

DISSERTATION

ANALYSIS OF THE RELATIONSHIP BETWEEN GENOMIC INSTABILITY,  
HETEROZYGOSITY LEVELS AND PHENOTYPE  
IN *SACCHAROMYCES CEREVISIAE*

Submitted by

Nadia Maria Vieira Sampaio

Graduate Degree Program in Cell and Molecular Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2018

Doctoral Committee:

Advisor: Juan Lucas Argueso

Laurie A. Stargell  
John K. McKay  
Kenneth F. Reardon

Copyrighted by Nadia Maria Vieira Sampaio 2018

All Rights Reserved

## ABSTRACT

### ANALYSIS OF THE RELATIONSHIP BETWEEN GENOMIC INSTABILITY, HETEROZYGOSITY LEVELS AND PHENOTYPE IN *SACCHAROMYCES CEREVISIAE*

Understanding the forces that mediate genome evolution is a central problem in genetics, with implications for diverse processes that range from speciation, to biotechnological applications, to human disease. The central theme of my dissertation was the characterization of two forces, genomic instability and natural selection, that significantly impact genome structure by influencing the levels of genomic heterozygosity. While genomic instability processes can act to erode heterozygosity from the genome, natural selection may favor the maintenance of heterozygous alleles in cases where there is a positive correlation between heterozygosity and higher fitness.

In Chapter I, I reviewed different types of mitotic mutations that can result in the appearance of tracts of homozygosity in genomes and recent discoveries about the temporal accumulation of such events. I also introduce the concept of heterosis, a phenomenon characterized by a positive correlation between genomic heterozygosity and phenotype in many species, and its potential role in contributing to the long-term maintenance of genomic heterozygosity.

In Chapter II, I describe the characterization of a mechanism of systemic genomic instability in yeast that challenges the conventional model of gradual and independent accumulation of mutations. We showed that a subset of mitotic cells within

a population experience bursts of genomic instability, which results in multiple independent events of loss-of-heterozygosity (LOH) accumulating over one or a few generations of mitotic cell division. We named this outcome “systemic genomic instability”. The occurrence of this phenomenon was initially identified in the heterozygous yeast strain JAY270, and then validated in a conventional laboratory strain background, whose genome is almost fully homozygous. Elevated rates of coincident LOH was also observed in mutant strains incapable of entering meiosis, indicating cryptic initiation of meiotic recombination followed by return-to-growth in a few cells in the population was not responsible for the higher than expected rates of coincident LOH. This finding brings to light a novel and intriguing mechanism of genomic instability in yeast that has relevant parallels to bursts of accumulation of copy number alterations in the human genome, providing a powerful experimental model system to dissect the fundamental mechanisms responsible for the generation of rapid changes in chromosome structure.

In Chapter III, we explored the role that genomic heterozygosity plays on the superior industrial traits of the JAY270 strain. In the previous Chapter we showed that mitotic recombination leading to LOH occurs at a high frequency during JAY270’s clonal propagation. These LOH events act against the long-term maintenance of genomic heterozygosity, yet about 60% of JAY270’s genome has remained heterozygous over time. We hypothesized that specific heterozygous alleles may have a positive impact on the traits of this strain and therefore were maintained through selection. We generated a collection of inbred strains derived from JAY270, and assessed them phenotypically under different growth conditions. Our results demonstrated that genomic

heterozygosity indeed has a substantial impact on two important industrial traits of this strain – heat stress tolerance and growth kinetics. We identified several genomic regions potentially associated with those traits and conducted experiments to investigate the bulk contributions of heterozygosity blocks in three specific chromosomes. This study revealed candidate regions containing loci that potentially underlie important industrial traits of JAY270 and details on the extent to which heterozygosity may impact JAY270's genome evolution and phenotype.

The combined results of these research projects provide important insights about the role of genomic instability mechanisms and their phenotypic outcomes in determining genome evolution, contributing discoveries that may have important practical implications for diverse fields, including biotechnology, cancer development and evolution, as well as genome sciences.

## ACKNOWLEDGMENTS

The work presented in this dissertation reflects the support and contributions of many people. First, I am thankful to my advisor, Dr. Lucas Argueso. His guidance and continuous excitement for our career and research have guided me through graduate school and will certainly continue to guide my future career. It was a privilege and a joy to be trained by him.

I thank the members of my graduate advisory committee, Dr. John McKay, Dr. Kenneth Reardon and Dr. Laurie Stargell, for pushing me to think critically about my projects and for all suggestions that helped to improve the quality of my research. I am also profoundly grateful to Dr. Carol Wilusz, a brilliant teacher, scientist and mentor. All lessons I have learned from you will remain an inspiration for the rest of my scientific career. Thank you for your support. I would also like to thank Dr. Cris Argueso for her support and continuous encouragement. Dr. Chris Allen, Dr. Claudia Wiese, Dr. Howard Liber, Dr. Mark Stenglein, Dr. David Maranon, thank you for sharing your knowledge with me and for making my research much easier.

I am grateful for the help and support of all current and past members of the Argueso Lab. I would like to especially thank Dr. Hailey Conover Sedam. I could not have completed my PhD without your daily dose of laughter. Thank you for your friendship and for teaching me so much, about science and life! I also thank our lab angel, Ruth Watson. Thank you for all your help and dedication during the completion of my projects.

I would like to thank the Cell and Molecular Biology Program, a great intellectual and cooperative environment. I especially thank Charlene Spencer and Lori Williams for being always so patient and willing to help.

I would like to give special thanks to Marcelo Bassalo. This work would not have been possible without his continuous encouragement and love. Thank you for being the biggest supporter of my career. It has been incredibly fun to walk this journey by your side.

Finally, I would like to express my profound gratitude to my family. I thank my parents, Sergio and Nilza Sampaio, and my siblings, Andre and Ana Sampaio. Thank you for your love, for believing in me and for all the personal sacrifice you have made to give me the chance to pursue my career goals. This dissertation is dedicated to you.

## TABLE OF CONTENTS

ABSTRACT.....	II
ACKNOWLEDGMENTS .....	V
LIST OF TABLES .....	IX
LIST OF FIGURES.....	X
<b>CHAPTER I .....</b>	<b>1</b>
<b>Introduction and Background.....</b>	<b>1</b>
General classes of mitotic mutations .....	1
Mechanisms leading to loss-of-heterozygosity (LOH) .....	2
Incremental steps or large jumps in the accrual of chromosomal alterations? .....	5
A tug of war in the regulation of genomic heterozygosity in yeast .....	10
<b>REFERENCES .....</b>	<b>14</b>
<b>CHAPTER II .....</b>	<b>19</b>
<b>Mitotic systemic genomic instability in yeast .....</b>	<b>19</b>
Summary .....	19
Introduction.....	20
Results .....	22
Appearance of altered colony morphology derivatives of JAY270.....	22
Genetic basis of the rough colony phenotype.....	26
Analysis of Chr12 LOH in spontaneous rough colony isolates .....	27
Analysis of selected Chr12 LOH.....	31
Genome-wide analysis LOH and CNA in the JAY664 isolate .....	35
Investigation of systemic genomic instability .....	37
Validation and quantification of coincident LOH .....	41
Discussion .....	43
Rough colony phenotype and mutation in <i>ACE2</i> .....	43
mitSGL and precedents of coincident recombination .....	44
mitSGL-like observations in human disease.....	48
Possible mechanisms underlying mitSGL .....	49
Materials and Methods .....	52
Growth media .....	52
Yeast genetic backgrounds and procedures.....	52
Isolation of spontaneous rough colonies derived from JAY270 .....	53
Allele replacement and complementation tests .....	54
Construction of strains used in single and double LOH assays.....	54
Quantitative LOH rate assays.....	55
Genome Sequencing Analyses .....	56
Genotyping specific HetSNPs through PCR-RFLP and molecular karyotyping .....	62
<b>REFERENCES .....</b>	<b>63</b>

<b>CHAPTER III .....</b>	<b>69</b>
<b>Characterization of phenotypic consequences of heterozygosity in yeast.....</b>	<b>69</b>
Summary .....	69
Introduction.....	70
Results .....	75
Controlled reduction of heterozygosity in the JAY270 genome through inbreeding.....	75
Characterization of phenotypic variation in the inbred diploid collection.....	79
Identification of genomic regions associated with phenotypic variation .....	87
Phenotypic consequences of chromosome-scale LOH .....	89
Discussion .....	94
Heterozygosity and fitness in yeast hybrids.....	94
Cryptic phenotypic variation in JAY270 .....	95
UPD strategy for phenotypic testing of confined LOH .....	97
Materials and Methods .....	99
Growth media .....	99
Yeast genetic backgrounds and microbiology procedures .....	100
Construction of a collection of partial inbred diploids derived from JAY270 ..	100
Construction of a GFP-tagged JAY270 derivative .....	101
Phenotypic assessment of the inbred collection .....	101
Identification of allele combinations enriched in high and low fitness strains.	103
Construction of UPD strains .....	105
<b>REFERENCES .....</b>	<b>108</b>
<b>CHAPTER IV .....</b>	<b>114</b>
<b>Conclusions and future directions.....</b>	<b>114</b>
Systemic genomic instability and resulting rapid karyotype evolution.....	114
Effects of heterozygosity on phenotype and its influence in genome configuration	117
<b>REFERENCES .....</b>	<b>119</b>
<b>APPENDIX A: SUPPLEMENTAL FIGURES.....</b>	<b>120</b>
<b>APPENDIX B: SUPPLEMENTAL TABLES .....</b>	<b>140</b>

## LIST OF TABLES

<i>Table S2.1. Yeast strains used in this study.</i>	140
<i>Table S2.2. Oligonucleotides used in this study.</i>	143
<i>Table S2.3. List of phased JAY270 HetSNPs and hemizygous sequences and respective detection methods.</i>	145
<i>Table S2.4. Smooth and rough clones isolation specifics and analysis of chromosome size polymorphisms by PFGE.</i>	146
<i>Table S2.5. Summary of WGS analysis of rough colony isolates.</i>	149
<i>Table S3.1. Yeast strains used in this study.</i>	152
<i>Table S3.2. Oligonucleotides used in this study.</i>	155
<i>Table S3.3. Growth conditions tested through plate spotting assay.</i>	157
<i>Table S3.4. Summary of phenotypic data measure in competition assays (CA) and tolerance to high temperature assays (HT) for each inbred strain.</i>	158
<i>Table S3.5. List of phased JAY270 HetSNPs interrogated for confirmation of UPD strains.</i>	162
<i>Table S3.6. Details on genomic regions showing significant association to the growth kinetics and heat tolerance phenotypes.</i>	163

## LIST OF FIGURES

<i>Figure 1.1. Genetic outcomes of interhomolog mitotic homologous recombination. ....</i>	<i>5</i>
<i>Figure 1.2. Models of mutation accumulation. ....</i>	<i>8</i>
<i>Figure 1.3. The levels of heterozygosity in the yeast genome are regulated by the balance between the counter-acting forces of heterosis and loss-of-heterozygosity. ....</i>	<i>13</i>
<i>Figure 2.1. Smooth and rough colony morphologies, mother-daughter cell attachment, and phenotypes of diploids derived from mating specific haploids. ....</i>	<i>23</i>
<i>Figure 2.2. LOH tract maps of Chr12 from five original rough colony isolates. ....</i>	<i>30</i>
<i>Figure 2.3. Quantitative analyses of LOH. ....</i>	<i>33</i>
<i>Figure 2.4. Analysis of unselected chromosomal changes in JAY664. ....</i>	<i>36</i>
<i>Figure 2.5. Genome-wide map of LOH tracts in spontaneous rough colony isolates. ....</i>	<i>40</i>
<i>Figure 3.1. Panel of inbred diploids derived from JAY270. ....</i>	<i>79</i>
<i>Figure 3.2. Inbred strains show heterogeneous levels of tolerance to high temperature stress. ....</i>	<i>81</i>
<i>Figure 3.3. Phenotypic assessment through growth competition assays. ....</i>	<i>83</i>
<i>Figure 3.4. Genome-wide association analyses to identify potential loci underlying heat stress tolerance. ....</i>	<i>88</i>
<i>Figure 3.5. Construction and relative growth kinetics profiles of Chr-UPD strain pairs. ....</i>	<i>93</i>
<i>Figure S2.1. Draft map of heterozygosity in the JAY270 genome. ....</i>	<i>120</i>
<i>Figure S2.2. HetSNP co-segregation genetic mapping approach and ace2-A7 complementation tests. ....</i>	<i>121</i>
<i>Figure S2.3. Sanger DNA sequencing analysis of ACE2 locus and genome-wide copy number in JAY664. ....</i>	<i>122</i>
<i>Figure S2.4. PCR genotyping of LOH and CNV in JAY664. ....</i>	<i>123</i>
<i>Figure S2.5. Analysis of Chr12 LOH in selected smooth and rough clones. ....</i>	<i>125</i>
<i>Figure S2.6. Detailed maps of Chr12 LOH in the 25 sequenced spontaneous rough colony isolates. ....</i>	<i>127</i>
<i>Figure S2.7. PFGE profiles of 29 independent smooth derivatives of JAY270. ....</i>	<i>129</i>
<i>Figure S2.8. PFGE profiles of smooth clones obtained through two bottlenecking lineages of JAY270. ....</i>	<i>130</i>
<i>Figure S2.9. PFGE profile of 27 independent rough colony derivatives of JAY270. ....</i>	<i>133</i>
<i>Figure S3.1. Genotype distribution in the inbred collection. ....</i>	<i>134</i>
<i>Figure S3.2. Inbred strains show uniform growth at 30°C and wide variation in tolerance to high temperature stress at 39°C. ....</i>	<i>135</i>
<i>Figure S3.3. Flow cytometry data analysis. ....</i>	<i>136</i>
<i>Figure S3.4. Venn Diagrams. ....</i>	<i>137</i>
<i>Figure S3.5. Genotype confirmation of UPD strains. ....</i>	<i>138</i>

## CHAPTER I

### **Introduction and Background**

#### **General classes of mitotic mutations**

A genome contains all of the information necessary for cellular function. Cells devote a repertoire of enzymatic pathways to accurately replicate and maintain their DNA molecules and the information encoded by its nucleotide sequence. Even though these pathways are extremely reliable, failure to faithfully replicate the DNA or to appropriately repair DNA damage does occur occasionally, and can lead to mutations that alter the DNA sequence and, consequently, the way cells behave and interact with an ever-changing environment.

When mutations occur in the germline they are passed on to the offspring and have long-term consequences on a species. On the other hand, mutations that occur in mitotically dividing somatic cells are not heritable, but they may have important functional outcomes to a cell's clonal lineage. For instance, mutations that occur in somatic cells are an underlying cause of cancer in humans (Hanahan and Weinberg 2000) and, in single celled microorganisms, they have a significant impact in adaptive evolution to new environmental conditions (Payen et al. 2016; Venkataram et al. 2016).

Spontaneous single nucleotide mutations and small insertions or deletions commonly result from DNA damage and replication errors that evade repair (Boiteux and Jinks-Robertson 2013; Zhu et al. 2014). They represent are an essential source of genetic variability, but in diploid cells, the potential detrimental effects of non-

synonymous mutations are usually compensated by the presence of a second intact wild type allele of the mutated gene. In contrast, mutational events that result in structural chromosomal alterations have a greater potential for leading to phenotypic changes because a single event can alter the genotype of multiple loci at once (Putnam and Kolodner 2017). These structural alterations can cause, for instance, gene dosage imbalances as a result of segmental deletions and amplifications (Putnam and Kolodner 2017). In addition, chromosomal alterations that result from recombination events between homologous chromosomes often lead to extensive copy-neutral loss-of-heterozygosity (LOH) (Symington, Rothstein, and Lisby 2014; Putnam and Kolodner 2017), which may unmask recessive mutant alleles. This dissertation focuses on the characterization of mutational events that lead to structural chromosomal alterations, particularly LOH, and their phenotypic consequences.

