressed.

Our screen identified *CDC20* as the gene whose overexpression triggers the greatest increase in the frequency of LOH. In *S. cerevisiae*, the essential *CDC20* gene encodes an activator of ubiquitin-protein ligase activity and is required for the metaphase-anaphase transition. At this stage of the mitotic cycle, Cdc20p enables the anaphase-promoting complex to properly degrade securin, allowing the degradation of the cohesin rings that link the two sister chromatids. Several studies have shown that degradation of Cdc20p is an essential and conserved mechanism of spindle assembly checkpoint maintenance (46, 47). Therefore, the sustained presence of the Cdc20 protein could perturb genome integrity and our observation that overexpression of *C. albicans CDC20* favors LOH, especially through chromosome loss when these types of events can occur, is consistent with this hypothesis. Conditional *CDC20* deletion mutants of *C. albicans* have been characterized by Chou et al. (43). Their work suggested that, similar to *ScCDC20*, *CaCDC20* was important for the metaphase-to-anaphase transition and mitotic exit. Interestingly, *CDC20* has a distinct role in morphogenesis in *C. albicans*, which is not the case in *S. cerevisiae*.

Rad53 and Rad51, two proteins involved in DNA damage checkpoint maintenance and DNA repair, were identified as causing an increase in LOH upon overexpression. Rad53 is a kinase involved in DNA damage checkpoint maintenance whose deletion causes a growth defect, an increased sensitivity to several DNA-damaging agents, and a defect in genotoxic-stress-induced filamentation in *C. albicans* (45). An increase in LOH could also be observed in *CaRAD53* deletion mutants (16). The recombinase Rad51 plays a major role in homologous recombination during DNA double-strand break repair by searching for sequence homology and promoting strand pairing. Deletion of *RAD51* in *C. albicans* results in a decreased growth rate and increased sensitivity to various DNA-damaging agents (44). On the other hand, genome integrity has not been investigated in *CaRAD51* deletion mutants. In *S. cerevisiae*, *RAD51* overexpression has been shown to promote genome instability, probably by inhibiting the accurate repair of double-strand breaks by homologous recombination (48). Similarly, *RAD51* overexpression in the mouse leads to an increase in genome instability, notably, LOH events (49). Our observation that, in addition to increasing the frequency of LOH, overexpression of *RAD51* and *RAD53* favored MCO/BIR events is consistent with their role in DNA double-strand break repair, their overexpression being likely to unbalance the DNA damage response and favor recombination-mediated events.

In this work, a structural component of the microtubule skeleton of yeast encoded by *BIM1* has been identified as increasing the frequency of LOH upon overexpression. While *ScBIM1* deletion has been shown to result in increased sensitivity to various DNA-damaging agents (50) and increased chromosome instability (51), the effect of *BIM1* overexpression on genome integrity has not been investigated, probably in part because of the strong growth defect associated with *BIM1* overexpression in *S. cerevisiae* (52–54). Similarly, we also observed a growth defect associated with *BIM1* overexpression in *C. albicans* (data not shown). Our observation that *BIM1* overexpression, in addition to promoting LOH, favors MCO/BIR events was unexpected, as Bim1 is involved in chromosome segregation, and therefore its overexpression would be more likely to cause chromosome loss. We hypothesized that the overexpression-induced reduced growth rate associated with *BIM1* overexpression could favor the occurrence of DNA breaks and thus explain the increase in MCO/BIR events or chromosome truncations in the *BIM1* overexpression mutant.

