A FACS-optimized screen identifies regulators of genome stability in Candida albicans.
Raphaël Loll-Krippleber, Adeline Feri, Marie Nguyen, Corinne Mauvais, Jennifer Yansouni, Christophe d’Enfert, Mélanie Legrand

To cite this version:

HAL Id: pasteur-01515479
https://hal-pasteur.archives-ouvertes.fr/pasteur-01515479
Submitted on 27 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives| 4.0 International License
A FACS-optimized screen identifies regulators of genome stability in *Candida albicans*

Raphaël Loll-Krippleber\textsuperscript{a, b, c*}, Adeline Feri\textsuperscript{a, b, c}, Marie Nguyen\textsuperscript{d}, Corinne Maufrais\textsuperscript{e}, Jennifer Yansouni\textsuperscript{f}, Christophe d’Enfert\textsuperscript{a, b} and Mélanie Legrand\textsuperscript{a, b##}

Unité Biologie et Pathogénicité Fongiques, Département Mycologie, Institut Pasteur, Paris, France\textsuperscript{a}; INRA USC2019, Paris, France\textsuperscript{b}; Univ. Paris Diderot, Sorbonne Paris Cité, Paris, France\textsuperscript{c}; Plate-Forme de Cytométrie, Imagopole, Institut Pasteur, Paris, France\textsuperscript{d}; Centre d’Informatique en Biologie, Institut Pasteur, Paris, France\textsuperscript{e}; BIOASTER, Genomics and transcriptomics, Lyon, France\textsuperscript{f}

Running head: Loss Of Heterozygosity screen in *Candida albicans*

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

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

**ABSTRACT**
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 BFP and GFP fluorescent 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 implementing new screens to identify genes that are important for genome stability in C. albicans and more generally in eukaryotic cells.

INTRODUCTION

Normally found as a harmless commensal organism, Candida albicans is also the 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 mostly found as a diploid organism, haploid and tetraploid forms have been observed in the laboratory and upon passage of C. albicans in animal models of infection (2-4). The formation of tetraploids results from mating between diploids of opposite mating type 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 LOH frequencies 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 knock-out mutations in genes involved in base-excision repair, nucleotide-excision repair and mismatch repair had little impact on LOH frequencies 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 frequencies of LOH (13, 15, 16).

Evaluation of LOH frequencies in C. albicans has up to now relied on the use of several counter-selectable 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, measuring the number of 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 LOH frequencies in these strains. Although robust, these systems are labor-intensive, 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 knock-out constructs or overexpression plasmids, such screens are becoming amenable to *C. albicans* (20-22) and could provide new insights in the mechanisms that control genome stability in this species and other eukaryotes.

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 LOH frequencies triggered by different physical and chemical stresses, and knock-out mutations. Moreover, using a newly developed collection of overexpression plasmids for 124 genes whose orthologs in *S. 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 *C. albicans* success as a commensal and 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 *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.

For the LOH assay by FACS, aliquots of prototrophic cells were taken from the -80°C stock and grown in 96-well plates overnight in YPD at 30°C with rotatory shaking for recovery.
The next day, these cultures were inoculated in SC-His-Arg to eliminate the cells that had undergone LOH and have lost the BFP-HIS1 or GFP-ARG4 cassettes as a result of the freezing/thawing step. The SC-His-Arg cultures were grown overnight at 30°C with rotatory shaking. At this 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, Fisher Bioblock Scientific, 3 μg/ml) in YPD or SD 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 overnight at 30°C, 37°C or 39°C and allowed to recover overnight in YPD before flow cytometry analysis.

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

**Plasmid constructions**

Oligonucleotides used in this work are listed in Tables S1 and S2 in the supplemental material. Plasmids used or generated in this work are listed in Table S3 in the supplemental material. Plasmid constructions are described in the supplemental material (Text S1) and yielded six plasmids harboring 1) either the GFP or BFP genes placed under the control of the constitutively-active promoter of the C. albicans TDH3 gene (25) and 2) either the ARG4 (26), HIS1 (27), URA3 (26) or SAT1 (27) transformation markers (Figure S1).