### **Mechanisms leading to loss-of-heterozygosity (LOH)**

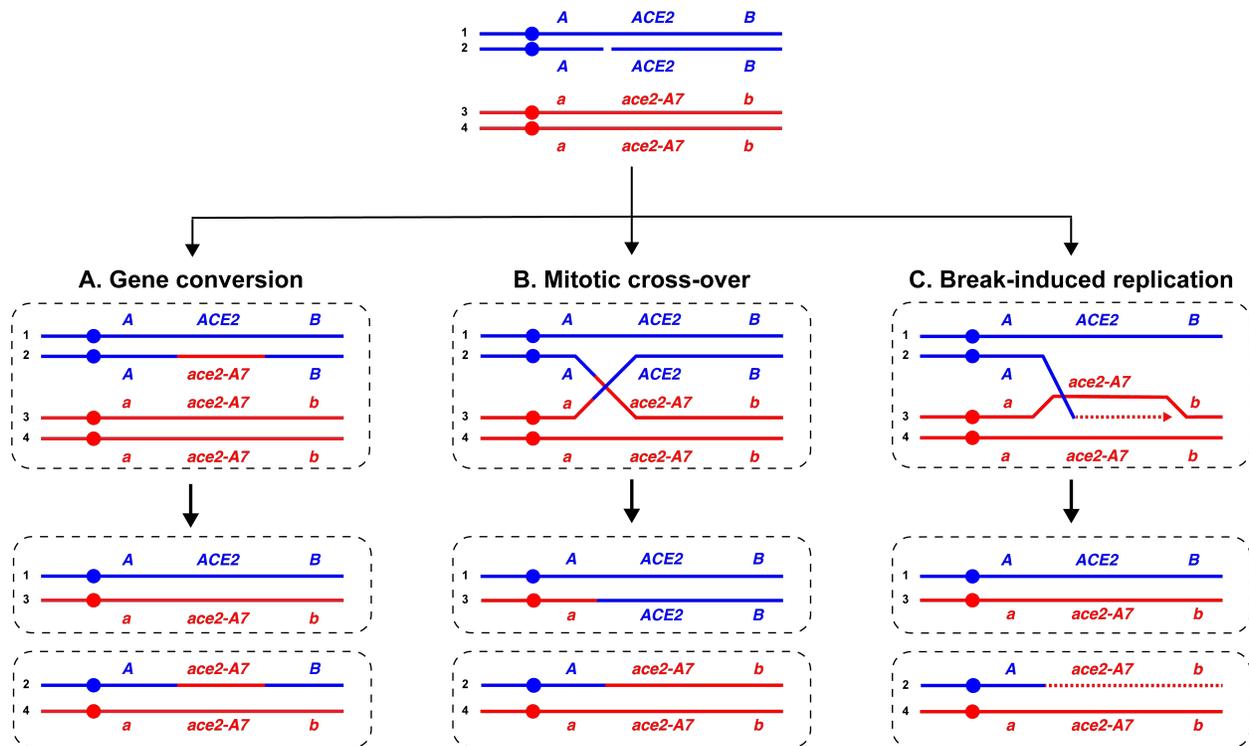
Generally, structural chromosomal alterations arise as a result of inappropriate repair of DNA double-strand breaks (DSBs), triggered by local DNA lesions and/or replication fork collapse (Pâques and Haber 1999; Putnam and Kolodner 2017). In yeast, as well as in human cells, several repair mechanisms compete hierarchically to repair eventual DSBs and, although different organisms favor different repair pathways their genetic outcomes are essentially identical. Typically, these mechanisms are classified in two categories. Cells may repair DSBs by simply re-joining the two broken chromosome ends through a non-homologous end joining pathway (NHEJ) (Chang et al. 2017). Alternatively, cells may repair DSBs using pathways that rely on an unbroken

DNA template through homologous recombination (HR). In this scenario, if the perfectly identical and allelic sister chromatid is the chosen homologous template for repair, the initial DNA lesion will have no genetic consequence (Pâques and Haber 1999; Symington, Rothstein, and Lisby 2014). However, in heterozygous diploid cells, the homologous chromosome (non-sister chromatid) may also be used as the donor sequence for repair, providing the potential for LOH tract accumulation and other changes in the chromosome structure to occur. For instance, interstitial and terminal copy-neutral LOH tracts, ranging from a few to hundreds of kilobases, may arise as a result of interhomolog allelic mitotic homologous recombination (Symington, Rothstein, and Lisby 2014).

The HR pathway is initiated by 5'-3' resection at the DSB. One of the resulting 3' single-stranded DNA (ssDNA) tails invades the double stranded DNA (dsDNA) of the homologous template to form a displacement loop (D-loop). The 3' end of the resected tail is then used to prime the synthesis using the complementary donor strand as template. The displaced donor strand is then available for annealing to the other resected tail on the broken molecule. Through DNA synthesis, branch migration and ligation, these structures eventually progress to form a double Holliday junction (dHJ) intermediate. Resolution of the dHJ may occur by dissolution and re-annealing of recipient DNA, or by endonucleolytic cleavage. The orientation of the cleavage of the the dHJ may lead to the formation of either crossover (perpendicular orientation) or non-crossover (parallel orientation) outcomes. Formation of a non-crossover product, *i.e.* a gene conversion tract, will result in a daughter cell with an interstitial copy neutral LOH tract after the ensuing cell division (Chen et al. 2007) (Figure 1A). Alternatively, the dHJ

may be resolved to yield a crossover product, which can result in reciprocal terminal LOH tracts if the recombinant chromatids segregate to different daughter cells (Symington, Rothstein, and Lisby 2014) (Figure 1B).

An identical terminal LOH genetic outcome may also arise as a result of break-induced replication (BIR), a synthesis-dependent recombination pathway for repair of single-ended DSBs that results in the non-reciprocal transfer of DNA from the donor homologous chromosome to the broken, recipient molecule (Llorente, Smith, and Symington 2008; Malkova and Ira 2013; Donnianni and Symington 2013) (Figure 1C). This pathway is also initiated by single strand invasion and D-loop formation. However, in this case, the homologous chromosome is used as a template for conservative DNA synthesis that may replicate > 100 kb of the template until the end of the chromosome, resulting in extensive LOH. Although the genetic outcome of repair through BIR is indistinguishable from interhomolog allelic mitotic recombination resolved by crossover, it has been demonstrated that BIR is suppressed when both ends of a DSB are available to initiate repair (Stark and Jasin 2003; Llorente, Smith, and Symington 2008) and the efficiency of this repair pathway is greatly reduced (<40%) when the length of DNA to be synthesized is greater than ~100 Kb (Donnianni and Symington 2013). In addition to the pathways described above, LOH can be generated by other types of mutational events. For example point mutations may lead to localized LOH and segmental deletions and whole chromosome loss may lead to extensive LOH due to the loss of one haplotype (Symington, Rothstein, and Lisby 2014; Putnam and Kolodner 2017).



**Figure 0.1. Genetic outcomes of interhomolog mitotic homologous recombination.**

Cells may repair DSB lesions through different pathways that rely on homologous recombination. Circles indicate centromeres. Three different heteroalleles are represented by A/a, ACE2/ace2-A7 and B/b. Duplicated homologous chromosomes in diploid cells are represented in colors blue and red. A) Gene conversion. Unidirectional genetic exchange between a donor homolog (red) and a recipient homolog (blue) leads to a non-reciprocal interstitial tract of LOH in only one of the resulting daughter cells. B) Crossover. Recombination resolved by a crossover yields extensive LOH if the recombinant molecules segregate to opposite cells in the subsequent cell division. C) Break-induced replication (BIR). One-ended DSBs repaired through conservative synthesis lead to extensive LOH in one of the resulting daughter cells. Note: all LOH events shown above span the ACE2 locus, which will be relevant for the experiments described in Chapters II and III of this dissertation.

### Incremental steps or large jumps in the accrual of chromosomal alterations?

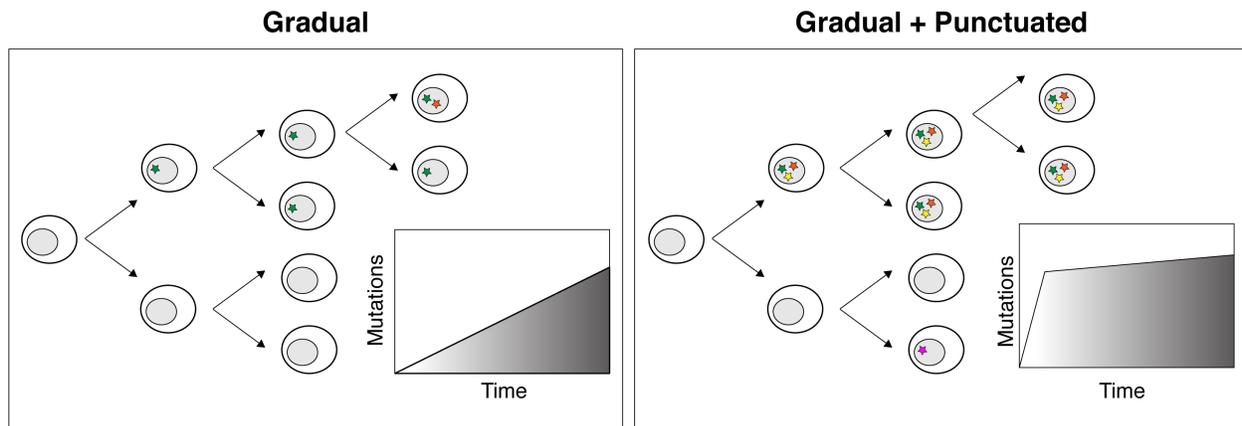
Over the last four decades, the development and advancement of whole-genome sequencing technologies have enabled a revolution in our understanding of the organization and evolution of genomes (Shendure et al. 2017). The systematic and comprehensive characterization of thousands of cancer and microbial genomes to date

have revealed many details about the nature, frequency and distribution of mutations (Ciriello et al. 2013; Alexandrov and Stratton 2014; Peter et al. 2018). The landscape of somatic mutations found in most cancer genomes is astounding when compared to normal healthy cells. An average cancer genome contains about 1000 - 10000 point mutations, 10 - 100 small insertions and deletions and 1 - 10 large scale chromosomal rearrangements (reviewed in Zhang and Pellman 2015). Large-scale genomic studies have revealed the existence of mutation hotspots, genomic regions including late replicating and heterochromatic regions that are more prone to accumulate mutations. In addition, most cancer types are associated with a characteristic mutational signature, which suggests cancers may be driven by different mutational processes (Ciriello et al. 2013; Alexandrov and Stratton 2014). For instance, colorectal carcinoma, uterine carcinoma and acute myeloid leukemia are frequently associated with point mutations, while ovarian and breast carcinoma are characterized by copy number changes (Ciriello et al. 2013). Although it is now evident that the type and distribution of mutations are not uniform throughout the genome, their frequency of appearance is estimated to be low and these events are generally thought to accumulate independently of each other over time, in a random and gradual fashion (Stratton, Campbell, and Futreal 2009). While literature overwhelmingly supports this general mechanism, mounting evidence is emerging for the co-existence of punctuated bursts of mutation accumulation, in which multiple mutational events take place over a single or a few cell division cycles (Zhang et al. 2015; Zhang and Pellman 2015; Sottoriva et al. 2015; Gao et al. 2016; Field et al. 2018). One relatively well understood example is chromothripsis, a mutational phenomenon characterized by extensive chromosomal rearrangements confined to a

single chromosome, a chromosome segment or occasionally a few chromosomes (Stephens et al. 2011; Leibowitz, Zhang, and Pellman 2015). These massive localized rearrangements have been shown to arise over a single cell division as a result of nuclear reincorporation of mis-segregated chromosomes (Zhang et al. 2015) and also as a result of telomere crisis caused by rupture and fragmentation of dicentric chromosomes (Maciejowski et al. 2015). Other examples include chromoplexy (Baca et al. 2013), firestorms (Hicks et al. 2006) and kataegis (Roberts et al. 2012), all of which are associated to numerous point mutations or chromosomal rearrangements confined to one or a couple limited genomic regions. In addition to these examples of bursts of clustered mutations, a few recent studies have also reported evidence of punctuated bursts of copy number alterations (CNAs) distributed on a genome-wide scale in some types of cancer and neurodevelopmental disorders (Gao et al. 2016; Liu et al. 2017; Field et al. 2018).

Recent genomic analysis of thousands of individual cells extracted from tumors of patients with triple-negative breast cancer revealed that cells extracted from the same tumor could be broadly clustered into only 1 to ~3 discrete phylogenetic subpopulations based on their copy number profiles (Gao et al. 2016). Each phylogenetic branch was composed of tens of individual cells with highly rearranged genomes compared to normal cells, yet with very similar alterations to each other indicating stable clonal expansion. Importantly, cells with intermediate rearranged karyotypes were not detected in any of the 12 tumor samples analyzed, which should have been observed if the CNAs accumulated gradually or sequentially. Instead, advanced mathematical modeling pointed to a scenario through which the majority of CNAs were acquired at early stages

of tumor evolution, in short punctuated bursts, followed by stable clonal expansions that formed the tumor mass.



**Figure 0.2. Models of mutation accumulation.**

Left) Traditional model of gradual accumulation of mutations, where new mutations emerge in incremental, sequential steps. Dashed lines indicate the time of appearance of new mutations. Right) Punctuated model of mutation accumulation, in which multiple mutations accumulate over a short period of crisis, followed by stasis and clonal expansion of mutant cells. Insets represent the rate of mutation accumulation over time for each model.

Similar conclusions were drawn from single cell whole genome sequencing data and mathematical modeling approaches used to investigate genomic heterogeneity in uveal melanoma (UM), the most frequent type of primary eye cancer, which often leads to metastatic death (Field et al. 2018). UM is often associated with inactivation of one allele of the *BAP1* tumor suppressor gene, followed by loss of the second functional allele through an LOH event spanning the *BAP1* locus on chromosome 3. The patterns of genome-wide CNA and LOH of thousands of single cells from 151 different tumor samples were strikingly reminiscent of those found in breast cancer: single genomes from the same tumor sample could be clustered into few and narrowly defined phylogenetic branches, whereas cells with intermediate rearrangements were not found.

Combined, these results strongly imply that the rearrangements present in each tumor subpopulation most likely accumulated rapidly, through an early burst of genome-wide instability in one or a few founding cells. Those highly rearranged cells then expanded stably to give rise to genomically uniform subpopulations that were ultimately found in the advanced stage tumors.

The discovery of rare bursts of chromosomal alterations has the potential to transform our understanding of the initial steps of cancer formation and phenotype-genotype evolution in general. However, mechanisms and cellular pathways governing these bursts are yet well characterized and their investigation is technically very challenging. In Chapter II of this dissertation, we provide experimental evidence that an analogous phenomenon of systemic genomic instability leading to punctuated bursts of chromosomal alterations also occurs in the yeast *Saccharomyces cerevisiae*. The significance of this finding is mainly threefold, (1) it indicates that the punctuated evolutionary model might be widespread among higher eukaryotes, (2) it provides a powerful model system that may facilitate the investigation of the underlying molecular mechanisms leading to punctuated bursts of genomic instability and (3) it reveals a novel mutational mechanism that can contribute to rapid changes in phenotype and adaptive evolution of natural populations of yeast. We discuss in Chapter II plausible underlying causes for transient genomic crises and, in Chapter IV, we propose future perspectives for the research in this field and some of the technical challenges that will need to be overcome.

## **A tug of war in the regulation of genomic heterozygosity in yeast**

Loss of a functional tumor suppressor allele through LOH can lead to uncontrolled cell proliferation and tumorigenesis (Ryland et al. 2015). For instance, this mechanism accounts for a large portion of the cases of retinoblastoma, which are frequently associated with loss of a functional copy of the Rb allele (Cavenee et al. 1983). LOH is not only an outcome of genomic instability that plays an important role in tumorigenesis, but these events are also a driving force in the evolution of unicellular microorganisms (Magwene et al. 2011). Mechanisms leading to LOH have an elevated potential for leading to phenotypic changes because one single mutational event has the capability of altering the genotype of long stretches of DNA harboring hundreds of genes. Population genomic analyses of many yeast isolates have shown that heterozygosity is common among wild diploid strains and LOH might represent a path of easy access to new allelic combinations and phenotypes (Magwene et al. 2011). For example, LOH has been correlated to recurrent acquisition of drug resistance (Coste et al. 2006), adaptation to nutrient-limited conditions (Smukowski Heil et al. 2017), among others.

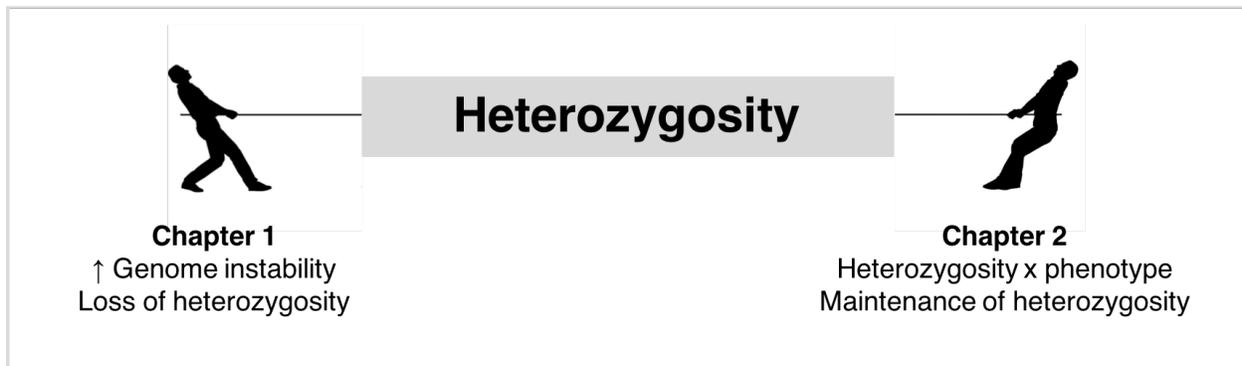
An important feature of this type of mutational event is that the resulting LOH tract can never be reverted back to the original heterozygous genotype. The high rates of LOH events during mitotic cell division and the irreversible nature of these events lead to many intriguing questions: How is heterozygosity maintained overtime in the genomes of wild yeast strains? How does heterozygosity influence phenotype? What are the phenotypic consequences of LOH?