Our SNP typing analysis of sorted cells revealed an additional important aspect. Indeed, we have observed that the molecular mechanisms giving rise to LOH can be homolog specific, with chromosome loss never affecting Ch4A. Notably, loss of all or part of Ch4B was accompanied by variations in colony size. As cells in small colonies have almost all arisen through chromosome loss (Fig. 3D) and yield large colonies (data not shown), we hypothesize that small colonies reflect a lower growth rate of cells that are monosomic for Ch4A and eventually duplicate this chromosome. Chromosome homozygosis was long thought to be hampered in *C. albicans* by the presence of recessive lethal mutations dispersed throughout the genome. The existence of haploids (2), completely homozygous *rad52*-derived isolates (55), and partly homozygous diploid parasexual progenies (56) demonstrated that homozygosis of each one of the eight chromosomes can occur. Nevertheless, a bias in the homolog being retained was reported for certain chromosomes (2, 55). This was the case for Ch4, one homolog of which could never be lost (2, 55). Our observation that Ch4A could never be lost suggests the presence of at least one recessive lethal allele on Ch4B. Because MCO/BIR or chromosome truncation-mediated LOH occurred in mono-GFP cells, we can narrow the recessive lethal allele(s)-carrying region to the right arm of Ch4B or the left arm of Ch4B between the *PGA59*/*62* locus and the centromere. Genome sequencing of these mono-GFP cells will reveal the extent of the LOH and allow refinement of the position of this hypothetical recessive lethal allele(s). While homozygosis along the entire smaller chromosomes, Ch5, Ch6, and Ch7, was often observed, MCO/BIR events or chromosome truncations have been reported to be responsible for most LOH in both Ch1 homologs (12, 13). Altogether, these observations suggest that the mechanisms underlying LOH in *C. albicans* could be chromosome specific, as well as homolog specific. The presence of essential alleles that cannot be lost might indeed select for cells that have repaired double-strand breaks through BIR, especially in the case of Ch1, one of the largest chromosomes in *C. albicans* (3.2 Mb). On the other hand, the loss of one Ch4 homolog, which is half the size of Ch1 (1.6 Mb), might be less detrimental to the cell. Alternatively, Ch4 might be more prone to mitotic nondisjunction events or to unrepaired DNA lesions.

Some expected genes, such as *TUB2*, were not recovered in our screen (57). This may reflect the inherent limitations of a genetic screen. First, a number of the usual suspects were shown to play a role in genome stability on the basis of deletion mutant study, which does not imply that their overexpression triggers an increase in the frequency of LOH. Second, the level of overexpression achieved in this study might be insufficient to cause a significant phenotype. Because the proteins are produced in an untagged form, we have no means of verifying their production. Third, the lack of phenotype could result from the presence of peptides encoded by the *att* sequences at the amino- and carboxy-terminal ends of the proteins, which could alter their proper folding and activity.

Living cells have evolved different mechanisms in order to ensure genome integrity (58). Interestingly, these mechanisms are conserved across species, and cancer genetic studies have linked malfunctions in these processes to the genome instability observed in human cancer cells (59, 60). Therefore, the characterization of all of the players involved in genome integrity maintenance and a better understanding of the associated signaling pathways are very important from the human perspective. In this respect, functional genomic approaches to model eukaryotic species such as *S. cerevisiae* have uncovered many components required for genome maintenance and integrity (61). Genome-wide genetic screens of this species have examined deletion mutants for a growth defect upon exposure to different DNA-damaging agents (62–65) or for synthetic genetic interaction with other mutant genes important for the cellular response to DNA damage (66–69). Other screens identified mutants with increased genomic instability by investigating the mutation rates within a specific gene (70) or the rates of gross chromosomal rearrangements (51, 71–75). While several studies have used *S. cerevisiae* to study genome stability, drastic genome changes are not well tolerated by this organism, thus correlating with a high fitness cost (76). Interestingly, *C. albicans* is more tolerant of genome changes, which is reminiscent of the situation in cancer cells that continue to divide rapidly despite undergoing massive genome rearrangements. The tools presented here, in association with the *C. albicans* ORFeome we are currently constructing with the group of Carol Munro in Aberdeen (20), will allow *C. albicans* to become a model of choice for the study of mechanisms involved in eukaryotic genome integrity.

ACKNOWLEDGMENTS

We thank Christiane Bouchier and Laurence Ma from the Pasteur Plateforme Genomique (PF1) for Illumina sequencing of the overexpression plasmids. We are thankful to members of the BIOASTER genomics and transcriptomics core facility, namely, Lilia Boucinha and Frédéric Reynier, for help with genome sequencing. We are also grateful to other members of Fungal Biology and Pathogenicity Unit for helpful discussions.

This work was supported by the French Government's Investissement d'Avenir program, Laboratoire d'Excellence Integrative Biology of Emerging Infectious Diseases (grant ANR-10-LABX-62-IBEID). R.L.-K. was the recipient of a Ph.D. grant from INRA and the Institut Pasteur. M.L. was the recipient of a postdoctoral fellowship from the Institut Pasteur (Bourse Roux).