**C. albicans strains construction**

*Multiple heterozygous markers strain construction*
One allele of the \textit{GAL1} gene was replaced with the \textit{CmLEU} marker gene by PCR-mediated transformation of SN148 (24). The resulting strain was named CEC1027. The $P_{TDH3}$-$BFP$-$HISI$ 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 homologous sequences to the genomic DNA located between \textit{PGA59} and \textit{PGA62} on chromosome 4 (Ch4). The PCR products were sequentially transformed in CEC1027 according to Gola et al. (26) yielding CEC2814. For activation of the $P_{TET}$ promoter, strain CEC2814 was eventually transformed with \textit{SacII,KpnI}-digested pNIMX that 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) using oligonucleotides K7 yFP-Fwd and K7 yFP-Rev. CEC2683 was subsequently transformed with \textit{StuI-linearized CIp30} (28) to generate CEC3194, the prototrophic parental strain.

A \textit{mec1} null mutant was constructed by replacing both alleles with the \textit{LEU2} and \textit{HIS1} markers using a PCR-based method (29) in CEC2683, yielding first the heterozygous $\textit{MEC1/mec1}$ mutant CEC3172 and then the null $\textit{mec1/mec1}$ mutant CEC3183. CEC3172 and CEC3183 were then transformed with \textit{StuI-linearized CIp30} (28) or CIp10 (30) to generate CEC3203 and CEC3216, the prototrophic \textit{mec1} heterozygous and homozygous mutants.

\textit{C. albicans} overexpression strains

\textit{StuI}-digested or \textit{I-SceI}-digested CIp10-$P_{TET}$-GTW derivatives were transformed into strain CEC2824 as described in Chauvel et al. (21), resulting in 124 overexpression strains (see Table S2 in the supplemental material) which together with the CEC3989 control strain derived from CEC2824 transformed with \textit{StuI-linearized CIp10-$P_{TET}$-GTW} will be referred to as the 3R overexpression collection.
All *C. albicans* transformants were verified by PCR using primers inside the transformed plasmid/cassette and primers in the genomic DNA regions of insertion in order to assess proper integration of the plasmid/cassette in the *C. albicans* genome.

**Flow cytometry analysis**

Cells from overnight cultures were diluted at 1:1000 in 1X PBS in BD Falcon tubes (Ref 352054) and $10^6$ cells were analyzed by flow cytometry in 96-well plate format using a MACSQuant (Miltenyi) flow cytometer. We used the 405 nm 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**

*Primary screen*

Strains in the 3R overexpression collection were screened for LOH frequencies by flow cytometry following overnight growth at 30°C in YPD 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 sample and applied these gates to the rest of the data set for the control cultures. The percentage of Bfp+ Gfp- cells and Bfp- Gfp+ cells was exported in excel format. The mean was calculated, as well as the standard deviation (SD). The same gates were applied to all the mutants, resulting in a percentage of Bfp+ Gfp- cells and Bfp- Gfp+ cells for each mutant strain. For each mutant, a Z-score was calculated and we selected mutants that showed a Z-score superior to 5 for the Bfp- Gfp+ population (corresponding to five standard deviations from the mean value of the control strain).
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 mutants identified in the primary screen were grown in triplicates in YPD and YPD+ATc (3 μg/ml) for 8 hours and re-inoculated in YPD alone to recover overnight at 30°C. The next day, 10⁶ 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 hours and allowed to recover overnight in YPD alone. Fluorescence-activated cell sorting of these cultures, using a FACSaria III cell sorter (BD Biosciences), was performed into 1.5 ml Eppendorf tubes at a rate of 10,000 evt/s. 1000 cells from the Bfp- Gfp+ and Bfp+ Gfp- populations were sorted, recovered in 400 μl YPD, and immediately plated on YPD plates. The plates were incubated at 30°C for 2 days. For the Bfp- Gfp+ population, 48 colonies were inoculated in 800 μl YPD in 96-well plates and grown overnight at 30°C. For the Bfp+ Gfp- population, we observed a colony size phenotype and therefore 24 small colonies and 24 large colonies were inoculated in 800 μl YPD in 96-well plates and grown overnight at 30°C. Cultures were then analyzed by flow cytometry to validate mono-fluorescence and spotted on YPD, 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 SNP-typing of 16 isolates, when possible. Genomic DNA was extracted from the sorted clones and PCR-RFLP was performed.
to assess the extent of the LOH. We used SNPs 95 and 156, which are part of \textit{Alu}I and \textit{Taq}I 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 \textit{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 base pairs and Illumina adapters were blunt-end ligated to the fragments. The final libraries were amplified by PCR followed by sequencing on a MiSeq platform, using the v3 chemistry. 300 nt paired-end reads were aligned to the \textit{C. albicans} strain SC5314 reference genome assembly 21, available from CGD (32, 33) using BWA (34). Single nucleotide variants (SNVs) between the sequenced genomes and the reference genome were identified using GATK (35) at positions with a sequencing depth equal or above 18X. 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 in Loll-Krippleber et al. (16). Homozygous SNP density maps were constructed by counting the number of homozygous positions per 10 kb region and plotting these values.