One possibility is that human-associated environments facilitate outcrossing between unrelated strains, thus contributing to the maintenance of heterozygosity in strains isolated from those settings (Magwene et al. 2011). Another possibility, and the one explored in this dissertation, is that heterozygosity itself may be advantageous to natural yeast populations. A positive correlation between heterozygosity and fitness has been described in many other species, a phenomenon named hybrid vigor or heterosis (Birchler et al. 2010; Melchinger et al. 2007). This relationship has been characterized in a few studies showing evidence that *S. cerevisiae* strains also benefit from heterozygous genomes. Using a yeast collection with strains originated from two major groups, the “domesticated” strains, which were selected from environments associated with human activity (laboratory, industry, clinic) and the “wild” strains, which were isolated from natural habitats, Plech *et al.* 2014 showed that heterosis is prevalent in the group of domesticated yeast strains (Plech, de Visser, and Korona 2014). In addition, Shapira *et al.* 2014 successfully demonstrated that yeast heterosis is governed by the combined effects of different genetic interactions – dominance, overdominance and epistasis (Shapira et al. 2014).

Although these studies established the initial background necessary for understanding how this complex phenomenon manifests itself in yeast, the genomic regions and specific genes associated with the heterotic phenotype were not investigated. In addition, the hybrid strains analyzed in those studies were artificially created by mating spores derived from homozygous unrelated strains, *i.e.* the heterotic phenotypes analyzed are not observed in nature. In Chapter III of this dissertation, we contribute insights into this problem by demonstrating that heterozygosity likely plays a

role in the fitness of a natural hybrid *S. cerevisiae* strain. We use different approaches to reduce the abundance and genomic distribution of heterozygous alleles and show that phenotype is altered each time heterozygosity is reduced.

In summary, my doctoral research focused on the investigation of two counter-acting forces that regulates the levels of heterozygosity on the genome and, consequently, contribute to shaping genome structure. In **Chapter II**, I describe a novel phenomenon of genomic instability that can very rapidly erode heterozygosity from the genome. In **Chapter III**, I investigate the impact of heterozygosity on fitness, and how this relationship may contribute to the maintenance of heterozygous genomes in wild isolates of *S. cerevisiae* (Figure 3). The combined results presented here provide an experimental model system to further dissect the fundamental mechanisms responsible for bursts of systemic genomic instability that might underlie cancer and genomic disorders. In addition, they also provide insights into the roles of heterozygosity and mitotic recombination in shaping the genome architecture of *S. cerevisiae*.



**Figure 0.3. The levels of heterozygosity in the yeast genome are regulated by the balance between the counter-acting forces of heterosis and loss-of-heterozygosity.**

In Chapter II of this dissertation, we described the occurrence of genome instability episodes that lead to the accumulation of multiple tracts of loss-of-heterozygosity over one or a few mitotic cell divisions. In Chapter III, we investigated the role that heterozygous alleles play on fitness, a relationship that could be contributing to the long-term maintenance of highly heterozygous genomes in many yeast strains.

## REFERENCES

- Alexandrov, Ludmil B., and Michael R. Stratton. 2014. "Mutational Signatures: The Patterns of Somatic Mutations Hidden in Cancer Genomes." *Current Opinion in Genetics and Development* 24 (1): 52–60. <https://doi.org/10.1016/j.gde.2013.11.014>.
- Baca, Sylvan C., Davide Prandi, Michael S. Lawrence, Juan Miguel Mosquera, Alessandro Romanel, Yotam Drier, Kyung Park, et al. 2013. "Punctuated Evolution of Prostate Cancer Genomes." *Cell* 153 (3): 666–77. <https://doi.org/10.1016/j.cell.2013.03.021>.
- Birchler, James A, Hong Yao, Sivanandan Chudalayandi, Daniel Vaiman, and Reiner Veitia. 2010. "Heterosis." *The Plant Cell* 22 (July): 2105–12. <https://doi.org/10.1105/tpc.110.076133>.
- Boiteux, Serge, and Sue Jinks-Robertson. 2013. "DNA Repair Mechanisms and the Bypass of DNA Damage in *Saccharomyces Cerevisiae*." *Genetics* 193 (4): 1025–64. <https://doi.org/10.1534/genetics.112.145219>.
- Cavenee, W. K., T. P. Dryja, R. A. Phillips, W. F. Benedict, R. Godbout, B. L. Gallie, A. L. Murphree, L. C. Strong, and R. L. White. 1983. "Expression of Recessive Alleles by Chromosomal Mechanisms in Retinoblastoma." *Nature* 305 (5937): 779–84. <https://doi.org/10.1038/305779a0>.
- Chang, Howard H. Y., Nicholas R. Pannunzio, Noritaka Adachi, and Michael R. Lieber. 2017. "Non-Homologous DNA End Joining and Alternative Pathways to Double-Strand Break Repair." *Nature Reviews Molecular Cell Biology* 18 (8): 495–506. <https://doi.org/10.1038/nrm.2017.48>.
- Chen, Jian-Min, David N. Cooper, Nadia Chuzhanova, Claude Férec, and George P. Patrinos. 2007. "Gene Conversion: Mechanisms, Evolution and Human Disease." *Nature Reviews Genetics* 8 (10): 762–75. <https://doi.org/10.1038/nrg2193>.
- Ciriello, Giovanni, Martin L Miller, Bülent Arman Aksoy, Yasin Senbabaoglu, and Chris Sander. 2013. "Emerging Landscape of Oncogenic Signatures across Human Cancers." *Nature Genetics* 45 (10): 1127–33. <https://doi.org/10.1038/ng.2762>.Emerging.
- Coste, Alix, Vincent Turner, Françoise Ischer, Joachim Morschhäuser, Anja Forche, Anna Selmecki, Judith Berman, Jacques Bille, and Dominique Sanglard. 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 (4): 2139–56. <https://doi.org/10.1534/genetics.105.054767>.

Donnianni, R. A., and L. S. Symington. 2013. "Break-Induced Replication Occurs by Conservative DNA Synthesis." *Proceedings of the National Academy of Sciences* 110 (33): 13475–80. <https://doi.org/10.1073/pnas.1309800110>.

Field, Matthew G., Michael A. Durante, Hima Anbunathan, Louis Z. Cai, Christina L. Decatur, Anne M. Bowcock, Stefan Kurtenbach, and J. William Harbour. 2018. "Punctuated Evolution of Canonical Genomic Aberrations in Uveal Melanoma." *Nature Communications* 9 (1). <https://doi.org/10.1038/s41467-017-02428-w>.

Gao, Ruli, Alexander Davis, Thomas O. McDonald, Emi Sei, Xiuqing Shi, Yong Wang, Pei Ching Tsai, et al. 2016. "Punctuated Copy Number Evolution and Clonal Stasis in Triple-Negative Breast Cancer." *Nature Genetics* 48 (10): 1119–30. <https://doi.org/10.1038/ng.3641>.

Hanahan, D, and R A Weinberg. 2000. "The Hallmarks of Cancer." *Cell* 100 (1): 57–70. <https://doi.org/10.1007/s00262-010-0968-0>.

Hicks, James, Alexander Krasnitz, B Lakshmi, Nicholas E Navin, Michael Riggs, Evan Leibu, Diane Esposito, et al. 2006. "Novel Patterns of Genome Rearrangement and Their Association with Survival in Breast Cancer Novel Patterns of Genome Rearrangement and Their Association with Survival in Breast Cancer." *Genome Research*, 1465–79. <https://doi.org/10.1101/gr.5460106>.

Leibowitz, Mitchell L., Cheng-Zhong Zhang, and David Pellman. 2015. "Chromothripsis: A New Mechanism for Rapid Karyotype Evolution." *Annual Review of Genetics* 49 (1): 183–211. <https://doi.org/10.1146/annurev-genet-120213-092228>.

Liu, Pengfei, Bo Yuan, Claudia M B Carvalho, Arthur Wuster, Klaudia Walter, Tomasz Gambin, Zechen Chong, et al. 2017. "An Organismal CNV Mutator Phenotype Restricted to Early Human Development." *Cell* 168 (5): 830–42. <https://doi.org/10.1016/j.cell.2017.01.037.An>.

Llorente, Bertrand, Catherine E. Smith, and Lorraine S. Symington. 2008. "Break-Induced Replication: What Is It and What Is It For?" *Cell Cycle* 7 (7): 859–64. <https://doi.org/10.4161/cc.7.7.5613>.

Maciejowski, John, Yilong Li, Nazario Bosco, Peter J. Campbell, and Titia De Lange. 2015. "Chromothripsis and Kataegis Induced by Telomere Crisis." *Cell* 163 (7): 1641–54. <https://doi.org/10.1016/j.cell.2015.11.054>.

Magwene, Paul M., Ömür Kayıkçı, Joshua A. Granek, Jennifer M. Reininga, Zackary Scholl, and Debra Murray. 2011. "Outcrossing, Mitotic Recombination, and Life-History Trade-Offs Shape Genome Evolution in *Saccharomyces Cerevisiae*." *Proceedings of the National Academy of Sciences* 108 (5): 1987–92. <https://doi.org/10.1073/pnas.1012544108>.

Malkova, Anna, and Grzegorz Ira. 2013. "Break-Induced Replication: Functions and Molecular Mechanism." *Current Opinion in Genetics and Development* 23 (3): 271–79. <https://doi.org/10.1016/j.gde.2013.05.007>.

Melchinger, a. E., H. F. Utz, H. P. Piepho, Z. B. Zeng, and C. C. Schön. 2007. "The Role of Epistasis in the Manifestation of Heterosis: A Systems-Oriented Approach." *Genetics* 177 (November): 1815–25. <https://doi.org/10.1534/genetics.107.077537>.

Pâques, F, and J E Haber. 1999. "Multiple Pathways of Recombination Induced by Double-Strand Breaks in *Saccharomyces Cerevisiae*." *Microbiology and Molecular Biology Reviews : MMBR* 63 (2): 349–404. <https://doi.org/10.1128/MMBR.63.2.349-404.1999>.

Payen, Celia, Anna B. Sunshine, Giang T. Ong, Jamie L. Pogachar, Wei Zhao, and Maitreya J. Dunham. 2016. "High-Throughput Identification of Adaptive Mutations in Experimentally Evolved Yeast Populations." *PLoS Genetics* 12 (10): 1–24. <https://doi.org/10.1371/journal.pgen.1006339>.

Peter, Jackson, Matteo De Chiara, Anne Friedrich, Jia-xing Yue, David Pflieger, Anders Bergström, Anastasie Sigwalt, et al. 2018. "Genome Evolution across 1,011 *Saccharomyces Cerevisiae* Isolates." *Nature* 556. <https://doi.org/10.1038/s41586-018-0030-5>.

Plech, Marcin, J. Arjan G. M. de Visser, and Ryszard Korona. 2014. "Heterosis Is Prevalent Among Domesticated but Not Wild Strains of *Saccharomyces Cerevisiae*." *G3 Genes Genomes Genetics* 4 (2): 315–23. <https://doi.org/10.1534/g3.113.009381>.

Putnam, Christopher D, and Richard D Kolodner. 2017. *Pathways and Mechanisms That Prevent Genome Instability In Saccharomyces Cerevisiae. Genetics*. Vol. 206. <https://doi.org/10.1534/genetics.112.145805>.

Roberts, Steven A., Joan Sterling, Cole Thompson, Shawn Harris, Deepak Mav, Ruchir Shah, Leszek J. Klimczak, et al. 2012. "Clustered Mutations in Yeast and in Human Cancers Can Arise from Damaged Long Single-Strand DNA Regions." *Molecular Cell* 46 (4): 424–35. <https://doi.org/10.1016/j.molcel.2012.03.030>.

Ryland, Georgina L., Maria A. Doyle, David Goode, Samantha E. Boyle, David Y.H. Choong, Simone M. Rowley, Jason Li, et al. 2015. "Loss of Heterozygosity: What Is It Good For?" *BMC Medical Genomics* 8 (1): 1–12. <https://doi.org/10.1186/s12920-015-0123-z>.

Shapira, R, T Levy, S Shaked, E Fridman, and L David. 2014. "Extensive Heterosis in Growth of Yeast Hybrids Is Explained by a Combination of Genetic Models." *Heredity*, no. October 2013: 1–11. <https://doi.org/10.1038/hdy.2014.33>.

Shendure, Jay, Shankar Balasubramanian, George M. Church, Walter Gilbert, Jane Rogers, Jeffery A. Schloss, and Robert H. Waterston. 2017. "DNA Sequencing at 40: Past, Present and Future." *Nature* 550 (7676). <https://doi.org/10.1038/nature24286>.

Smukowski Heil, Caiti S., Christopher G. DeSevo, Dave A. Pai, Cheryl M. Tucker, Margaret L. Hoang, and Maitreya J. Dunham. 2017. "Loss of Heterozygosity Drives Adaptation in Hybrid Yeast." *Molecular Biology and Evolution* 34 (7): 1596–1612. <https://doi.org/10.1093/molbev/msx098>.

Sottoriva, Andrea, Haeyoun Kang, Zhicheng Ma, Trevor A. Graham, Matthew P. Salomon, Junsong Zhao, Paul Marjoram, et al. 2015. "A Big Bang Model of Human Colorectal Tumor Growth." *Nature Genetics* 47 (3): 209–16. <https://doi.org/10.1038/ng.3214>.

Stark, Jeremy M, and Maria Jasin. 2003. "Extensive Loss of Heterozygosity Is Suppressed during Homologous Repair of Chromosomal Breaks." *Molecular and Cellular Biology* 23 (2): 733–43. <https://doi.org/10.1128/MCB.23.2.733-743.2003>.

Stephens, Philip J., Chris D. Greenman, Beiyuan Fu, Fengtang Yang, Graham R. Bignell, Laura J. Mudie, Erin D. Pleasance, et al. 2011. "Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development." *Cell* 144 (1): 27–40. <https://doi.org/10.1016/j.cell.2010.11.055>.

Stratton, Michael R., Peter J. Campbell, and P. Andrew Futreal. 2009. "The Cancer Genome." *Nature* 458 (7239): 719–24. <https://doi.org/10.1038/nature07943>.

Symington, L. S., R. Rothstein, and M. Lisby. 2014. "Mechanisms and Regulation of Mitotic Recombination in *Saccharomyces Cerevisiae*." *Genetics* 198 (3): 795–835. <https://doi.org/10.1534/genetics.114.166140>.

Venkataram, Sandeep, Barbara Dunn, Yuping Li, Atish Agarwala, Jessica Chang, Emily R. Ebel, Kerry Geiler-Samerotte, et al. 2016. "Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast." *Cell* 166 (6): 1585–1596.e22. <https://doi.org/10.1016/j.cell.2016.08.002>.

Zhang, Cheng-Zhong, and David Pellman. 2015. "From Mutational Mechanisms in Single Cells to Mutational Patterns in Cancer Genomes." *Cold Spring Harbor Symposia on Quantitative Biology* 80: 117–37. <https://doi.org/10.1101/sqb.2015.80.027623>.

Zhang, Cheng-Zhong, Alexander Spektor, Hauke Cornils, Joshua M Francis, Emily K Jackson, Shiwei Liu, Mattew Meyerson, and David Pellman. 2015. "Chromothripsis from DNA Damage in Micronuclei." *Nature* 522: 179–84. <https://doi.org/10.1016/j.cogdev.2010.08.003>. Personal.

Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov. 2014. "Precise Estimates of Mutation Rate and Spectrum in Yeast." *Proceedings of the National Academy of Sciences* 111 (22): E2310–18. <https://doi.org/10.1073/pnas.1323011111>.

## CHAPTER II<sup>1</sup>

### Mitotic systemic genomic instability in yeast

#### Summary

Conventional models of genome evolution generally include the assumption that mutations accumulate gradually and independently over time. We characterized the occurrence of sudden spikes in the accumulation of genome-wide loss-of-heterozygosity (LOH) in *Saccharomyces cerevisiae*, suggesting the existence of a mitotic systemic genomic instability process (mitSGI). We characterized the emergence of a rough colony morphology phenotype resulting from an LOH event spanning a specific locus (*ACE2/ace2-A7*). Surprisingly, half of the clones analyzed also carried unselected secondary LOH tracts elsewhere in their genomes. The number of secondary LOH tracts detected was 20-fold higher than expected assuming independence between mutational events. Secondary LOH tracts were not detected in

---

<sup>1</sup> This chapter is an adaptation of a preprint manuscript available on BioRxiv, the figures have been renumbered to indicate both chapter and figure number.

Reference for the full article:

Rodrigues Prause A, **Sampaio NMV**, Ajith VP, Gurol TM, Chapman MJ, Malc EP, Chakraborty P, Duarte FM, Aguirre GM, Tizei PA, Pereira GAG, Mieczkowski PA, Nishant KT, Argueso JL. Mitotic systemic genomic instability in yeast. BioRxiv 161869; doi: <https://doi.org/10.1101/161869>

Contributions to this research are as follows:

Identification and genetic characterization of rough colony phenotype: ARP, PAT, TMG, GAGP. Quantitative LOH assay: **NMVS**, MJC, GMA. Karyotype analysis: **NMVS**, FMD. Whole Genome Sequencing: **NMVS**, VPA, PC, PAM, KTN, JLA. Data Analysis: **NMVS**, ARP, VPA, KTN, JLA. Strain Construction: **NMVS**, ARP. Manuscript Preparation: **NMVS**, JLA

The copyright holder for this preprint is the author/funder. All rights reserved. No reuse allowed without permission.

control clones without a primary selected LOH event. We then measured the rates of single and double LOH at different chromosome pairs and found that coincident LOH accumulated at rates 30-100 fold higher than expected if the two underlying single LOH events occurred independently. These results were consistent between two different strain backgrounds, and in mutant strains incapable of entering meiosis. Our results indicate that a subset of mitotic cells within a population experience systemic genomic instability episodes, resulting in multiple chromosomal rearrangements over one or few generations. They are reminiscent of early reports from the classic yeast genetics literature, as well as recent studies in humans, both in the cancer and genomic disorder contexts, all of which challenge the idea of gradual accumulation of structural genomic variation. Our experimental approach provides a model to further dissect the fundamental mechanisms responsible for mitSGI.