REFERENCES

1. **Edmond MB, Wallace SE, McClish DK, Pfaller MA, Jones RN, Wenzel RP.** 1999. Nosocomial bloodstream infections in United States hospitals: a three-year analysis. Clin Infect Dis **29:**239–244. http://dx.doi.org/10.1086/520192.
2. **Hickman MA, Zeng G, Forche A, Hirakawa MP, Abbey D, Harrison BD, Wang YM, Su CH, Bennett RJ, Wang Y, Berman J.** 2013. The 'obligate diploid' Candida albicans forms mating-competent haploids. Nature **494:**55–59. http://dx.doi.org/10.1038/nature11865.
3. **Hull CM, Raisner RM, Johnson AD.** 2000. Evidence for mating of the "asexual" yeast Candida albicans in a mammalian host. Science **289:**307–310. http://dx.doi.org/10.1126/science.289.5477.307.
4. **Magee BB, Magee PT.** 2000. Induction of mating in Candida albicans by construction of MTLa and MTLalpha strains. Science **289:**310–313. http://dx.doi.org/10.1126/science.289.5477.310.
5. **Miller MG, Johnson AD.** 2002. White-opaque switching in Candida albicans is controlled by mating-type locus homeodomain proteins and allows efficient mating. Cell **110:**293–302. http://dx.doi.org/10.1016/S0092-8674(02)00837-1.
6. **Bennett RJ, Johnson AD.** 2003. Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. EMBO J **22:**2505–2515. http://dx.doi.org/10.1093/emboj/cdg235.
7. **Selmecki A, Forche A, Berman J.** 2010. Genomic plasticity of the human fungal pathogen *Candida albicans*. Eukaryot Cell **9:**991–1008. http://dx.doi.org/10.1128/EC.00060-10.
8. **Diogo D, Bouchier C, d'Enfert C, Bougnoux ME.** 2009. Loss of heterozygosity in commensal isolates of the asexual diploid yeast Candida albicans. Fungal Genet Biol **46:**159–168. http://dx.doi.org/10.1016/j.fgb.2008.11.005.
9. **Forche A, Magee PT, Selmecki A, Berman J, May G.** 2009. Evolution in Candida albicans populations during a single passage through a mouse host. Genetics **182:**799–811. http://dx.doi.org/10.1534/genetics.109.103325.
10. **Coste A, Turner V, Ischer F, Morschhauser J, Forche A, Selmecki A, Berman J, Bille J, Sanglard D.** 2006. A mutation in Tac1p, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in Candida albicans. Genetics **172:**2139–2156. http://dx.doi.org/10.1534/genetics.105.054767.
11. **Dunkel N, Blass J, Rogers PD, Morschhauser J.** 2008. Mutations in the multi-drug resistance regulator MRR1, followed by loss of heterozygosity, are the main cause of MDR1 overexpression in fluconazole-resistant Candida albicans strains. Mol Microbiol **69:**827–840. http://dx.doi.org/10.1111/j.1365-2958.2008.06309.x.
12. **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. http://dx.doi.org/10.1128/mBio.00129-11.
13. **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. http://dx.doi.org/10.1128/EC.00299-07.
14. **Legrand M, Chan CL, Jauert PA, Kirkpatrick DT.** 2008. Analysis of base excision and nucleotide excision repair in Candida albicans. Microbiology **154:**2446–2456. http://dx.doi.org/10.1099/mic.0.2008/017616-0.
15. **Legrand M, Chan CL, Jauert PA, Kirkpatrick DT.** 2011. The contribution of the S-phase checkpoint genes MEC1 and SGS1 to genome stability maintenance in Candida albicans. Fungal Genet Biol **48:**823–830. http://dx.doi.org/10.1016/j.fgb.2011.04.005.
16. **Loll-Krippleber R, d'Enfert C, Feri A, Diogo D, Perin A, Marcet-Houben M, Bougnoux ME, Legrand M.** 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. http://dx.doi.org/10.1111/mmi.12471.
17. **Boeke JD, LaCroute F, Fink GR.** 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. http://dx.doi.org/10.1007/BF00330984.
18. **Forche A, May G, Beckerman J, Kauffman S, Becker J, Magee PT.** 2003. A system for studying genetic changes in Candida albicans during infection. Fungal Genet Biol **39:**38–50. http://dx.doi.org/10.1016/S1087-1845(02)00585-6.
19. **Gorman JA, Gorman JW, Koltin Y.** 1992. Direct selection of galactokinase-negative mutants of Candida albicans using 2-deoxy-galactose. Curr Genet **21:**203–206. http://dx.doi.org/10.1007/BF00336842.
20. **Cabral V, Chauvel M, Firon A, Legrand M, Nesseir A, Bachellier-Bassi S, Chaudhari Y, Munro CA, d'Enfert C.** 2012. Modular gene over-