RESULTS
A new reporter system for LOH
As an alternative to the \textit{GAL1} and \textit{URA3} systems routinely used for LOH studies in \textit{C. albicans} (12, 18), we developed a new system combining the use of fluorescent markers and flow cytometry, more suitable for high-throughput studies. This system relies on the construction of an artificial locus on Ch4 in the intergenic region between the \textit{PGA59} and \textit{PGA62} ORFs (36). This 9 kb region has been used previously as a platform for integration of \textit{C. albicans} two-hybrid plasmids (37) and its modification is thought to be neutral to \textit{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 coding for the Green Fluorescent Protein (GFP; (38)) under the control of the promoter of the constitutively, highly-expressed \textit{TDH3} gene (25) and linked to an auxotrophic marker (\textit{ARG4, HIS1, URA3}) or the nourseothricin resistance marker (\textit{SAT1}). The other homolog was engineered to carry the gene coding for the Blue Fluorescent Protein (BFP; Evrogen) placed under the control of the \textit{TDH3} promoter and linked to another auxotrophic marker or the nourseothricin resistance marker. A series of cassettes have been developed that associate the \textit{BFP} or \textit{GFP} genes to different selection markers (\textit{SAT1, HIS1, URA3} and \textit{ARG4}), allowing their use in various genetic backgrounds (see Fig. S1 in the supplemental material). In this set-up, LOH events can be detected by flow cytometry since cells that undergo LOH at the \textit{PGA59-62} locus will produce only one of the two fluorescent proteins. As shown in Figure 1B, FACS analysis of strain CEC2683, which carries the \textit{BFP-SAT1/GFP-ARG4} system showed that the majority of the cells (>99%) co-expressed 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 from 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. 95% (72/76) and 94% (49/52) of the cells recovered from both populations were mono-fluorescent (data not
shown). Moreover, spotting assays on appropriate media indicated that all mono-fluorescent 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 2-3×10^{-4} at this locus. Interestingly, the cells recovered from the Bfp- Gfp- population did not grow, indicating that these correspond to non-viable cells. Fluorescence plots of mono-BFP (CEC4007), mono-GFP (CEC3998), unlabeled (SN100) and BFP/GFP-labeled strains are presented in Figure S2A in the supplemental material.

LOH increase at the PGA59-62 locus in mecl 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 Figure 2A and Table 2 showed a statistically significant increase in the proportion of mono-GFP or mono-BFP cells in mecl null mutants as compared to the wild-type strain. Interestingly, we also observed an increase in the proportion of Bfp- Gfp- cells, suggesting that inactivation of MEC1 increases cell mortality (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 LOH frequencies by flow cytometry upon inactivation of the MEC1 gene.

LOH increase at the PGA59-62 locus in response to different stresses
Forche et al. (12) have shown an increase in 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°C, 37°C or 39°C. Although cells were filamenting after 7 hours growth at 37°C or 39°C, microscopic analysis of cells after an overnight growth at 37°C 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 before flow cytometry analysis. Similarly 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 Figure S2B confirmed that the distribution of mono-labeled cells remains the same in absence or presence of heat stress.

We also treated the control strain carrying the BFP/GFP system with the genotoxic agent MMS (Methyl Methanesulfonate). After a 30-minute treatment at different concentrations of MMS, the cells were allowed to recover in rich medium overnight. Flow cytometry analysis revealed a LOH frequency 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 to detect increases in LOH frequencies resulting from physical or chemical stresses.

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 wild-type LOH frequencies. To this aim, we generated a collection of C. albicans
overexpression strains focused on genes which, based on their GO annotations or 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, 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 in *C. albicans* strain CEC2824 that 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 OE strains were obtained indicating an overall 60.8% success rate, slightly below that reported previously by Chauvel et al. (21) for a non overlapping set of overexpression plasmids, possibly because of the larger size of the cloned ORFs in our study (27% above 2500 bp vs 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 will be referred to below as the 3R overexpression collection.