## **Introduction**

Heterozygosity is often associated with beneficial phenotypes in a variety of multicellular eukaryotes ranging from plants, to livestock, and even humans (Chen 2013). At the organismal level, heterozygosity can be promoted and maintained through breeding between unrelated individuals, and conversely, can be lost through inbreeding (Charlesworth et al. 2009). It can also be lost at the cellular level through allelic mitotic recombination between homologous chromosomes. Such loss-of-heterozygosity (LOH) events typically have negative consequences, such as somatic mosaicism or loss of tumor suppressor genes (Lapunzina and Monk 2011), but unless these mutations occur

in the germline, they are not heritable and do not have long term consequences for the species.

Single cell eukaryotes including various yeast species also benefit from heterozygous genomes (Magwene 2014; D'Enfert et al. 2017). However, maintaining heterozygosity is more challenging in these cases as a mitotic LOH event leads to immediate fixation of the homozygous state in a clonal cell lineage. High levels of genomic heterozygosity have been described in several *Saccharomyces cerevisiae* strains (Argueso et al. 2009; Magwene et al. 2011; Borneman et al. 2011). One of the first examples to be characterized was the JAY270/PE-2 strain used in bioethanol production (Argueso et al. 2009). This heterothallic diploid was originally isolated as a robust and highly productive contaminant at a sugarcane distillery (Basso et al. 2008). Similarly isolated wild strains are also heterothallic and heterozygous (Babrzadeh et al. 2012), and genomic heterozygosity is suspected to contribute to their industrial traits. Interestingly, in most of the strains described above, including JAY270/PE-2, heterozygosity is not evenly distributed across the genome. Heterozygous regions are interspersed with stretches of homozygosity indicating the occurrence of LOH events in clonal ancestors (Magwene et al. 2011). However, it is still unclear what, if any, consequences these LOH events may have on their general fitness.

The relationship between genomic heterozygosity, LOH, and phenotypic consequences in yeast is better understood in the human pathogen *Candida albicans*. In that system, LOH events have been shown to have a profound effect on clinically relevant traits, particularly drug resistance (D'Enfert et al. 2017; Bennett, Forche, and Berman 2014). For example, LOH leading to homozygosis of a hyperactive form of

Tac1, a transcription factor that regulates the multidrug transporter genes, results in increased efflux of azole antifungal drugs (Coste et al. 2006). In addition to providing a recurrent path to drug resistance, LOH has been shown to play a significant role in the evolution of the *C. albicans* genome (Hirakawa et al. 2015; Ford et al. 2015).

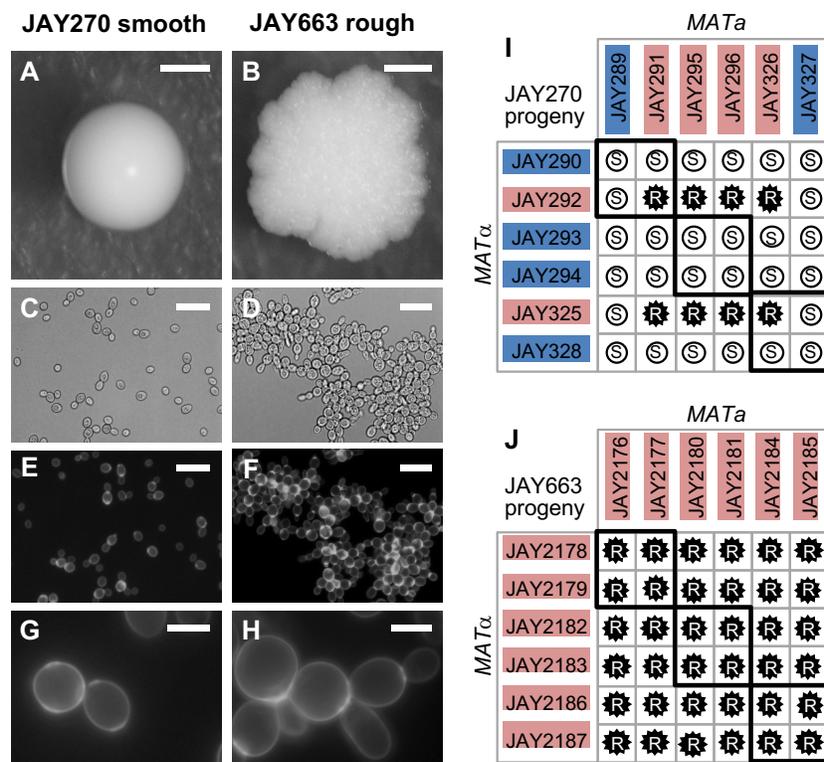
In this study, we identified and characterized a specific and easily discernible phenotypic transition in the *S. cerevisiae* JAY270/PE-2 strain, from smooth to rough yeast colony morphology, caused by an LOH event spanning the *ACE2* locus on chromosome XII (Chr12). Whole genome analyses of rough clones selected for carrying this specific Chr12 LOH event revealed that additional unselected recombination events were often present elsewhere in the genome. This initial observation was validated by direct measurements of coincident LOH rates at different chromosomes, suggesting the existence of a mitotic systemic genomic instability (mitSGI) process. The high rate of coincident LOH uncovered in our study resembles the bursts of accumulation of copy number alterations (CNAs) in human cancer (Gao et al. 2016) and genomic disorders (Liu et al. 2017). The results reported here have important ramifications for the characterization of mitSGI mechanisms that contribute to structural genomic variation.

## **Results**

### **Appearance of altered colony morphology derivatives of JAY270**

One of the most desirable features of the JAY270/PE-2 bioethanol production strain (henceforth referred to simply as JAY270) is that it does not normally aggregate during industrial sugarcane extract fermentation (*i.e.* cells stay in suspension in liquid culture). Accordingly, JAY270 produces normal hemispherical colonies with smooth

surfaces and edges when grown in solid agar medium (Fig. 2.1A). While this is the phenotype typically observed, over the course of our studies using this strain we noticed the sporadic occurrence of colonies clonally derived from JAY270 that displayed altered morphology: relatively flat-growing colonies with rough surfaces and edges (Fig. 2.1B). Under bright field microscopic examination, yeast cells derived from such rough colonies appeared to grow in chains, showing a budding pattern consistent with a defect in the separation of the daughter cells from their mother (Fig. 2.1C-D). We stained these cells with calcofluor white to visualize the chitin-rich ring septa, confirming the attachment of mother and daughter cells at the budding neck site (Fig. 2.1E-H).



**Figure 0.1. Smooth and rough colony morphologies, mother-daughter cell attachment, and phenotypes of diploids derived from mating specific haploids.** **A-H** show images of the JAY270 smooth parent diploid strain (left panels) and its spontaneous rough derivative JAY663 (right panels). **A** and **B**, colony morphologies on YPD agar after 3 days growth at 30C. **C** and **D**, bright field, and **E-H**, fluorescence microscopy of cells stained with calcofluor white to highlight chitin septa and the mother-

daughter cell attachment. Scale bars are 1mm (**A-B**), 20 $\mu$ m (**C-F**) and 5 $\mu$ m (**G-H**). **I** and **J**, Smooth (S, white circles) and rough (R, black stars) phenotypes of diploids formed by crossing the indicated MAT $\alpha$  and MAT $\alpha$  haploids isolated from three tetrads of each JAY270 (**I**) and JAY663 (**J**). Thick black lines indicate the four diploids derived from matings of intra-tetrad sibling haploids. The colored backgrounds for each haploid correspond to their inferred genotype (Blue, dominant wild type allele; Red, recessive mutant allele). All 12 haploids from panel **I** had their whole genomes sequenced. Co-segregation analysis with JAY270 HetSNPs (Fig. S2.1) was used for identification of the causal mutation at the ACE2 locus (Figs. S2.2 and S2.3). Main results in this figure were generated by: ARP.

We initially isolated five independent examples of such rough colonies for genetic characterization (JAY663, JAY664, JAY665, JAY912 and JAY913), all of which were derived either directly from JAY270 or from JAY270-isogenic strains. The phenotype of these isolates was stably maintained and was not reversible over several clonal generations, suggesting that it was likely hard-wired genetically and not caused by a transient transcriptional or post-transcriptional state. We estimated that these five rough colony isolates appeared spontaneously from a pool of ~50,000 smooth colonies. Assuming a genetic origin and based on this high frequency of occurrence in diploid cells, we reasoned that this phenotype was unlikely to be caused by a rare dominant *de novo* nucleotide point mutation, but instead, mitotic recombination leading to loss-of-heterozygosity (LOH) provided a more plausible mechanism.

In a parallel project, we observed that crossing two specific haploid descendants of JAY270 (JAY291 MAT $\alpha$  and JAY292 MAT $\alpha$ ) resulted in diploid cells with the same rough colony morphology and mother-daughter cell attachment pattern observed in the five rough-colony isolates above. This was despite the fact that JAY291, JAY292, and all other haploid derivatives of JAY270 have the normal smooth colony phenotype. This indicated that the rough colony phenotype was diploid-specific, and the ability to

consistently reproduce the mutant phenotype in controlled crosses between specific haploids opened an avenue to investigate its genetic basis.

We previously reported the whole genome sequence of the JAY291 haploid (Argueso et al. 2009). Since then we have sequenced the genomes of 55 additional JAY270-derived haploids. This genome sequence dataset, comprising fourteen sets of four-spore tetrads, was generated in a project to characterize the abundance, distribution and phasing of heterozygous loci in the JAY270 genome, the full results of which will be described elsewhere. These haploid genomic sequences were used to create a draft map of phased heterozygous single nucleotide polymorphisms (HetSNPs) containing 12,023 loci unevenly distributed across the genome (Fig. S2.1).

We carried out crosses between twelve sequenced haploid descendants of JAY270 (3 tetrads; Fig. 2.11). All possible  $MATa \times MAT\alpha$  crosses were performed producing 36 different diploids. Among them, we found eight with rough and 28 with smooth colony surfaces, in a pattern that was consistent with recessive inheritance of a trait controlled by a single gene. Even though the rough colony phenotype was not observed in any of the haploid parents, the phenotypes of their respective diploid combinations allowed us to infer which allele was present in the parents: either the wild type dominant allele or the recessive mutant allele.

In addition, we induced sporulation of one of the spontaneous rough-colony isolates, JAY663, dissected tetrads, and examined the phenotypes of the haploid derivatives. None of the resulting haploids displayed the rough colony phenotype; they were all smooth (~100 examined). We then took twelve of these haploids (JAY2176

through JAY2187 comprising three full tetrads, determined their mating types, and conducted all possible mating combinations between them (Fig. 2.1J). In this case, all 36 crosses resulted in rough colony diploids. This result was consistent with JAY663 being homozygous for the causal recessive mutant allele, and supported the hypothesis that copy neutral LOH could be responsible for the sporadic appearance of the mutant phenotype in JAY270.

### **Genetic basis of the rough colony phenotype**

Based on the interpretation that the rough colony phenotype was associated with monogenic recessive inheritance of a diploid-specific trait, we divided the sequenced JAY270-derived haploids from Fig. 2.1I into two groups according to their inferred genotype. Group 1 included the six haploids inferred to carry the mutant recessive allele, whereas group 2 included the six haploids with the wild type dominant allele. We then compared the genome sequences of the twelve haploids to the draft JAY270 HetSNPs map. We interrogated each of the HetSNPs searching for alleles that co-segregated in all six individuals within group 1, and that conversely, had the other allele co-segregating in all six individuals within group 2. This analysis identified two candidate regions that fit the strict co-segregation criterion (Fig. S2.2A-C). One of the candidate regions corresponded to ~30 Kb on Chr11, including thirteen genes; and the other spanned ~15 Kb containing nine genes on the right arm of Chr12, located ~50 Kb centromere proximal to the ribosomal DNA genes tandem repeats (rDNA).

We reviewed the annotations of the 22 candidate genes, and identified a gene located in the Chr12 region, *ACE2*, which encodes a transcription factor that controls

the expression of genes involved in the mother-daughter cell separation process (Weiss 2012). In cells lacking Ace2p, the daughter cell remains attached to the mother cell wall at the bud neck, resulting in the accumulation of multicellular clusters. Importantly, a diploid-specific rough colony phenotype is observed in *ace2/ace2* homozygous mutant strains in certain genetic backgrounds (Voth et al. 2005).

We inspected the genomic sequence of the *ACE2* gene in JAY291, and compared it to the sequence in the S288c reference genome. Only one difference was identified: The wild type *ACE2* allele in S288c contains a homopolymer run of eight adenine nucleotides, while the mutant allele in JAY291 has seven adenines in this region, resulting in a -1 frameshift mutation and a stop codon shortly downstream. Hence, we named the mutant allele *ace2-A7*. We then conducted reciprocal complementation tests to formally demonstrate that *ace2-A7* was the causal mutation responsible for the rough colony phenotype. The mutant alleles in haploids JAY291 and JAY292 were replaced with the wild type allele, resulting respectively in the isogenic *ACE2* strains JAY1051 and JAY1039. When these allele replacement strains were crossed to *ace2-A7* strains (Fig. S2.2D), the resulting diploids displayed the smooth colony phenotype, thus confirming that the wild type *ACE2* allele fully complemented the *ace2-A7* mutation in heterozygous diploids.

### **Analysis of Chr12 LOH in spontaneous rough colony isolates**

After identifying the association between the *ACE2* locus and the rough colony phenotype, we determined its sequence in the five spontaneous rough colony derivatives isolated earlier in the study. We PCR-amplified and Sanger-sequenced the

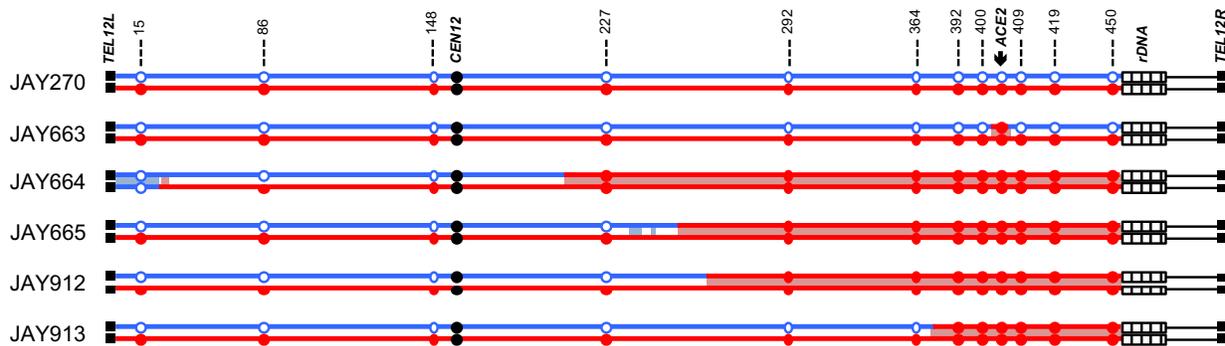
region containing the adenine homopolymer run in *ACE2* from JAY270, from the haploid derivatives JAY290 and JAY291, and from the rough colony isolates (Fig. S2.3A-B). This analysis confirmed the presence of a run of 8 adenines in *ACE2* (JAY290) and 7 adenines in *ace2-A7* (JAY291). The chromatogram in the JAY270 heterozygous diploid was consistent with a mixture of *ACE2* and *ace2-A7* DNA templates being present in the sequencing reaction: single nucleotide peaks were observed at positions primer-proximal to the homopolymer run, and out-of-register double peaks were seen downstream of the seventh adenine nucleotide. The chromatograms for all five rough colony isolates showed the presence of the *ace2-A7* frameshift mutation and absence of the *ACE2* allele. The loss of the *ACE2* allele in the diploid rough-colony isolates may be explained by either a copy-neutral LOH mechanism such as inter-homolog mitotic recombination, by a segmental deletion spanning *ACE2*, or Chr12 monosomy. To distinguish between these scenarios, we conducted tetrad analysis with the five spontaneous rough colony isolates. Four of them produced tetrads that had four viable haploid spores, and each spore had a copy of the *ACE2* locus as determined by PCR (data not shown). One of the isolates, JAY664, produced tetrads with two viable and two inviable spores, indicating the presence of a recessive-lethal mutation. We performed array-CGH analysis on JAY664 and determined that two copies of Chr12, including the *ACE2* locus, were present (Fig. S2.3C-D). Together, these results showed that all five rough colony isolates were homozygous for *ace2-A7*, in agreement with the initial hypothesis that the high frequency of smooth to rough colony morphology transitions among JAY270 derivatives was caused by interhomolog recombination leading to copy-neutral LOH. Unexpectedly, the JAY664 array-CGH also showed that

this rough colony isolate did carry copy number alterations in genomic regions other than Chr12. In particular, a terminal deletion on the right arm of Chr6 spanned multiple essential genes and explained the 2:2 spore viability phenotype (Fig. S2.3E).