expression strategies for Candida albicans. Methods Mol Biol **845:**227–244. http://dx.doi.org/10.1007/978-1-61779-539-8_15.
21. **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. http://dx.doi.org/10.1371/journal.pone.0045912.
22. **Noble SM, French S, Kohn LA, Chen V, Johnson AD.** 2010. Systematic screens of a Candida albicans homozygous deletion library decouple morphogenetic switching and pathogenicity. Nat Genet **42:**590–598. http://dx.doi.org/10.1038/ng.605.
23. **Rose M, Winston F, Hieter P.** 1990. Methods in yeast genetics. a Cold Spring Harbor Laboratory course manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, NY.
24. **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. http://dx.doi.org/10.1128/EC.4.2.298-309.2005.
25. **Delgado ML, Gil ML, Gozalbo D.** 2003. Candida albicans TDH3 gene promotes secretion of internal invertase when expressed in Saccharomyces cerevisiae as a glyceraldehyde-3-phosphate dehydrogenase-invertase fusion protein. Yeast **20:**713–722. http://dx.doi.org/10.1002/yea.993.
26. **Gola S, Martin R, Walther A, Dunkler 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. http://dx.doi.org/10.1002/yea.1044.
27. **Schaub Y, Dunkler A, Walther A, Wendland J.** 2006. New pFA-cassettes for PCR-based gene manipulation in Candida albicans. J Basic Microbiol **46:**416–429. http://dx.doi.org/10.1002/jobm.200510133.
28. **Dennison PM, Ramsdale M, Manson CL, Brown AJ.** 2005. Gene disruption in Candida albicans using a synthetic, codon-optimised Cre-loxP system. Fungal Genet Biol **42:**737–748. http://dx.doi.org/10.1016/j.fgb.2005.05.006.
29. **Wilson RB, Davis D, Mitchell AP.** 1999. Rapid hypothesis testing with *Candida albicans* through gene disruption with short homology regions. J Bacteriol **181:**1868–1874.
30. **Murad AM, Lee PR, Broadbent ID, Barelle CJ, Brown AJ.** 2000. CIp10, an efficient and convenient integrating vector for Candida albicans. Yeast **16:**325–327. http://dx.doi.org/10.1002/1097-0061(20000315)16:4<325::AID-YEA538>3.0.CO;2-#.
31. **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. http://dx.doi.org/10.1111/j.1567-1364.2009.00542.x.
32. **Binkley J, Arnaud MB, Inglis DO, Skrzypek MS, Shah P, Wymore F, Binkley G, Miyasato SR, Simison M, Sherlock G.** 2014. The Candida Genome Database: the new homology information page highlights protein similarity and phylogeny. Nucleic Acids Res **42:**D711–D716. http://dx.doi.org/10.1093/nar/gkt1046.
33. **van het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, Hogues H, Cuomo C, Berriman M, Scherer S, Magee BB, Whiteway M, Chibana H, Nantel A, Magee PT.** 2007. Assembly of the Candida albicans genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol **8:**R52. http://dx.doi.org/10.1186/gb-2007-8-4-r52.
34. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:**1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.
35. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:**1297–1303. http://dx.doi.org/10.1101/gr.107524.110.
36. **Moreno-Ruiz E, Ortu G, de Groot PW, Cottier F, Loussert C, Prevost MC, de Koster C, Klis FM, Goyard S, d'Enfert C.** 2009. The GPI-modified proteins Pga59 and Pga62 of Candida albicans are required for cell wall integrity. Microbiology **155:**2004–2020. http://dx.doi.org/10.1099/mic.0.028902-0.
37. **Stynen B, Van Dijck P, Tournu H.** 2010. A CUG codon adapted two-hybrid system for the pathogenic fungus Candida albicans. Nucleic Acids Res **38:**e184. http://dx.doi.org/10.1093/nar/gkq725.
38. **Cormack BP, Bertram G, Egerton M, Gow NA, Falkow S, Brown AJ.** 1997. Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans. Microbiology **143**(Pt 2)**:**303–311. http://dx.doi.org/10.1099/00221287-143-2-303.