### FACS-optimized LOH screens upon gene overexpression

Prior to analyzing the overexpression mutants for LOH frequencies under inducing conditions, we evaluated the LOH frequency in the 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 analyzed and the mean frequency of LOH and 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 an overnight induction. Thirty-three overexpression strains were selected that had a Z-score (for the frequency of Bfp- Gfp+ cells) above 5 (see Fig. S3 in the supplemental material). In the secondary screen, we allowed the cells to recover in rich media after an 8 hours induction in YPD + ATc3 and re-analyzed the 33 overexpression strains identified in the primary screen in triplicates. While we saw no difference in Bfp+ Gfp- and Bfp- Gfp+ frequencies between the control strain and the overexpression mutants in absence of ATc treatment, we confirmed the increase in LOH at the PGA59-62 locus after ATc treatment for 4 of these genes, namely $BIM1$ (10-fold for the Bfp+ Gfp- cells and 13-fold for the 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 LOH frequency possibly owing to the relatively low LOH frequency 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) or break-induced replication (BIR), chromosome truncation or chromosome loss. Further to identifying genes whose overexpression increased LOH frequencies, we tested whether overexpression of these genes could also alter the balance between these different types of events as observed in the control strain.

Fluorescence-activated cell sorting was used in order to recover individual mono-fluorescent cells and subsequently analyze these cells by SNP-RFLP typing that can reveal the molecular mechanisms – gene conversion, MCO-BIR or chromosome truncation, chromosome loss – underlying the LOH event responsible for the loss of the BFP or GFP marker on Ch4 (Fig.
Cells sorted from the Bfp- Gfp+ control and mutant populations and plated on YPD 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 arisen at somewhat equal frequency through BIR/MCO or chromosome truncation, these two types of events being undistinguishable by SNP-RFLP typing, and gene conversion (Fig. 3B). A complete loss of the homolog carrying the BFP marker was never observed. A similar pattern was observed upon overexpression of CDC20 and RAD53. In contrast, we detected an increase in the proportion of BIR/MCO or chromosome truncation events versus gene conversions resulting in the loss of the BFP marker in the RAD51 and BIM1 overexpression mutants (Fig. 3B).

Notably, colonies of different size were observed when cells sorted from the Bfp+ Gfp- populations were plated on YPD (data not shown). This included colonies of a size similar to those obtained from the Bfp- Gfp+ populations and smaller colonies. This small colony phenotype was observed for the control strain and all 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 the Bfp- Gfp+ population, chromosome loss was the major mechanism responsible for the loss of the GFP marker in the small colony-derived cells in 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 3E). SNP-RFLP typing data for
both types of colonies were normalized to the proportion of small vs large colonies and combined to represent the percentage of each molecular mechanism (gene conversion, BIR/MCO or chromosome truncation, chromosome 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. In order to determine the extent of the LOH, we performed whole genome sequencing of 3 small Bfp+ Gfp- colonies obtained from an overexpression mutant that did not show any change in LOH frequency relative to the control and in which LOH were 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 3 small colonies had lost the entire Ch4B carrying the GFP marker (Fig. 3G). The observation that Ch4 had remained disomic (data not shown) indicates that the other homolog 4A has been reduplicated. Homozygosis of Ch4A also accounts for the histidine auxotrophy displayed by the 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 only observed for Ch4B; (ii) this was reflected by the occurrence of small colony size clones when all or part of Ch4B was lost, (iii) when chromosome loss occurred, overexpression of CDC20 favored this mechanism, (iv) RAD51 overexpression favored BIR/MCO events, (v) BIM1 overexpression favored BIR/MCO events
unless chromosome loss was possible; and (vi) \textit{RAD53} overexpression favored BIR/MCO events when chromosome loss was possible.

**DISCUSSION**

Here, we have presented a FACS-optimized genetic system to detect LOH in \textit{C. albicans} and its application to the identification of genes whose overexpression results in an increase in LOH frequency and changes in the frequencies of the molecular events that are at the origin of LOH.