Interestingly, the breakpoint for the Chr6 deletion occurred at a position immediately distal to *FAB1*, where a tRNA gene and Ty1 retrotransposon sequences are found in the S288c reference genome and in the JAY270 maternal Chr6 homolog, which sustained the deletion. We analyzed the status of HetSNP markers flanking the breakpoint and found that a proximal marker (Chr6 - 185 Kb) remained heterozygous, while a distal marker (Chr6 - 229 Kb) lost heterozygosity through a deletion mechanism (Fig. S2.4A-B). Even though we did not characterize the precise sequences that were joined at the deletion breakpoint in JAY664, this pattern was consistent with non-allelic homologous recombination (HR) involving Ty retrotransposon repeats, a major class of gross chromosomal rearrangements observed in *S. cerevisiae* (Argueso et al. 2008; Putnam and Kolodner 2017).

LOH is typically a regional, rather than local, mutational mechanism. Interstitial tracts of homozygosity can span tens of kilobases, and terminal tracts are even longer, extending all the way to the telomeres (St Charles and Petes 2013). Therefore, in addition to being homozygous for *ace2-A7*, the rough colony isolates might also be homozygous for flanking HetSNPs. We tested this model initially at low resolution by determining the genotypes at eleven Chr12 HetSNPs using PCR (Table S2.3). The results of this analysis were compiled to produce the LOH tract maps shown in Fig. 2.2. As expected, JAY270 was heterozygous for all eleven markers tested. Notably, Chr12 in this strain is only heterozygous for positions to the left of the rDNA cluster (Fig. S2.1).

This pattern is similar to that described previously for other heterozygous diploid *S. cerevisiae* genomes and is suggestive of ancestral LOH events mediated by rDNA instability (Magwene et al. 2011).



**Figure 0.2. LOH tract maps of Chr12 from five original rough colony isolates.**

The genotypes at twelve phased JAY270 Chr12 HetSNP marker loci were determined using PCR and RFLP or Sanger sequencing analyses (Table S2.3). The approximate coordinates of the markers are shown in Kb. The Chr12 homolog containing the *ace2-A7* allele was arbitrarily designated as maternal (Chr12-M, red) and the homolog containing the wild type *ACE2* allele as paternal (Chr12-P, blue). JAY270 was heterozygous at all markers, and all rough colony isolates were homozygous for the *ace2-A7* allele. White boxes distal to the 450 Kb HetSNP represent ~1.5 Mb of ribosomal DNA repeats (rDNA). Chr12 regions distal to the rDNA do not contain any heterozygous markers in JAY270. The red or blue shading corresponds to the directions (M or P, respectively) and approximate breakpoint positions of the LOH tracts determined at high resolution using whole genome sequencing (detailed in Fig. S2.6). Main results in this figure were generated by: ARP, MJC, NMVS.

Analysis of the JAY663 isolate showed that, while it was homozygous for the *ace2-A7* mutation, it remained heterozygous at all other flanking markers, including those immediately proximal and immediately distal to the *ACE2* locus. A mitotic gene conversion tract limited to the 8.7 Kb region between these HetSNPs could explain this result. Alternatively, a *de novo* -1 contraction mutation in the adenine homopolymer run of the *ACE2* allele could also account for the JAY663 genotype (Tran et al. 1997).

The four remaining isolates (JAY664, JAY665, JAY912, and JAY913) were homozygous for regions well beyond the *ACE2* locus. In this low-resolution map, all four LOH tracts were unidirectional, continuous, and homozygous for SNPs present in the Chr12 homolog that contained the *ace2-A7* allele, which we arbitrarily designated the maternal homolog (Chr12-M; red in all figures). Subsequent high resolution LOH mapping using whole genome sequencing (WGS; below) confirmed the initial results and revealed additional complexities to the tracts. The centromere-proximal breakpoints of the LOH tracts were roughly mapped to positions ranging from 39 Kb (JAY913) to 184 Kb (JAY664) from the *ACE2* locus. On the distal side, these isolates had an additional 45 Kb of LOH that extended to the HetSNP at position 450 Kb, located 1.4 Kb proximal to the rDNA repeats. From this point, Chr12 contains ~1.5 Mb of rDNA repeats plus another ~0.6 Mb of distal homozygous single copy sequences. Since the 450 Kb HetSNP was the most distal marker in Chr12, we could not distinguish if these LOH tracts were generated as very long interstitial gene conversion events, or if they extended to the right telomere. This initial PCR-based analysis also revealed an unexpected secondary LOH event on the left arm of Chr12 in JAY664, but in this case, it was associated with homozygosity for the SNPs from the paternal homolog (Chr12-P, blue in all figures; see below).

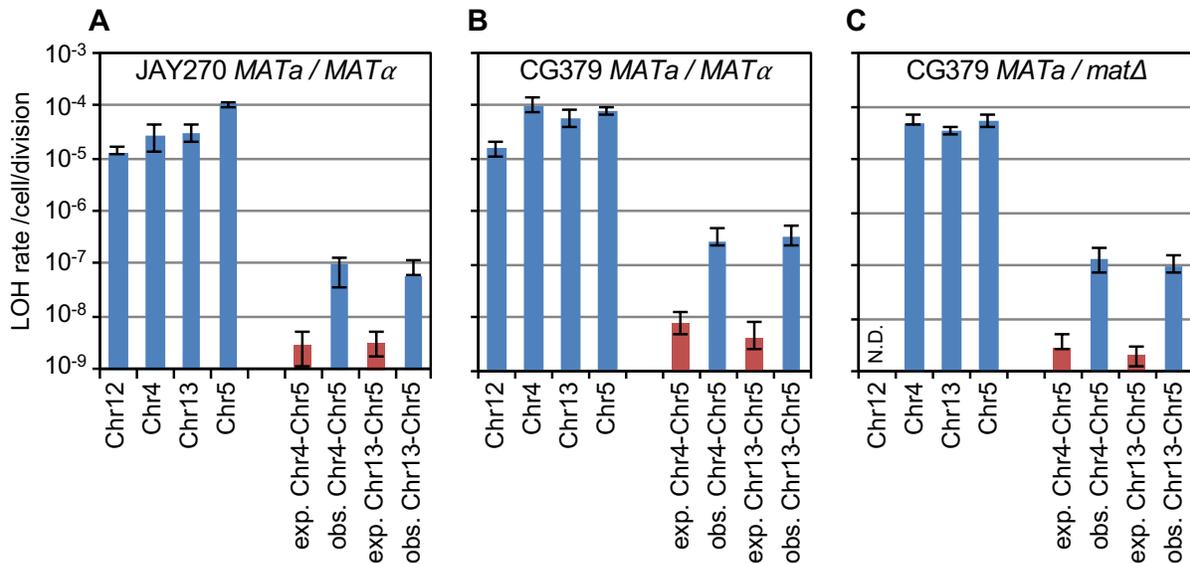
### **Analysis of selected Chr12 LOH**

Taken together, the results described above showed that the majority of the isolates with altered colony morphology were homozygous not only at the *ACE2* locus, but also for surrounding regions, indicating that interhomolog mitotic recombination was

frequent in JAY270 and that it likely had substantial effect on the genetic makeup of this strain. In addition, the distribution of HetSNPs in the genome is notably uneven (Fig. S2.1), with long tracts of homozygosity, suggesting that abundant LOH occurred in the JAY270 lineage.

To gain a deeper understanding of the impact of genome instability processes on the present genetic composition of JAY270, we conducted experiments to directly measure the rate of LOH in this strain (Fig. 2.3A). Starting with a homozygous *ura3/ura3* derivative of JAY270 (FGY050; gift from F. Galzerani), we introduced one copy of the  $\kappa$ *URA3-scURA3-KanMX4* CORE2 counter selectable cassette (Zhang et al. 2013) at a position immediately proximal to *ACE2* (1.3 Kb from the adenine homopolymer run). We grew cultures of strains carrying this insertion and plated the cells in media containing 5-FOA to identify clones that had lost the cassette. We performed this assay in a derivative of JAY270 carrying the hemizygous CORE2 insertion in Chr12-M. The frequency of homozygosity for the *ACE2* allele was  $1.2 \times 10^{-4}$ , comparable to the unselected frequency of *ace2-A7* homozygosity (5 in ~50,000) estimated earlier in the study. We also used hemizygous CORE2 insertions to measure LOH rates at two other positions in the genome (Chr4 near *SSF2*, and Chr13 near *ADH6*), and on Chr5 by deleting one allele of the *CAN1* gene (*can1* $\Delta$ ::*NatMX4/CAN1*), and selecting for loss of the remaining WT allele in clones resistant to canavanine. In order to provide a reference for comparison of LOH rates from JAY270, we introduced these same four constructs in a standard laboratory yeast strain background routinely used to study genome instability mechanisms, including LOH (CG379; Fig. 2.3B) (Morrison et al. 1991; Conover et al. 2015). The rates of LOH were not significantly different between

the two strain backgrounds at Chr12 ( $p=0.225$ ), and were slightly higher in JAY270 at Chr5 ( $p<0.001$ ) and slightly higher in CG379 at Chr13 ( $p=0.021$ ) and Chr4 ( $p<0.001$ ). In general, this analysis indicated that the two strains backgrounds have comparable levels of chromosome stability, and therefore JAY270 genome is not inherently unstable.



### Figure 0.3. Quantitative analyses of LOH.

The measured median rates (blue bars) of single and double LOH events are shown in panels **A** for the JAY270 strain background, **B** for the CG379 strain background, and **C** for CG379  $MATa / mat\Delta$  strains. In the X axis, Chr12, Chr4, or Chr13 indicate diploids with hemizygous insertions of the CORE2 cassette ( $k_{i}URA3-s_{c}URA3-KanMX4$ ) at each of those chromosomes. Chr5 indicates hemizygous deletion of *CAN1*. These strains were used to determine the rates of single LOH individually at each CORE2 insertion or at *CAN1*. Error bars indicate 95% confidence intervals (CI) for each rate measurement. The rates expected for independent double LOH events (red bars; exp.) were calculated by multiplying the respective rates of single LOH at Chr4 and Chr5 or Chr13 and Chr5. The matching rates of observed (obs.) double LOH for each loci pair are shown in blue to the right, measured from strains carrying one copy of CORE2 and one copy of *CAN1*. The expected 95% CI for independent double LOH rates were calculated by multiplying upper or lower 95% CI boundaries from the corresponding single rates, therefore they likely overestimate the expected 95% CI. The rate of single Chr12 LOH for the strain background in panel **C** was not determined (N.D.). Main results in this figure were generated by: NMVS, GMA, MJC.

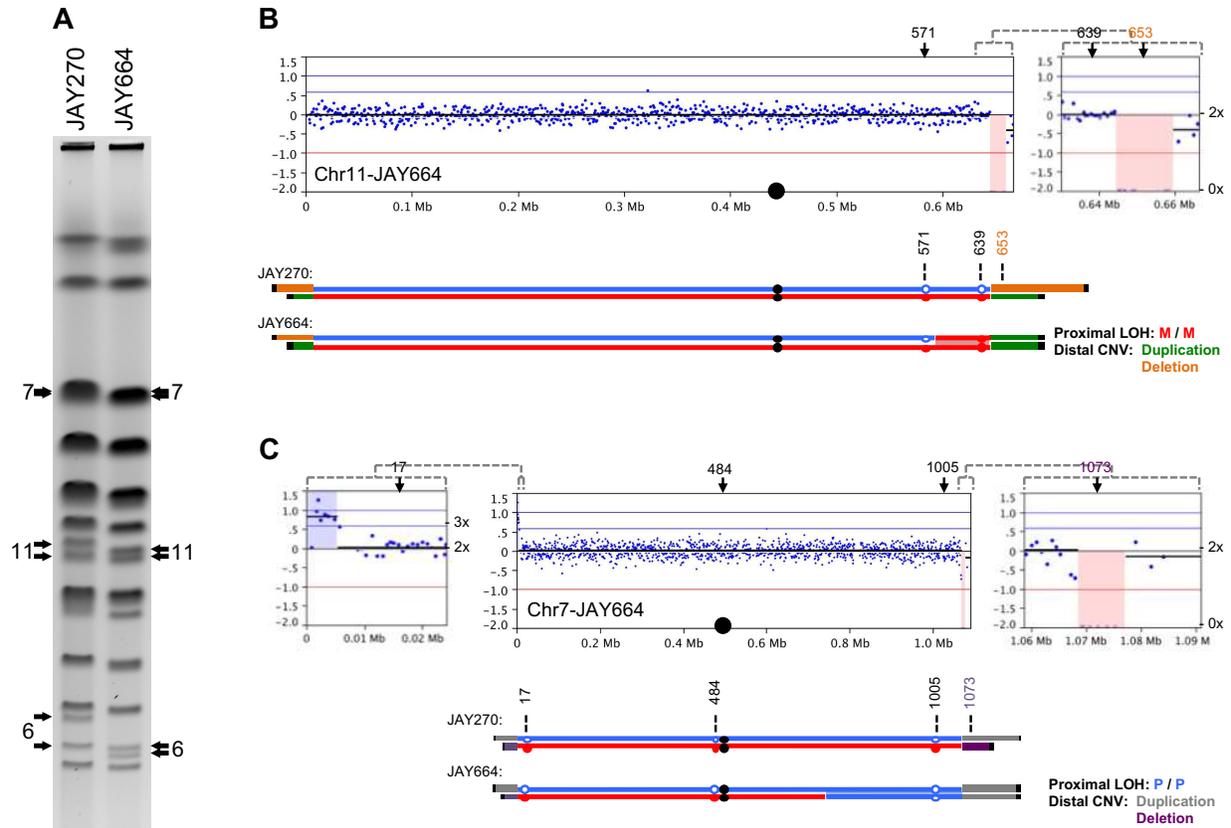
We also used the 5-FOA<sup>R</sup> selection approach to characterize the qualitative nature of the Chr12 LOH tracts in a larger set of independent clones (Fig. S2.5A-B). We used nine PCR-RFLP markers to map LOH tracts from 5-FOA<sup>R</sup> clones derived from CORE2 insertions on Chr12-M and Chr12-P. The patterns observed were similar between the two LOH directions, and resembled the tracts observed in the five initial spontaneous rough colony isolates (Fig. 2.2). All 41 selected Chr12 LOH clones were heterozygous for the left arm, and 39 had uninterrupted unidirectional LOH tracts on the right arm, starting at positions between *CEN12* and *ACE2*, and extending up to the 450 Kb HetSNP. This predominant tract pattern was consistent with a simple interhomolog mitotic crossover mechanism.

The region between *CEN12* and *ACE2* was divided in five intervals delimited by HetSNP markers. The distribution of breakpoints found at these intervals was not significantly different between the strains carrying the CORE2 insertion at Chr12-M or Chr12-P ( $\chi^2 = 0.855$ ;  $p = 0.93$ ), suggesting that both homologs shared similar mitotic recombination properties. We pooled the observed breakpoint distribution data from the 19 5-FOA<sup>R</sup> clones derived from the Chr12-P insertion, and the breakpoints from 23 spontaneous rough isolates that were mapped using WGS (below; Fig. S2.5C and Fig. S2.6). We compared the total number of breakpoints leading to *ace2-A7/ace2-A7* LOH observed within each interval to the expected distribution if breakpoints were allocated purely as a function of the size of the physical interval. This analysis indicated that the observed breakpoint distribution was not significantly different from this simple model ( $\chi^2=6.846$ ;  $p=0.1442$ ) (Fig. S2.5D). Since LOH breakpoints in this region of the genome

were relatively evenly distributed, it does not appear that a mitotic recombination hotspot (*i.e.* fragile site) was present.

### **Genome-wide analysis LOH and CNA in the JAY664 isolate**

In addition to the two copy-neutral Chr12 LOH events and the Chr6 terminal deletion discussed above, we also detected additional structural alterations in the genome of the rough colony isolate JAY664. Using pulse-field gel electrophoresis (PFGE) we detected a size reduction in the long homolog of Chr11 (Fig. 2.4A), and array-CGH showed a 15 Kb full deletion (0 copies) near the right end of that chromosome. (Fig. 2.4B). Analysis of PCR markers proximal to (Chr11 - 639 Kb) and within (Chr11 - 653 Kb) the deletion showed that the deleted region was hemizygous in the JAY270 genome (Fig. S2.4C-E). In JAY664, however, the proximal marker became homozygous, and the distal hemizygous region was lost. The combination of array-CGH and PCR genotyping showed that JAY664 experienced a copy-neutral LOH event with a breakpoint proximal to the Chr11 - 639 Kb marker, leading to homozygosity for the haplotype lacking the hemizygous region represented in the microarray. A similar pattern of proximal LOH and distal CNA was also detected for the right end of Chr7 (Fig. 2.4C and Fig. S2.4F-I). Altogether in JAY664, including PFGE, array-CGH and WGS methods, we detected structural alterations at eight independent regions of the genome.