39. **Kato R, Ogawa H.** 1994. An essential gene, ESR1, is required for mitotic cell growth, DNA repair and meiotic recombination in Saccharomyces cerevisiae. Nucleic Acids Res **22:**3104–3112. http://dx.doi.org/10.1093/nar/22.15.3104.
40. **Zhou BB, Elledge SJ.** 2000. The DNA damage response: putting checkpoints in perspective. Nature **408:**433–439. http://dx.doi.org/10.1038/35044005.
41. **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. http://dx.doi.org/10.1186/gb-2013-14-9-r97.
42. **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. http://dx.doi.org/10.1016/j.fgb.2007.08.008.
43. **Chou H, Glory A, Bachewich C.** 2011. Orthologues of the anaphase-promoting complex/cyclosome coactivators Cdc20p and Cdh1p are important for mitotic progression and morphogenesis in *Candida albicans*. Eukaryot Cell **10:**696–709. http://dx.doi.org/10.1128/EC.00263-10.
44. **García-Prieto F, Gómez-Raja J, Andaluz E, Calderone R, Larriba G.** 2010. Role of the homologous recombination genes RAD51 and RAD59 in the resistance of Candida albicans to UV light, radiomimetic and antitumor compounds and oxidizing agents. Fungal Genet Biol **47:**433–445. http://dx.doi.org/10.1016/j.fgb.2010.02.007.
45. **Shi QM, Wang YM, Zheng XD, Lee RT, Wang Y.** 2007. Critical role of DNA checkpoints in mediating genotoxic-stress-induced filamentous growth in Candida albicans. Mol Biol Cell **18:**815–826. http://dx.doi.org/10.1091/mbc.E06-05-0442.
46. **Stevenson LF, Kennedy BK, Harlow E.** 2001. A large-scale overexpression screen in Saccharomyces cerevisiae identifies previously uncharacterized cell cycle genes. Proc Natl Acad Sci U S A **98:**3946–3951. http://dx.doi.org/10.1073/pnas.051013498.
47. **Yu L, Pena Castillo L, Mnaimneh S, Hughes TR, Brown GW.** 2006. A survey of essential gene function in the yeast cell division cycle. Mol Biol Cell **17:**4736–4747. http://dx.doi.org/10.1091/mbc.E06-04-0368.
48. **Paffett KS, Clikeman JA, Palmer S, Nickoloff JA.** 2005. Overexpression of Rad51 inhibits double-strand break-induced homologous recombination but does not affect gene conversion tract lengths. DNA Repair **4:**687–698. http://dx.doi.org/10.1016/j.dnarep.2005.03.003.
49. **Richardson C, Stark JM, Ommundsen M, Jasin M.** 2004. Rad51 overexpression promotes alternative double-strand break repair pathways and genome instability. Oncogene **23:**546–553. http://dx.doi.org/10.1038/sj.onc.1207098.
50. **Kapitzky L, Beltrao P, Berens TJ, Gassner N, Zhou C, Wuster A, Wu J, Babu MM, Elledge SJ, Toczyski D, Lokey RS, Krogan NJ.** 2010. Cross-species chemogenomic profiling reveals evolutionarily conserved drug mode of action. Mol Syst Biol **6:**451. http://dx.doi.org/10.1038/msb.2010.107.
51. **Yuen KW, Warren CD, Chen O, Kwok T, Hieter P, Spencer FA.** 2007. Systematic genome instability screens in yeast and their potential relevance to cancer. Proc Natl Acad Sci U S A **104:**3925–3930. http://dx.doi.org/10.1073/pnas.0610642104.
52. **Schwartz K, Richards K, Botstein D.** 1997. BIM1 encodes a microtubule-binding protein in yeast. Mol Biol Cell **8:**2677–2691. http://dx.doi.org/10.1091/mbc.8.12.2677.
53. **Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B.** 2006. Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell **21:**319–330. http://dx.doi.org/10.1016/j.molcel.2005.12.011.
54. **Yoshikawa K, Tanaka T, Ida Y, Furusawa C, Hirasawa T, Shimizu H.** 2011. Comprehensive phenotypic analysis of single-gene deletion and overexpression strains of Saccharomyces cerevisiae. Yeast **28:**349–361. http://dx.doi.org/10.1002/yea.1843.
55. **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. http://dx.doi.org/10.1111/j.1365-2958.2011.07532.x.
56. **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. http://dx.doi.org/10.1371/journal.pbio.0060110.