The new BFP/GFP system displays appealing features over other genetic systems currently used to study LOH in \textit{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 \textit{GAL1} or \textit{URA3} markers, which require large amounts of expensive drugs, the BFP/GFP system does not require any costly consumables. The use of drugs such as 2-deoxygalactose or 5-fluoroorotic acid also raises the issue of exposing the cells to a selective pressure, which might distort the evaluation of LOH frequencies. In contrast, the BFP/GFP system allows the measurement of spontaneous LOH frequencies since cells are not exposed to any stress. As 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 LOH frequencies 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 \textit{C. albicans} overexpression mutants to identify genes whose overexpression would result in an increase in LOH frequency. Our screen identified 4 candidates as regulators of genome stability in \textit{C. albicans}. Null or conditional deletion mutants have been obtained and studied for 3 of these genes: \textit{CDC20} (43), \textit{RAD51} (44) and \textit{RAD53} (45). Yet, the role of these genes in genome integrity...
was never addressed, except for **RAD53** whose deletion increases LOH frequency (16).

Reports of *C. albicans* overexpression mutants of the 4 candidates genes could not be found in the literature. In contrast, null or conditional deletion mutants, as well as overexpression mutants, have been described in *S. cerevisiae* for the 4 candidate genes. Although genome stability was investigated in the deletion mutants, the question of genome integrity was not addressed in the overexpression mutants.

Our screen identified **CDC20** as the gene whose overexpression triggers the highest increase in LOH frequency. In *S. cerevisiae*, the essential **CDC20** gene encodes an activator of ubiquitin-protein ligase activity and is required for metaphase/anaphase transition. At this stage of the mitotic cycle, Cdc20p enables the APC (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 to maintain the spindle assembly checkpoint (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 have been characterized in *C. albicans* by Chou et al. (43). Their work suggested that, similarly 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 and DNA repair, were identified as causing an increase in LOH upon overexpression. Rad53 is a kinase involved in the DNA-damage checkpoint whose deletion in *C. albicans* causes a growth defect, an increased sensitivity to several DNA-damaging agents and a defect in genotoxic-stress induced filamentation (45). An increase in LOH could also be observed in **CaRAD53** deletion
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 \textit{RAD51} in \textit{C. albicans} results in a decreased growth rate and an increased sensitivity to various DNA-damaging agents (44). On the other hand, genome integrity has not been investigated in the \textit{CaRAD51} deletion mutants. In \textit{S. cerevisiae}, \textit{RAD51} overexpression has been shown to promote genome instability, probably by inhibiting the accurate repair of double-strand breaks by homologous recombination (48). Similarly, \textit{RAD51} overexpression in the mouse leads to an increase in genome instability, notably LOH events (49). Our observation that, in addition to increasing LOH frequency, overexpression of \textit{RAD51} and \textit{RAD53} favored BIR/MCO 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 \textit{BIM1} has been identified as increasing LOH frequency upon overexpression. While \textit{ScBIM1} deletion has been shown to result in increased sensitivity to various DNA-damaging agents (50) and increased chromosome instability (51), the effect of \textit{BIM1} overexpression on genome integrity has not been investigated, probably in part because of the strong growth defect associated with \textit{BIM1} overexpression in \textit{S. cerevisiae} (52-54). Similarly, we also observed a growth defect associated to \textit{BIM1} overexpression in \textit{C. albicans} (data not shown). Our observation that \textit{BIM1} overexpression, in addition to promoting LOH, favors BIR/MCO 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 \textit{BIM1} overexpression could favor the occurrence of DNA breaks and thus explain the increase in BIR/MCO events or chromosome truncations in the \textit{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 slower 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 8 chromosomes can occur. Nevertheless, a bias in the homolog being retained was reported for certain chromosomes (2, 55). This was the case for Ch4, for which one homolog 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 BIR/MCO 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 to refine the position of this or these hypothetical recessive lethal alleles. While homozygosis along the entire smaller chromosomes, Ch5, Ch6 and Ch7, were often observed, MCO-BIR events or chromosome truncations have been reported to be responsible for most LOH on 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 DSB 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 of Ch4 homologs, 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). Yet, 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 based on deletion mutant study, which does not
imply that their overexpression triggers an increase in LOH frequency. 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 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 in
model eukaryotic species such as S. cerevisiae have uncovered many components required for
genome maintenance and integrity (61). Genome-wide genetic screens in this species have
examined deletion mutants for 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 in this organism thus correlating with a high fitness cost (76). Interestingly, C. albicans is more tolerant to 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 Dr Carol Munro in Aberdeen (20), will allow C. albicans to become a model of choice to study 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). RL was the recipient of a PhD grant from INRA and Institut Pasteur. ML was the recipient of a postdoctoral fellowship from Institut Pasteur (Bourse Roux).