### Figure 0.4. Analysis of unselected chromosomal changes in JAY664.

**A** shows a PFGE of JAY270 and the JAY664 rough colony isolate. Arrows indicate the chromosomal bands that differed between the two strains, with the respective chromosome number next to them. **B** shows the Chr11 array-CGH copy number plot for JAY664 relative to JAY270. The X-axis indicates the chromosomal coordinates and Y-axis has the  $\text{Log}_2(\text{Cy5-JAY664}) / (\text{Cy3-JAY270})$  ratio, with the corresponding DNA copy number (*i.e.* 0x, 2x, 3x) indicated to the right. Blue dots correspond to individual probes in the array and their  $\text{Log}_2$  Cy5/Cy3 ratios and positions along the chromosome. The inset to the right shows a close-up view of the right end of Chr11 where a full deletion of the probes in the region (pink shading) was detected. The positions of three PCR markers evaluated in the analysis are indicated, and their genotypes are indicated by blue and red circles according to the presence of paternal or maternal alleles (see Table S2.3 and Fig. S2.4). Note that only the sequences from the *S. cerevisiae* S288c reference genome are represented in the array. The probes within the array-CGH deletion signal are hemizygous in JAY270 (orange, 1x), including the 653 Kb PCR marker. JAY664 is homozygous for the maternal right terminal segment of Chr11, and therefore completely lost the signal for the paternal probes in the array, while duplicating maternal JAY270 hemizygous sequences (green). The maternal hemizygous sequences are not present in the S288c genome and thus not represented in the array. The position of the Chr11 LOH breakpoint determined by WGS was near the 581 Kb

HetSNP. Panel **C** shows a similar analysis of Chr7 in JAY664. Gray and purple lines represent the hemizygous regions at the ends of the paternal and maternal homologs, respectively. A proximal LOH event (WGS breakpoint at 784 Kb) and distal CNV event with deletion of one hemizygous region (purple, including the 1073 marker) and duplication of sequences (gray) not present in the array. The inset to the left shows the amplification (blue shading) of probes near the left telomere, but no proximal LOH occurred on that region of Chr7. Instead, the sequences represented by these probes are actually present as a hemizygous insertion near the left telomere of Chr12 in JAY270. An LOH event at that region (see JAY664 Fig. 2.2 and Fig. S2.6) caused them to be duplicated, while deleting hemizygous regions not represented in the array. Main results in this figure were generated by: NMVS, JLA.

### **Investigation of systemic genomic instability**

The remarkably high number of coincident rearrangements observed in JAY664's genome suggested that they might not have accumulated sequentially or independently. Instead, a simpler scenario would be that the observed karyotype arose during a burst of systemic genomic instability in one or a few ancestor cells in the JAY664 lineage. If this model were correct, then the existence of a primary LOH event should increase the likelihood of existence of secondary LOH events, thus other rough colony isolates with a primary selected Chr12 LOH event should also carry a higher than expected number of unselected LOH events elsewhere in their genomes.

To test this hypothesis, we initially analyzed 29 independent smooth clones derived from JAY270, all isolated after five transfer cycles in liquid culture without single cell bottlenecks (~57 cell generations; Methods). Since these clones were smooth, they should not carry LOH on Chr12, and indeed, PCR analysis confirmed that they were all heterozygous *ACE2/ace2-A7* (data not shown). We performed PFGE to detect unselected chromosomal rearrangements in these isolates lacking a primary LOH event. No visible rearrangements were observed in any of the 29 smooth clones (Fig. S2.7) for chromosomes other than Chr12 (rDNA) and Chr8 (*CUP1* tandem repeats).

This observation was corroborated by a second experiment in which we generated two independent mutation accumulation lineages of JAY270 that were cultured with 10 single colony bottlenecks for a total of ~220 cell generations. PFGE analysis of smooth clones from intermediate and final points along the two lineages also showed no visible chromosome size polymorphisms (Fig S2.8).

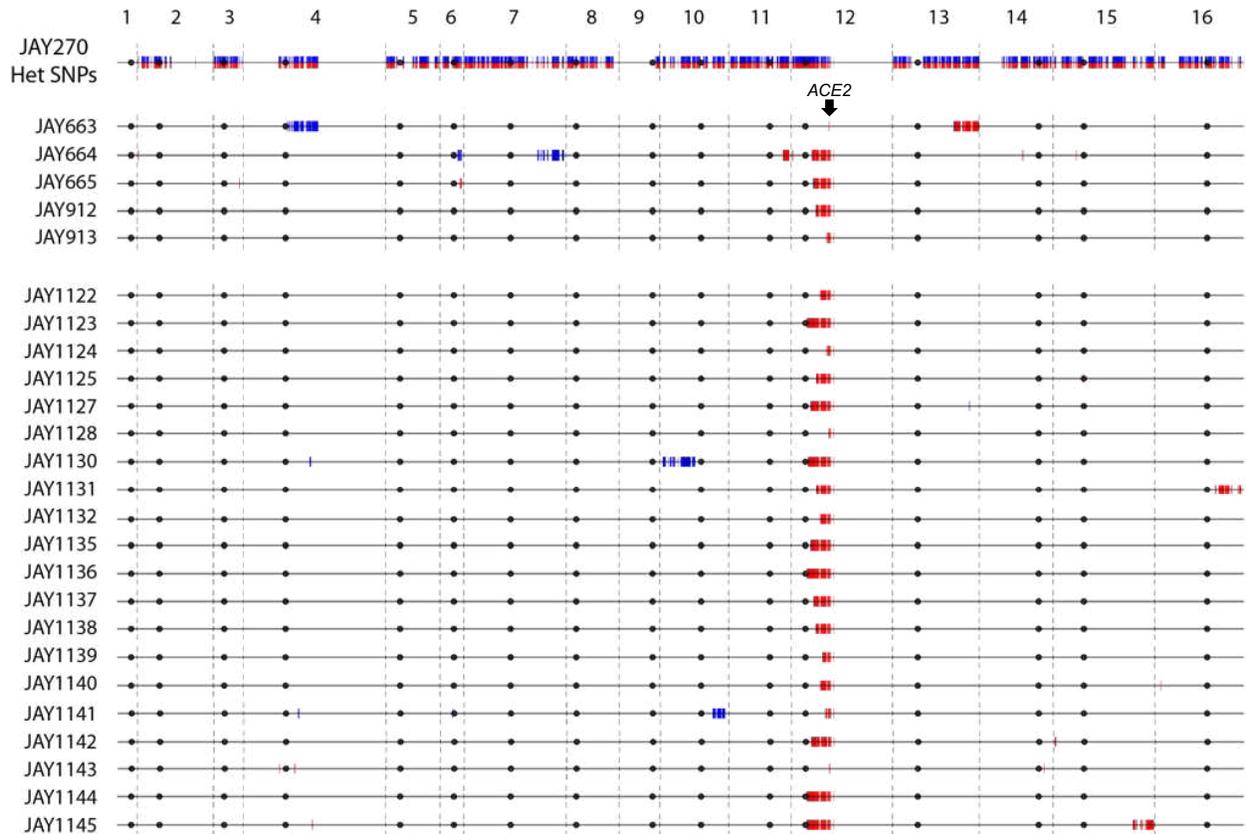
The rate of mitotic crossover for the entire yeast genome has been estimated to be  $\sim 6.2 \times 10^{-4}$  / cell / cell division (St. Charles and Petes 2013). Based on this value, we calculated that roughly one of the 29 smooth clones from the liquid growth regimen, and less than one of the two bottleneck lineages, should have had at least one LOH event. Although the PFGE approach can only detect a subset of structural chromosomal changes (interstitial events, LOH in the longer chromosomes and LOH in chromosomes without structural variation between homologs are not detectable), the lack of any visible chromosomal rearrangements in the smooth clones derived from these two experiments indicated that rearrangements were not abundant in JAY270 when a primary LOH event was not selected.

Next we performed a similar liquid growth without bottlenecking regimen analysis, this time plating ~1,000 cells after every passage cycle to identify rough colonies. Twenty independent spontaneous rough colonies were obtained relatively quickly using this approach. Eleven of them were isolated after ~43 or less cell generations, and only one of them was isolated after more than ~57 (JAY1127; ~85 generations, Table S2.4). PCR analysis of the new rough clones showed that all were homozygous *ace2-A7/ace2-A7* (data not shown). Combined with the original set of five (Fig. 2.2), a total of 25 independent rough colony isolates were analyzed by PFGE (Fig.

2.4 and Fig. S2.9). Strikingly, six of them showed at least one visible size polymorphism for chromosomes other than Chr12 and Chr8. This result showed that, despite its limitations, the PFGE approach was able to detect unselected chromosomal rearrangements in the rough clones, whereas none were seen among the smooth clones. This difference was especially notable since the rough clones were derived from fewer cell generations.

We then used WGS in order to quantify and characterize all LOH events present in the 25 rough colony isolates (Fig. 2.5). This analysis yielded detailed primary LOH tracts for Chr12 in all clones, and revealed a total of 27 unselected secondary LOH tracts. Based on the estimated rate of genome-wide mitotic crossover (St. Charles and Petes 2013), we calculated that only one clone out of the 25 should have had one unselected LOH event, in addition to the selected event spanning the *ACE2* locus ( $6.2 \times 10^{-4}$  crossovers/genome  $\times$  57 cell divisions  $\times$  25 clones). Note that this was an inflated estimate given that most rough clones analyzed were selected after 43 or less cell divisions (Table S2.4), and importantly, LOH is only detectable at the ~60% of the JAY270 genome that is heterozygous (Fig. S2.1). In addition to the high number of total secondary LOH events detected, the distribution of secondary LOH events per clone was also biased. Out of the thirteen clones with unselected LOH, four had two unselected events, three had three unselected events, and finally, one outlier (JAY664) had an astounding seven unselected LOH events. The probability of randomly identifying just one rough clone containing only two unselected LOH events is smaller than  $10^{-5}$ . The probability of having the distribution we actually found is far lower.

Therefore, our WGS analysis indicated the occurrence of systemic bursts of genomic instability leading to LOH.



**Figure 0.5. Genome-wide map of LOH tracts in spontaneous rough colony isolates.**

The top horizontal line is the linear end-to-end depiction of the 16 *S. cerevisiae* chromosomes in JAY270 with HetSNPs represented as paternal (blue) and maternal (red) markers. Chromosome numbers are indicated above, and the position of the *ACE2* locus is shown. Each horizontal line below corresponds to the genomes of the 25 spontaneous rough colony isolates sequenced. Only the HetSNPs that were homozygous P/P or M/M are shown (heterozygous markers are omitted to emphasize visualization of the LOH tracts). As expected from selection for the rough colony morphology, all clones were homozygous for the maternal *ace2-A7* allele (red). In addition, 27 unselected M/M or P/P LOH tracts were detected elsewhere in the genomes of the rough colony isolates. Plots were generated to scale in Python 2.7 using the matplotlib package and a custom script. For size reference, Chr1 is 230 Kb. Main results in this figure were generated by: NMVS, JLA, EPM, PAM, VPA, PC, KTN.

The detailed maps of Chr12 LOH in the 25 sequenced rough clones (Fig. S2.6) showed that 20 tracts were unidirectional, continuous and extended from a position between *CEN12* and the rDNA cluster. In five cases, the tracts had complex discontinuities and some even showed limited LOH for the paternal allele near the breakpoint. Among the unselected LOH tracts (Fig. 2.5), 15 were interstitial (median size 5.0 Kb) and 12 were terminal (median size 232.1 Kb) consistent with gene conversion and mitotic crossover mechanisms, respectively. We also identified 15 *de novo* point mutations among the sequenced clones. However, the small overall number detected was not sufficient to draw conclusions regarding nucleotide mutational signatures. The full WGS LOH and point mutation analyses in the rough clones is presented in Table S2.5.

### **Validation and quantification of coincident LOH**

The discovery of multiple LOH tracts in spontaneous rough colonies suggested the possibility that allelic mitotic recombination, and chromosomal rearrangements in general, may arise and accumulate during systemic genomic instability episodes, rather than independently of each other. Alternatively, it could also be possible that homozygosity at or near the Chr12 *ACE2* locus itself could destabilize the genome, thus increasing the likelihood of secondary rearrangements. We conducted a new round of experiments to validate and quantify the observation of systemic genomic instability, and did so under conditions that removed any possible influence from Chr12 status.

Earlier in the study we used diploid strains individually hemizygous for either the counter selectable *CORE2* cassette or *CAN1* gene, and measured the rate of LOH at

the individual loci by selection for resistance to 5-FOA or canavanine independently. We used the same approach to measure rates of coincident double LOH in diploids that were hemizygous for both *CORE2* and *CAN1*, and selected for simultaneous resistance to 5-FOA plus canavanine. If the occurrence of LOH at a *CORE2* insertion locus were completely independent from the occurrence of LOH at the *CAN1* locus, then the rate of LOH at both loci should be equal to the multiplicative product of the two individual rates. This independent mutation model predicts very low double LOH rates, in the  $10^{-9}$  /cell/division range (Fig.2.3A-C, red bars). The double LOH rates that we measured experimentally were 30-100 fold higher than expected. This was observed in two different strain backgrounds (JAY270 and CG379) for double LOH events at two different pairs of genomic regions (Chr4 and Chr5; Chr13 and Chr5), in all cases with Chr12 remaining unaltered.

Finally, we considered the possibility that the systemic genomic instability mechanism causing the high rates of double LOH could result not from a mitotic process, but instead could be due to sporadic and transient initiation of meiotic recombination in a few cells in the population followed by return-to-growth (RTG) (Laureau et al. 2016). Unlike mitotic recombination, which is sporadic and unscheduled, meiotic recombination is initiated as a well coordinated systemic and genome wide event (Keeney, Lange, and Mohibullah 2014). To distinguish between these mechanisms, we repeated the double LOH measurements in diploids deleted for the *MAT $\alpha$*  locus. These *MAT $\alpha$ /mat $\Delta$*  diploids behaved essentially like haploids. They efficiently mated to a *MAT $\alpha$*  haploid tester strain and lost the ability to sporulate (data not shown). Since *MAT $\alpha$ /mat $\Delta$*  diploids are unable to activate the meiotic developmental

program, they are also unable initiate meiotic recombination. The rate of single LOH in the *MATa/matΔ* diploids was similar or slightly lower than the rate in the *MATa/MATα* diploids, consistently with earlier studies (Pâques and Haber 1999). Importantly, the observed rate of double LOH was ~50 fold higher than expected if the single LOH events were initiated independently of each other. Taken together, our results support the existence of a mitotic systemic genomic instability (mitSGI) process in yeast cells, which is independent and qualitatively different from sporadic initiation of meiosis and RTG.

## Discussion

### **Rough colony phenotype and mutation in *ACE2***

We showed that a colony morphology transition in JAY270 resulted from an LOH event at a region of Chr12 heterozygous for a mutation in the *ACE2* gene. This gene encodes a transcription factor that controls the expression of genes involved in septum destruction and mother-daughter cell separation (Weiss 2012). The *ace2-A7 -1* frameshift allele found in JAY270 leads to a premature stop codon resulting in a truncated, likely inactive, protein that lacks three zinc finger domains and a nuclear localization sequence (McBride, Yu, and Stillman 1999). Interestingly, another heterozygous diploid industrial strain, FostersB used in brewing (Borneman et al. 2011), has a +1 frameshift variant in the same adenine homopolymer region of *ACE2* (*ace2-A9* allele). The *ACE2* gene has also been shown to be involved in the transition between the yeast and hyphal forms of *C. albicans*, an important trait for pathogenesis. In that context, inactivation of the *ACE2* ortholog contributes to the formation of cell filaments

during hyphal growth, and an alternative isoform of the Ace2 protein helps prevent inappropriate activation of cell detachment from hyphae (Calderón-Noreña et al. 2015). In *S. cerevisiae*, and in particular in JAY270 during bioethanol production, it is not known if or how the switch between dispersed to aggregated cell growth states has an effect on fitness. However, it is possible that JAY270's heterozygous *ACE2/ace-A7* genotype may offer an advantage by giving the population the ability to quickly access the aggregated state through LOH. This may provide a short-term adaptive solution as cells encounter various environmental challenges during industrial fermentation (Basso et al. 2008). Individuals in the population could then return to the dispersed state through expansion or contraction of the adenine run (Tran et al. 1997), or through sporulation and mating to an *ACE2* haploid.

### **mitSGI and precedents of coincident recombination**

Beyond the genetic characterization of the rough colony phenotype in JAY270, this study allowed us to uncover the mitSGI phenomenon through which multiple LOH events can accumulate in a cell lineage. Using PFGE, array-CGH and WGS, we determined that yeast clones carrying a primary selected LOH tract at Chr12 were more likely than expected to carry unselected LOH tracts. We also showed in quantitative LOH assays that combinations of double LOH at Chr5 and Chr4 or Chr13, occurred at rates 30-100 fold higher than expected if single LOH events occurred independently. We interpret these results as evidence for the occurrence of bursts of genomic instability leading to multiple LOH events over one or few mitotic cell generations.

Spontaneous mitotic recombination events like the ones described here are triggered by local DNA lesions and/or replication fork collapse episodes, which then lead to chromosomal breakage and allelic HR repair using the homolog as template (Symington, Rothstein, and Lisby 2014). Such precursor lesions are thought to occur randomly in vegetative cells, both spatially and temporally, therefore mitotic recombination events involving different chromosomes should not be coincident. In contrast, meiotic recombination is known to be a systemic genetic variation process, since it occurs simultaneously throughout the genome and involves intricate coordination between generation and repair of genome-wide double strand breaks (Keeney, Lange, and Mohibullah 2014).