57. **Burke D, Gasdaska P, Hartwell L.** 1989. Dominant effects of tubulin overexpression in *Saccharomyces cerevisiae*. Mol Cell Biol **9:**1049–1059.
58. **Harper JW, Elledge SJ.** 2007. The DNA damage response: ten years after. Mol Cell **28:**739–745. http://dx.doi.org/10.1016/j.molcel.2007.11.015.
59. **Pfau SJ, Amon A.** 2012. Chromosomal instability and aneuploidy in cancer: from yeast to man. EMBO Rep **13:**515–527. http://dx.doi.org/10.1038/embor.2012.65.
60. **Negrini S, Gorgoulis VG, Halazonetis TD.** 2010. Genomic instability—an evolving hallmark of cancer. Nat Rev Mol Cell Biol **11:**220–228. http://dx.doi.org/10.1038/nrm2858.
61. **Kolodner RD, Putnam CD, Myung K.** 2002. Maintenance of genome stability in Saccharomyces cerevisiae. Science **297:**552–557. http://dx.doi.org/10.1126/science.1075277.
62. **Bennett CB, Lewis LK, Karthikeyan G, Lobachev KS, Jin YH, Sterling JF, Snipe JR, Resnick MA.** 2001. Genes required for ionizing radiation resistance in yeast. Nat Genet **29:**426–434. http://dx.doi.org/10.1038/ng778.
63. **Birrell GW, Giaever G, Chu AM, Davis RW, Brown JM.** 2001. A genome-wide screen in Saccharomyces cerevisiae for genes affecting UV radiation sensitivity. Proc Natl Acad Sci U S A **98:**12608–12613. http://dx.doi.org/10.1073/pnas.231366398.
64. **Chang M, Bellaoui M, Boone C, Brown GW.** 2002. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. Proc Natl Acad Sci U S A **99:**16934–16939. http://dx.doi.org/10.1073/pnas.262669299.
65. **Hanway D, Chin JK, Xia G, Oshiro G, Winzeler EA, Romesberg FE.** 2002. Previously uncharacterized genes in the UV- and MMS-induced DNA damage response in yeast. Proc Natl Acad Sci U S A **99:**10605–10610. http://dx.doi.org/10.1073/pnas.152264899.
66. **Bellaoui M, Chang M, Ou J, Xu H, Boone C, Brown GW.** 2003. Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. EMBO J **22:**4304–4313. http://dx.doi.org/10.1093/emboj/cdg406.
67. **Mayer ML, Pot I, Chang M, Xu H, Aneliunas V, Kwok T, Newitt R, Aebersold R, Boone C, Brown GW, Hieter P.** 2004. Identification of protein complexes required for efficient sister chromatid cohesion. Mol Biol Cell **15:**1736–1745. http://dx.doi.org/10.1091/mbc.E03-08-0619.
68. **Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C.** 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science **294:**2364–2368. http://dx.doi.org/10.1126/science.1065810.
69. **Warren CD, Eckley DM, Lee MS, Hanna JS, Hughes A, Peyser B, Jie C, Irizarry R, Spencer FA.** 2004. S-phase checkpoint genes safeguard high-fidelity sister chromatid cohesion. Mol Biol Cell **15:**1724–1735. http://dx.doi.org/10.1091/mbc.E03-09-0637.
70. **Huang ME, Rio AG, Nicolas A, Kolodner RD.** 2003. A genomewide screen in Saccharomyces cerevisiae for genes that suppress the accumulation of mutations. Proc Natl Acad Sci U S A **100:**11529–11534. http://dx.doi.org/10.1073/pnas.2035018100.
71. **Alabrudzinska M, Skoneczny M, Skoneczna A.** 2011. Diploid-specific [corrected] genome stability genes of S. cerevisiae: genomic screen reveals haploidization as an escape from persisting DNA rearrangement stress. PLoS One **6:**e21124. http://dx.doi.org/10.1371/journal.pone.0021124.
72. **Andersen MP, Nelson ZW, Hetrick ED, Gottschling DE.** 2008. A genetic screen for increased loss of heterozygosity in Saccharomyces cerevisiae. Genetics **179:**1179–1195. http://dx.doi.org/10.1534/genetics.108.089250.
73. **Kanellis P, Gagliardi M, Banath JP, Szilard RK, Nakada S, Galicia S, Sweeney FD, Cabelof DC, Olive PL, Durocher D.** 2007. A screen for suppressors of gross chromosomal rearrangements identifies a conserved role for PLP in preventing DNA lesions. PLoS Genet **3:**e134. http://dx.doi.org/10.1371/journal.pgen.0030134.
74. **Smith S, Hwang JY, Banerjee S, Majeed A, Gupta A, Myung K.** 2004. Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A **101:**9039–9044. http://dx.doi.org/10.1073/pnas.0403093101.
75. **Strome ED, Wu X, Kimmel M, Plon SE.** 2008. Heterozygous screen in Saccharomyces cerevisiae identifies dosage-sensitive genes that affect chromosome stability. Genetics **178:**1193–1207. http://dx.doi.org/10.1534/genetics.107.084103.
76. **Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A.** 2007. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. Science **317:**916–924. http://dx.doi.org/10.1126/science.1142210.