REFERENCES


59. **Pfau SJ, Amon A.** 2012. Chromosomal instability and aneuploidy in cancer: from yeast to man. EMBO reports **13:**515-527.


**FIGURE LEGENDS**

**Figure 1. The BFP/GFP system.**

(A) Artificial heterozygous locus on Chromosome 4. BFP and GFP expression cassettes were integrated in the gene-free PGA59/PGA62 intergenic region on chromosome 4. One homolog carries the gene coding for the BFP while the other carries the gene coding for the GFP. Both 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 media, diluted in 1X PBS and analyzed on a MACSQuant cytometer (Miltenyi Biotech). 10^6 events are displayed.

**Figure 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. The WT BFP/GFP strain CEC3194 and the mec1ΔΔ mutant were grown overnight in YPD at 30°C and analyzed by flow cytometry. (B) Flow cytometry analysis upon heat treatments. The WT BFP/GFP
strain CEC3989 was grown overnight in YPD at 30°C, 37°C or 39°C and allowed to recover overnight in YPD before flow cytometry analysis. (C) **Flow cytometry analysis in presence of the DNA-damaging agent MMS.** The WT BFP/GFP strain CEC3989 was treated with increasing concentrations of MMS for 30 minutes. Cells were then allowed to recover overnight in fresh YPD and analyzed by flow cytometry.

**Figure 3. SNP typing reveals a shift in the molecular mechanisms leading to LOH upon overexpression of some of the candidate genes.**

(A) **Map of Chromosome 4 and localization of the BFP/GFP system.** SNPs used for RFLP characterization are indicated with black triangles. The telomere proximal SNP156 is part of a TaqI restriction site. One allele contains the TaqI site while the other does not. The centromere proximal SNP95 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. (B) **SNP-RFLP typing in the Bfp- Gfp+ cells.** Histogram presenting the proportion of BIR/MCO or chromosome truncation vs chromosome loss in the Bfp- Gfp+ population. BIR/MCO or chromosomal truncation events correspond to isolates that have maintained a heterozygous SNP95 and displayed a homozygous SNP156. Chromosome loss events correspond to isolates in which both SNP95 and SNP156 became homozygous. Gene conversion events correspond to isolates in which both SNP95 and SNP156 remained heterozygous. (C) **Proportion of small vs large colonies giving rise to true mono-BFP cells.** (D) **SNP-RFLP typing in the small colony-derived Bfp+ Gfp- cells.** (E) **SNP-RFLP typing in the large colony-derived Bfp+ Gfp- cells.** (F) **SNP-RFLP typing in the Bfp+ Gfp- cells.** The normalized percentage of each molecular mechanism, for instance gene conversion, was calculated as follow: normalized percentage GC = (% small colonies × % GC in small colonies) + (% large
colonies \times \% \text{GC in large colonies}. (G) **Heterozygous and homozygous SNV density maps** of 1 large and 3 small colony-derived cells that have undergone LOH at the BFP/GFP locus on Ch4. The number of heterozygous SNVs in each 10kb region along Ch4 was computed and is represented, revealing a 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 sequences and yielded similar heterozygous SNP density maps. Only the map resulting from mapping on HapB is shown (Left panel). The number of homozygous SNVs in each 10kb region along HapA (Middle panel) or HapB (Right panel) was computed and is represented. The high density of homozygous SNVs upon mapping on HapB indicates that chromosome loss, in the small colony-derived cells, affected Ch4B.

**TABLE FOOTNOTES**

**Table 1.** Yeast strains used in this study.

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

\(^a\) Three independent colonies were tested for the *mec1* mutants. Six independent cultures were tested for the physical and chemical stresses. Standard error mean (SEM) were determined with GraphPad Prism Software.

\(^b\) Ratio between the LOH frequency in a given mutant and that in the parental strain.

\(^c\) Ratio between the LOH frequency in presence of the stress and that in absence of the stress.

\(^d\) A Mann–Whitney test was performed to test if the BFP loss frequency in the different mutants were statistically different from the rate observed in the control strains/conditions.
Table 3. LOH quantification in the 4 candidate overexpression mutants validated in the secondary LOH screen.

a Ratio between the LOH frequency in a given mutant and that in the control strain CEC3989.