While our study is, to our knowledge, the first to describe the mitSGI phenomenon through the lens of high-resolution genome-wide analytical methods, there have been sparse reports of elevated coincident mitotic recombination in yeasts as well as in mammalian cells dating back decades (Fogel and Hurst 1963; Hurst and Fogel 1964; Minet, Grossenbacher, and Thuriaux 1980; Golin and Tampe 1988; Freeman and Hoffmann 2007; Forche et al. 2009; C. Y. Li, Yandell, and Little 1994; Grygoryev et al. 2014). The typical experimental design in those cases was to select clones for carrying a recombination event at a primary locus, and then screening the resulting clones for the occurrence of secondary unselected recombination at one or a limited number of unlinked loci. The same intriguing observation, shared in all cases, was a frequency of coincident recombination that was higher than that predicted assuming the individual events occurred independently.

In some of the yeast studies, the high coincident recombination rates were interpreted as being derived from a small number of cells within the replicating population that spuriously entered the meiotic developmental program, or transiently experienced a “para-meiotic” state, but reverted back to mitotic growth (Hurst and Fogel 1964; Minet, Grossenbacher, and Thuriaux 1980). A recent study specifically characterized this type of return-to-growth (RTG) event and the genome-wide recombination outcomes associated with it (Laureau et al. 2016). The authors often detected a large number of LOH tracts per clone (minimum of 5, average of ~30, and up to 87), indicating that the RTG induction leads to abundant and widespread recombination. Another notable finding was that while interstitial LOH (gene conversion-like; GC) tracts were frequent, their sizes were relatively constrained (2.3 Kb on average). This measurement is notable because it is consistent with GC tract sizes measured in haploids derived from complete meiotic divisions; ~2 Kb median size (Mancera et al. 2008). In contrast, GC tracts associated with mitotic recombination tend to be significantly longer, approximately 5-6 Kb median size (St Charles and Petes 2013). This variation in typical GC tract sizes is likely a reflection of subtle mechanistic differences in the processing of HR intermediates between meiotic and mitotic cells.

Despite the possibility discussed above, there are several reports of high coincident recombination in proliferating cells in which the induction of a full meiotic cycle, RTG or para-meiosis were either unlikely or could be ruled out entirely. One study in *S. cerevisiae* specifically measured the formation of spurious haploids from mitotic diploid cultures displaying high coincident intragenic recombination at unlinked pairs of heteroalleles (Freeman and Hoffmann 2007). The authors found that while haploids did

form in their cultures, the frequency was far below that needed to influence the formation of double recombinants, thus concluding that a low level of cryptic meiosis was not a likely contributor. In addition, one of the seminal studies (Forche et al. 2009) of LOH in *C. albicans* (a species devoid of a conventional sexual cycle (D'Enfert et al. 2017)), reported data that closely parallel our own observations. First, the authors selected clones for the presence of a primary LOH event at the *GAL1* locus on chromosome 1. Then, using a low resolution SNP-array platform, they detected frequent unselected secondary LOH tracts among clones carrying the primary event, but rarely in control clones still heterozygous at *GAL1*. In addition, selection for LOH at *GAL1* was associated with the emergence of altered colony morphology phenotypes, presumably derived from rearrangements elsewhere in the genome. Accordingly, clones displaying altered morphology were enriched for the presence of unselected LOH tracts when compared to clones with normal morphology.

Another important pair of precedents of mitSGL observations comes from experiments conducted in mammalian systems. These used either human TK6 lymphoblastoid cells in culture (C. Y. Li, Yandell, and Little 1994), or mouse kidney cells *in vivo* and in culture (Grygoryev et al. 2014). In both cases, the starting cells were heterozygous for mutations at the counter-selectable markers, *TK* and *Aprt*, respectively, enabling the selection of clones carrying a primary LOH event at those loci. Subsequently, the presence of secondary LOH tracts was assessed at roughly a dozen loci elsewhere in the human or mouse genomes. The two studies found that secondary LOH was more frequent in clones selected for carrying the primary LOH event than in controls clones that remained heterozygous. These studies demonstrated

that mitSGI also exists in metazoans, and can be detected in cells that are exclusively mitotic, thus ruling out a contribution from meiotic recombination, at least in these contexts.

The studies outlined above suggest that cryptic initiation of meiosis in a small number of cells can in some cases lead to systemic genomic instability, however, we favor the interpretation that the events analyzed in our study originated primarily from *bona fide* mitotic cells. The recent work by Laureau *et al.* clearly defined the features of systemic LOH caused by meiotic initiation followed by RTG. The pattern we detected in our study was different, and instead was consistent with mitotic patterns. The number of unselected interstitial LOH (GC) tracts per clone we detected was small (typically 1 or 2) and their sizes were long (median 5.0 Kb). This was reinforced by the observation that *MATa/matΔ* diploids, incapable of entering the meiotic developmental program, continued to display double LOH rates that were far higher than expected from independent events.

### **mitSGI-like observations in human disease**

In addition to the experimental examples above, our results also resemble recent reports of bursts of mitotic genomic instability in humans during cancer genome evolution and early development. Specifically, genome-wide copy number profiling of thousands of individual cells isolated from tumors in 12 patients with triple-negative breast cancer revealed that a large number of CNAs were acquired within a short period of time at the early stages of tumor development (Gao et al. 2016). Most of these CNAs were shared between several cells from a same tumor, suggesting the occurrence of a

burst of genomic instability in one or few initiating cells followed by a long period of stable clonal expansion. Although the study had power to detect gradual accumulation of mutations, no clones with intermediate CNA profiles were identified, suggesting a punctuated model of mutation accumulation.

Another pertinent parallel is the recent analysis of patients with genomic disorders that carry multiple *de novo* constitutional CNVs (MdnCNVs; (Liu et al. 2017)). Typically in those patients, only one of the structural variants was the primary event causing the symptoms associated with the disorder. The additional CNVs were secondary, occurred at unrelated regions, and apparently formed during a short burst of genomic instability at some point in the perizygotic time interval. The changes then propagated stably during development to be found in all cells in the patients. Taken together, these results suggest that mitSGI processes may be universal and may play an important role in human disease development.

### **Possible mechanisms underlying mitSGI**

Our results so far suggest that mitSGI is not likely associated with initiation of meiotic recombination and RTG. However, the specific causes for the existence of a small subset of recombination-prone cells within a normal mitotic population remain to be determined. This phenomenon could well have multiple and distinct origins, however, we favor two non-exclusive mechanisms, related to cellular ageing and stochastic gene expression. These two models are attractive because they are transient in nature, which would support stable transmission of rearranged genomic structures after the systemic vulnerability time window has passed.

The first scenario is that clones carrying multiple unselected LOH events originated from replicatively old mother cells. This model stems from the observation of a marked increase in the rate of LOH in daughter yeast cells budded from mothers that had undergone ~25 cell divisions (McMurray and Gottschling 2003), relatively old within the context of a maximum *S. cerevisiae* replicative lifespan of ~40. Subsequent work from the same group showed that this increase in nuclear genomic instability was strongly correlated with the initial appearance of mitochondrial DNA loss and/or damage in the old mother cells (Veatch et al. 2009). In our study, however, all of the spontaneous rough colony isolates analyzed retained normal respiratory activity (all were non-*petite*; grew on non-fermentable carbon sources), so they must have had integral mitochondrial genomes. They also did not show signs of continual genomic instability. Therefore, if replicative aging were an underlying factor in mitSGI, it would be through a pathway that does not involve loss of mitochondrial function.

Another explanation for a subpopulation of hyper-recombinogenic cells involves heterogeneities that exist even within an isogenic population. Specifically, cell-to-cell variation (*i.e.* noise) in gene expression has been reported in organisms ranging from prokaryotes, to yeast, to humans (Raj and van Oudenaarden 2008). It is plausible that stochastic variation in the expression of a broad class of genes involved in genome stability could cause specific protein levels to drop below those required for optimal function. A recent comprehensive genome stability network analysis identified 182 genes involved in suppression of gross chromosomal rearrangements (Putnam et al. 2016), and an earlier genetic screen identified 61 genes specifically involved in suppressing LOH (Andersen et al. 2008). In this scenario, rare individual cells that fail to

adequately express any of these genes could effectively behave as null mutants for a short period. Some of these genes act cooperatively, therefore concomitant loss of activity causes extreme levels of genomic instability. For example, double knockouts for *TEL1* and *MEC1*, encoding critically important DNA damage response proteins (orthologs of mammalian ATM and ATR, respectively), show marked increase in mitotic genomic instability (Craven et al. 2002), often accumulating multiple genome rearrangements (Serero et al. 2014). A similar extreme phenotype might be expected in a wild type cell that by chance simultaneously had a critically low level of transcription of two genome stability genes. Likewise, overexpression of single genes encoding a subunit of a genome stability multi-protein complex could lead to a dominant negative phenotype that temporarily impairs function. Importantly, the hyper-recombinogenic state of these individuals would be completely reversible once the descendant cells returned to the gene expression levels typical of most individuals in the population. This mechanism could explain the observed stable clonal expansion that followed mitSGI in our spontaneous rough colony clones, as well as in the recent *in vivo* human studies (Gao et al. 2016; Liu et al. 2017).

The analysis of spontaneous rough morphology clones provided unprecedented detailed information about the nature and frequency of secondary recombination events resulting from the mitSGI process. Our study also provides a unifying context for the interpretation of classic and recent reports of coincident recombination in yeasts, in mammalian experimental systems, and in human disease. The combination of whole genome analyses and the double LOH selection approach described here offer a

powerful experimental platform to further dissect the core mechanisms responsible for the mitSGI phenomenon.

## **Materials and Methods**

### **Growth media**

Yeast cells were grown in YPD (20 g/L glucose, 20 g/L peptone, 10 g/L yeast extract, 20 g/L bacteriological agar for solid media) and synthetic media (20 g/L glucose, 5 g/L ammonium sulfate, 1.7 g/L yeast nitrogen base without aminoacids, 1.4 g/L complete drop-out mix, 20 g/L bacteriological agar). Transformants carrying the *natMX* or *hphMX* cassettes were selected in YPD media supplemented with 200 mg/L of nourseothricin (Nat) or 300 mg/L of hygromycin, respectively. *URA3* transformants were selected in uracil drop-out synthetic media. Counter-selection against *URA3* and *CAN1* were performed in synthetic media supplemented with 1g/L of 5-Fluoroorotic Acid (5-FOA) and 60 ml/L of canavanine in arginine drop-out, respectively.

### **Yeast genetic backgrounds and procedures**

*Saccharomyces cerevisiae* strains used in this study descended from either the JAY270 or CG379 strain backgrounds (Table S2.1). JAY270 is a heterothallic diploid single colony isolate derived from the industrial bioethanol strain PE-2 (Argueso et al. 2009). CG379 diploid is related to the S288c laboratory strain background (Morrison et al. 1991; Argueso et al. 2008). Standard procedures for yeast transformation, crossing and sporulation were followed (Ausubel et al. 2003).

### **Isolation of spontaneous rough colonies derived from JAY270**

The five spontaneous rough colony isolates described early in the study were identified while visually inspecting colonies growing on YPD agar derived from plating aliquots of liquid cultures of JAY270, or JAY270 isogenic diploids. In some cases, the liquid cultures had been passaged more than once, although the precise number of growth cycles is not known.

The rough colonies obtained later in the study were derived from the following controlled procedure. JAY270 cells were refreshed from a frozen stock, streaked to single colonies in YPD plates and incubated for 48 hours at 30°C. 22 independent colonies were inoculated into 5 ml of liquid YPD cultures and grown for 24 hours, reaching saturation. At that point, a dilution was plated in YPD agar to screen for the presence of rough colonies, and a 50  $\mu$ l aliquot (1% of cell population) was transferred to fresh 5 ml liquid medium to start the next growth cycle. Whenever a rough colony was identified on the YPD plates, the isolate was frozen for further analysis and its respective downstream liquid culture lineage was discarded. Information on the number of cycles necessary for the identification of spontaneous rough colony isolates for each culture is provided in Table S2.4. A similar procedure was used for the selection of control smooth isolates, but with a predetermined growth regimen. Liquid cultures were started from smooth single colonies and passaged for 5 consecutive liquid YPD growth cycles, at which point a dilution was plated in solid YPD media and one random smooth colony per culture was frozen for further analysis.

### **Allele replacement and complementation tests**

A two-step allele replacement plasmid was constructed by cloning a segment of the wild type *ACE2* allele containing the 8-adenine homopolymer into the *URA3* pRS306 integrative vector (Sikorski and Hieter 1989). First, pRS306 was digested with *Bam*HI and *Kpn*I and ligated to a *Bgl*II- and *Kpn*I-digested segment of the wild type *ACE2* allele (PCR amplified from JAY289 genomic DNA using primers JAO904 and JAO905). The resulting plasmid, pAR1, was linearized with *Eco*RI and transformed into JAY291 and JAY292 haploids (both mutant *ace2-A7*). Pop-in Ura<sup>+</sup> integration transformants were selected, and pop-out events were then selected for resistance to 5-FOA. 5-FOA resistant candidates were screened by PCR and Sanger sequencing to verify the presence of the wild type *ACE2* allele. This procedure resulted in the *ACE2* strains JAY1051 and JAY1039, isogenic to JAY291 and JAY292 respectively. Complementation tests using these strains were performed as shown in Fig. S2.2D.

### **Construction of strains used in single and double LOH assays**

The JAY270 strains used in the LOH assays were constructed from a homozygous *ura3/ura3* derivative of JAY270 (JAY585; gift from F. Galzerani). The CORE2 cassette containing the *Kluyveromyces lactis URA3* gene, the *S. cerevisiae URA3* gene and the *kanMX* geneticin resistance marker (*KIURA3-ScURA3-KanMX4* CORE2) was amplified from pJA40 (Zhang et al. 2013) with primers targeting three different genomic regions (Table S2.2). Primers JAO1074 and JAO1075 were used for targeted integration proximal to the *ACE2* locus in Chr12. Integration in the maternal or paternal homolog was checked by PCR and Sanger genotyping of linked HetSNPs.

Primers JAO506 and JAO507 were used for integration of CORE2 distal to *SSF2* in Chr4, and JAO502 and JAO503 for integration proximal to *ADH6* in Chr13. Transformation of JAY585 with each cassette resulted in JAY865 (*SSF2/SSF2::CORE2*) and JAY868 (*ADH6/ADH6::CORE2*), which were used in the single 5-FOA<sup>R</sup> LOH assays. Subsequently, one native copy of the *CAN1* gene was deleted from each of these strains using the *natMX* cassette. The cassette was amplified from pAG25 (Goldstein and McCusker 1999) using primers JAO271 and JAO272 and transformed into JAY865 and JAY868, resulting in JAY1804 and JAY1812, respectively. These strains were used for Can<sup>R</sup> single LOH assays and for Can<sup>R</sup> - 5-FOA<sup>R</sup> double LOH assays. The same procedure was followed to build LOH assay strains in the CG379 background, resulting in strains JAY861 and JAY859 (5-FOA<sup>R</sup> single LOH assays), and strains JAY1567 and JAY1569 (Can<sup>R</sup> single LOH assays and Can<sup>R</sup> - 5-FOA<sup>R</sup> double LOH assays). We further manipulated JAY1567 and JAY1569 to create *MATa/matΔ* isogenic derivatives. The *hphMX* cassette was amplified from pAG32 (Goldstein and McCusker 1999) using the primers JAO1440 and JAO1441. This cassette was used to replace a segment of the *MATα* allele in JAY1567 and JAY1569, resulting in JAY1808 and JAY1809 respectively.

### **Quantitative LOH rate assays**

Yeast cells were streaked to single colonies on solid YPD medium and incubated at 30°C for two days. Single colonies were inoculated into 5 ml liquid YPD, and incubated for 24 hours at 30° C in a rotating drum. The cultures were serially diluted and plated on YPD (permissive), and 5-FOA (selective) and/or YPD plus canavanine

(selective). Colonies were counted after 2 days of growth on permissive and 4 days on selective plates, and colony counts were used to calculate recombination rates and 95% confidence intervals using the Lea & Coulson method of the median within the FALCOR web application [<http://www.keshavsingh.org/protocols/FALCOR.html>] ] (Lea and Coulson 1949; Hall et al. 2009). Statistical analyses of pairwise comparisons between recombination rates were performed using a two-sided nonparametric Mann Whitney test in GraphPad Prism software.

### **Genome Sequencing Analyses**

The genomes of the JAY270 parent strain and of 56 haploids derived from 14 complete JAY270 tetrads were sequenced using the Illumina short read whole genome sequencing platform. We also sequenced the genomes of 25 rough colony isolates derived from JAY270. The sequencing reads were used in the four analyses described below. Genome sequencing data associated with this study is available in the Sequence Read Archive (SRA) database under study number SRP082524.

#### *JAY270 draft HetSNP map construction*

We developed a map of heterozygous SNPs in JAY270 using a high stringency approach that would identify only high confidence sites. This was a conservative approach that is therefore it is likely to be missing some loci, but it is unlikely to contain any false calls. We took the reads from 44 haploid spores from 11 complete JAY270 tetrads and applied two parallel analyses. These 11 tetrads had been sequenced at the

same time and had uniform coverage and read lengths. The data from the additional 3 tetrads were not as homogeneous so they were not used for this aspect of the work.