Summary

The *Candida albicans* genome displays a high tolerance to rearrangements, notably loss-of-heterozygosity (LOH) events. These events most often result from the repair of DNA double-strand breaks (DSBs) and are known to play an important role in different aspects of *C. albicans* biology.

To study the molecular mechanisms leading to LOH, we have combined an I-*Sce*I meganuclease-dependent DSB-inducing system with a FACS-optimized reporter system of LOH. Our results show that expression of I-*Sce*I leads to a dramatic increase in the frequency of LOH events, mainly gene conversion events. Characterization of cells having undergone a LOH led us to identify recessive lethal and deleterious alleles present in the heterozygous state in the genome of the *C. albicans* laboratory strain. These alleles influence the nature of the LOH events that are observed following a DSB.

We also characterized the fidelity of the repair following an I-*Sce*I-induced DNA DSB. This revealed unexpected complex recombination events occurring upon both break-induced replication and gene conversion with crossover repair events.

In parallel, a collection of 564 overexpression plasmids for genes involved in signaling pathways, genome integrity and cell wall integrity has been transformed in a *C. albicans* strain harboring the I-*Sce*I-dependent DSB-inducing system and FACS-optimized LOH reporter. Analyses of the resulting transformants under conditions that allowed fro I-*Sce*I expression or not, led to the identification of genes whose overexpression results either in an increase of the basal LOH rate or in a reduction of the high LOH frequency associated to I-*Sce*I expression.

Résumé

La levure *Candida albicans* présente une tolérance élevée aux réarrangements de son génome et en particulier, aux pertes d'hétérozygotie (LOH). Ces LOH sont le plus souvent le résultat de la réparation d'une cassure double-brin de l'ADN (DNA DSB) et jouent un rôle important dans différents aspects de la biologie de *C. albicans*.

Afin d'étudier les mécanismes moléculaires à l'origine des LOH, nous avons combiné un système d'induction d'un DNA DSB par la méganucléase I-*Sce*I et un système rapporteur de LOH optimisé pour l'analyse par FACS. La surexpression de I-*Sce*I entraine une forte augmentation du taux de LOH, principalement des conversions géniques. La caractérisation des cellules ayant subi une LOH a permis d'identifier des allèles récessifs délétères et létaux présents à l'état hétérozygote dans le génome de la souche de laboratoire de *C. albicans*. Ces allèles influent sur la nature des LOH observés suite à un DNA DSB.

Nous avons également caractérisé la fidélité de la réparation d'un DNA DSB induit par I-*Sce*I chez *C. albicans*. Cette étude a permis de décrire des recombinaisons complexes et inattendues se déroulant pendant les événements de break-induced replication et de conversions géniques avec crossover.

En parallèle, une collection de 564 plasmides de surexpression pour des gènes impliqués dans les voies de signalisation et dans l'intégrité du génome et de la paroi a été transformée dans une souche possédant les deux systèmes mentionnés ci-dessus. L'analyse des transformants dans des conditions ou non d'expression de I-*Sce*I a permis d'identifier des gènes dont la surexpression augmente le taux basal de LOH ou diminue le taux de LOH élevé du à l'expression d'I-*Sce*I.