In the first analysis we combined all haploid reads to simulate an ultra deep coverage sequencing dataset from the JAY270 diploid. The data from each haploid was first processed to include only high quality reads ( $Q>30$ ), and the ends were trimmed to obtain 90 nt reads. Next we determined the haploid with the lowest number of reads within each tetrad. All the reads from this haploid and an equal number of random reads from each of its three sibling haploids were selected for the next phase. This ensured an equal number of reads contributed by each haploid within each tetrad. Finally, we combined all processed and intra-tetrad number-adjusted reads from the 44 haploids to generate the simulated JAY270 sequencing dataset. This set was composed of ~129 million reads for a mean depth coverage of ~800 reads per base. We aligned these reads to the *S. cerevisiae* S288c reference genome and independently called out SNPs using GATK (McKenna et al. 2010) and Samtools (H. Li et al. 2009), limiting the analysis to SNPs with coverage higher than 200 and allele frequency between 0.4 and 0.6. We then obtained a list of 13,594 candidate HetSNPs found by both approaches, all had allele frequencies close to 0.5.

For the second analysis we aligned the reads from each individual haploid to the reference genome and called the SNPs using GATK (McKenna et al. 2010), identifying 18,201 sites. Next we aligned the calls from each group of four haploids belonging to the same tetrad and determined the segregation ratio for each of the SNPs within each tetrad.

We then took the 13,075 sites that were discovered in both approaches and filtered to improve the confidence of the heterozygous calls. We retained only the sites that had a Mendelian 2:2 segregation in at least 9 of the 11 tetrads for sites located in central regions of chromosomes (defined by the first and last genes annotated as essential in SGD). We used a stricter filter of all 11 tetrads displaying 2:2 for sites located at distal regions to avoid confounding effects from sites present at subtelomeric repeated gene families. The segregation filtering resulted in 12,197 sites, which we then narrowed down manually (mostly by removing subtelomeric sites) to arrive at the final list of 12,023 high confidence HetSNPs shown in Fig S2.1. Note that this draft list is limited only to allelic sites that have one nucleotide that matches the S288c reference and the other that is a variant present in JAY270. It does not contain sites in which two nucleotide variants are present at the same site, nor short nucleotide insertion or deletions (with the exception of *ace2-A7*; Chr12\_405,714), nor larger structural variants. A comprehensive genome assembly of the JAY270 diploid genome will be described elsewhere.

#### *HetSNP phasing.*

The availability of genome sequencing data from multiple complete tetrads allowed us to deduce the phasing association between the JAY270 HetSNPs. To do so, we initially arbitrarily assigned the S288c reference bases to one phased haplotype and all alternative bases to the other haplotype. Next we aligned the genotypes of 56 haploids along the HetSNP list and determined the positions of crossover events between the two haplotypes within each of the respective 14 tetrads. Considering that

meiotic crossovers are very unlikely to occur at exactly the same position in different tetrads, we could make corrections to the arbitrary phasing to minimize the number of crossovers. In most cases, apparent 4-chromatid double crossovers were observed at the same interval in all tetrads, indicating an error in the arbitrary phasing. A simple correction at those sites resulted in the much more likely scenario of no crossovers at that interval in any of tetrads. In cases where actual crossovers occurred in one or a few of the tetrads, they typically were 2-chromatid single crossover events that we could also clearly identify and correct the arbitrary phasing accordingly to minimize the number of crossovers. We did the analysis and phasing corrections manually over three sequential iterations arriving at the phased HetSNP list. For most chromosomes where the physical distance between consecutive HetSNPs was short we were able to unambiguously deduce a single linkage group. In a few cases, either at long intervals delimited by distant consecutive HetSNPs, or sites of possible meiotic recombination hot spots, there was ambiguity in the phase calling, so we broke down the respective chromosome in multiple linkage groups. Overall for the 15 chromosomes with heterozygosity in them, 10 yielded a single linkage group, 4 had two groups, and 1 had three groups. No phasing could be done for Chr01 since it was fully homozygous. Once the phasing was completed and the haplotypes were defined, we arbitrarily named one of them maternal (M) and the other paternal (P) to facilitate the subsequent LOH tract analyses.

### Detection of LOH tracts

LOH tracts were detected using CLC genomics workbench software to map sequencing reads from the JAY270 parent strain and from 25 spontaneous rough colony isolates on to the S288c reference, and detecting SNPs across the whole genome. Low stringency detection parameters were set such that SNPs present at frequencies higher than 0.05 were identified. We then interrogated the 12,023 loci in the JAY270 HetSNP list, determining the nucleotides present at those positions and their relative frequency. When no SNPs were detected at those specific positions, the genotype was called as homozygous for the reference nucleotide. When the alternative nucleotide was detected at higher than 0.95 the genotype was called homozygous for the alternative nucleotide. Alternative nucleotides detected at frequencies between 0.1 and 0.9 resulted in a heterozygous call for that locus. After the genotypes were called they were then converted to the respective haplotype designations as MAMA/MAMA and PAPA/PAPA homozygous, and MAMA/PAPA heterozygous. All 12,023 loci were called as heterozygous from the analysis of JAY270 reads. LOH tract sizes were estimated by calculating the positions of breakpoints to the right and to the left, and subtracting the left side position from the right side. Breakpoint positions were calculated as the average position between the two HetSNPs that defined the transition from heterozygosity to homozygosity. For terminal LOH tracts the coordinates of the left or right telomeres were used as the breakpoint positions. LOH tracts were called even if they included a homozygous at a single marker HetSNP. Six such cases were identified. They were all interstitial and had median tract size was estimated as 5.4 Kb, consistent with gene conversion events at regions of low marker density. A subset of

these single marker LOH calls were independently validated by PCR and Sanger sequencing. The median size of interstitial LOH tracts involving two or more markers was very similar: 4.8 Kb.

#### *Mapping of the ace2-A7 mutation*

The position of the mutation responsible for the rough colony phenotype was determined by aligning the genotype calls at the 12,023 HetSNP sites for the 12 haploids used in the crosses in Fig. 2.11. Haploids were grouped according to whether they were inferred from the crossing phenotypic outcomes to carry the wild type or mutant allele. We filtered the genotype calls to identify genomic segments where all members of the wild type group had the same genotype, and conversely all members of the mutant group had the other genotype. Only the regions on Chr11 and Chr12 shown in Fig. S2.2 satisfied this strict co-segregation criterion. The *ACE2* gene was identified within the Chr12 region through analysis of its functional genomic annotation.

#### *Detection of de novo point mutations*

A list of *de novo* point mutations in the genome of the 25 rough colony isolates was obtained through multiple sample genotype calling of the JAY270 genome with the sequencing data of each clone using GATK (McKenna et al. 2010) haplotype caller. Reference and alternative allele counts were used for statistical analysis to determine the significance of the allelic variation for each position identified (Fisher test  $p < 0.0001$ ). All statistically significant mutations were further examined for strand bias using the Fisher Score and variants with  $FS > 50$  were filtered out of the list. We further refined the

list by excluding variants with total read count <200 for JAY270 and <30 for the rough colony derivatives. Mutations were further filtered based on gene start and end positions. Because the Fisher test has low accuracy when alternative allele frequency is low, we filtered the list for positions with alternate allele frequency >0.2. We compared these results to a list of mutations called using both Samtools (H. Li et al. 2009) and GATK (McKenna et al. 2010). Mutations shared across these lists were then manually validated by inspection of the read mappings to the reference genome (using CLC Genomics Workbench) and resulted in the final list of *de novo* point mutations shown in Table S2.5. Because of the very stringent parameters used, this list is unlikely to contain false positives, but some true mutations could have been missed.

### **Genotyping specific HetSNPs through PCR-RFLP and molecular karyotyping**

Genotyping was performed by PCR-amplification of regions containing the HetSNPs followed by restriction digestion and agarose gel analysis fragment length polymorphisms. In cases of markers where the HetSNPs were not associated with an RFPL, PCR products were Sanger-sequenced. A complete list of the HetSNPs coordinates, primers and restriction enzymes used is provided in Table S2.3. Pulsed-Field Gel Electrophoresis (PFGE) was used to identify the chromosomal length polymorphisms in rough and smooth, and array-CGH was used to evaluate copy number variation in JAY664 relative to the JAY270 parent strain. PFGE and array-CGH analyses were performed as described previously (Zhang et al. 2013).

## REFERENCES

- Andersen, Marguerite P., Zara W. Nelson, Elizabeth D. Hetrick, and Daniel E. Gottschling. 2008. "A Genetic Screen for Increased Loss of Heterozygosity in *Saccharomyces Cerevisiae*." *Genetics* 179 (3): 1179–95. <https://doi.org/10.1534/genetics.108.089250>.
- Argueso, Juan Lucas, Marcelo F. Carazzolle, Piotr A. Mieczkowski, Fabiana M. Duarte, Osmar V C Netto, Silvia K. Missawa, Felipe Galzerani, et al. 2009. "Genome Structure of a *Saccharomyces Cerevisiae* Strain Widely Used in Bioethanol Production." *Genome Research* 19 (12): 2258–70. <https://doi.org/10.1101/gr.091777.109>.
- Argueso, Juan Lucas, James Westmoreland, Piotr A Mieczkowski, Malgorzata Gawel, Thomas D Petes, and Michael A Resnick. 2008. "Double-Strand Breaks Associated with Repetitive DNA Can Reshape the Genome." *Proceedings of the National Academy of Sciences* 105 (33): 11845–50.
- Ausubel, Frederick M., Roger Brent, Robert E. Kingston, David D. Moore, J. G. Seidman, John A. Smith, and Kevin Struhl. 2003. *Current Protocols in Molecular Biology*. *Current Protocols in Molecular Biology*.
- Babrzadeh, F., R. Jalili, C. Wang, S. Shokralla, S. Pierce, A. Robinson-Mosher, P. Nyren, et al. 2012. "Whole-Genome Sequencing of the Efficient Industrial Fuel-Ethanol Fermentative *Saccharomyces Cerevisiae* Strain CAT-1." *Molecular Genetics and Genomics* : MGG 287 (6): 485–94.
- Basso, L.C., H.V. Amorim, A.J. Oliveira, and M.L. Lopes. 2008. "Yeast Selection for Fuel Ethanol Production in Brazil." *FEMS Yeast Research* 8 (7): 1155–63.
- Bennett, Richard J., Anja Forche, and Judith Berman. 2014. "Rapid Mechanisms for Generating Genome Diversity: Whole Ploidy Shifts, Aneuploidy, and Loss of Heterozygosity." *Cold Spring Harbor Perspectives in Medicine* 4 (10): 1–19. <https://doi.org/10.1101/cshperspect.a019604>.
- Borneman, A. R., B. A. Desany, D. Riches, J. P. Affourtit, A. H. Forgan, I. S. Pretorius, M. Egholm, and P. J. Chambers. 2011. "Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of *Saccharomyces Cerevisiae*." *PLoS Genetics* 7 (2).
- Calderón-Noreña, Diana M., Alberto González-Novo, Sara Orellana-Muñoz, Pilar Gutiérrez-Escribano, Yolanda Arnáiz-Pita, Encarnación Dueñas-Santero, M. Belén Suárez, et al. 2015. "A Single Nucleotide Polymorphism Uncovers a Novel Function for

the Transcription Factor Ace2 during *Candida Albicans* Hyphal Development.” *PLoS Genetics* 11 (4): 1–25. <https://doi.org/10.1371/journal.pgen.1005152>.

Charles, Jordan St., and Thomas D. Petes. 2013. “High-Resolution Mapping of Spontaneous Mitotic Recombination Hotspots on the 1.1 Mb Arm of Yeast Chromosome IV.” *PLoS Genetics* 9 (4). <https://doi.org/10.1371/journal.pgen.1003434>.

Charlesworth, Deborah, Deborah Charlesworth, John H Willis, and John H Willis. 2009. “The Genetics of Inbreeding Depression.” *Nature Reviews. Genetics* 10: 783–96. <https://doi.org/10.1038/nrg2664>.

Chen, Z Jeffrey. 2013. “Genomic and Epigenetic Insights into the Molecular Bases of Heterosis.” *Nature Reviews. Genetics* 14 (7): 471–82. <https://doi.org/10.1038/nrg3503>.

Conover, Hailey N., Scott A. Lujan, Mary J. Chapman, Deborah A. Cornelio, Rabab Sharif, Jessica S. Williams, Alan B. Clark, Francheska Camilo, Thomas A. Kunkel, and Juan Lucas Argueso. 2015. “Stimulation of Chromosomal Rearrangements by Ribonucleotides.” *Genetics* 201 (3): 951–61. <https://doi.org/10.1534/genetics.115.181149>.

Coste, Alix, Vincent Turner, Françoise Ischer, Joachim Morschhäuser, Anja Forche, Anna Selmecki, Judith Berman, Jacques Bille, and Dominique Sanglard. 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 (4): 2139–56. <https://doi.org/10.1534/genetics.105.054767>.

Craven, Rolf J., Patricia W. Greenwell, Margaret Dominska, and Thomas D. Petes. 2002. “Regulation of Genome Stability by TEL1 and MEC1, Yeast Homologs of the Mammalian ATM and ATR Genes.” *Genetics* 161 (2): 493–507.

D’Enfert, Christophe, Marie-elisabeth Bougnoux, Adeline Feri, Melanie Legrand, Raphael Loll-krippleber, Timea Marton, Corinne Maufrais, Jeanne Ropars, Natacha Sertour, and Emilie Sitterlé. 2017. “Genome Diversity and Dynamics in *Candida Albicans*.” *Cellular and Molecular Bioblogy*, 205–32. <https://doi.org/10.1007/978-3-319-50409-4>.

Fogel, S., and D. D. Hurst. 1963. “Coincidence Relations between Gene Conversion and Mitotic Recombination in *Saccharomyces*.” *Genetics* 48 (March): 321–28.

Forche, Anja, P. T. Magee, Anna Selmecki, Judith Berman, and Georgiana May. 2009. “Evolution in *Candida Albicans* Populations during a Single Passage through a Mouse Host.” *Genetics* 182 (3): 799–811. <https://doi.org/10.1534/genetics.109.103325>.

Ford, Christopher B., Jason M. Funt, Darren Abbey, Luca Issi, Candace Guiducci, Diego A. Martinez, Toni Delorey, et al. 2015. “The Evolution of Drug Resistance in

Clinical Isolates of *Candida Albicans*.” *ELife* 2015 (4): 1–27.  
<https://doi.org/10.7554/eLife.00662>.

Freeman, Kathryn M., and George R. Hoffmann. 2007. “Frequencies of Mutagen-Induced Coincident Mitotic Recombination at Unlinked Loci in *Saccharomyces Cerevisiae*.” *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* 616 (1–2): 119–32. <https://doi.org/10.1016/j.mrfmmm.2006.11.014>.

Gao, Ruli, Alexander Davis, Thomas O. McDonald, Emi Sei, Xiuqing Shi, Yong Wang, Pei Ching Tsai, et al. 2016. “Punctuated Copy Number Evolution and Clonal Stasis in Triple-Negative Breast Cancer.” *Nature Genetics* 48 (10): 1119–30.  
<https://doi.org/10.1038/ng.3641>.

Goldstein, Alan L., and John H. McCusker. 1999. “Three New Dominant Drug Resistance Cassettes for Gene Disruption in *Saccharomyces Cerevisiae*.” *Yeast* 15 (14): 1541–53. [https://doi.org/10.1002/\(SICI\)1097-0061\(199910\)15:14<1541::AID-YEA476>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K).

Golin, J. E., and H. Tampe. 1988. “Coincident Recombination during Mitosis in *Saccharomyces*: Distance-Dependent and -Independent Components.” *Genetics* 119 (3): 541–47.

Grygoryev, Dmytro, Cristian Dan, Stacey Gauny, Bradley Eckelmann, Anna P. Ohlrich, Marissa Connolly, Michael Lasarev, Gianfranco Grossi, Amy Kronenberg, and Mitchell S. Turker. 2014. “Autosomal Mutants of Proton-Exposed Kidney Cells Display Frequent Loss of Heterozygosity on Nonselected Chromosomes.” *Radiation Research* 181 (5): 452–63. <https://doi.org/10.1667/RR13654.1>.

Hall, Brandon M., Chang Xing Ma, Ping Liang, and Keshav K. Singh. 2009. “Fluctuation AnaLysis Calculator: A Web Tool for the Determination of Mutation Rate Using Luria-Delbück Fluctuation Analysis.” *Bioinformatics* 25 (12): 1564–65.  
<https://doi.org/10.1093/bioinformatics/btp253>.

Hirakawa, Matthew P., Diego A. Martinez, Sharadha Sakthikumar, Matthew Z. Anderson, Aaron Berlin, Sharvari Gujja, Qiandong Zeng, et al. 2015. “Genetic and Phenotypic Intra-Species Variation in *Candida Albicans*.” *Genome Research* 25 (3): 413–25. <https://doi.org/10.1101/gr.174623.114>.

Hurst, D. D., and S Fogel. 1964. “Mitotic Recombination and Heteroallelic Repair in *Saccharomyces Cerevisiae*.” *Genetics*, 435–58.

Keeney, Scott, Julian Lange, and Neeman Mohibullah. 2014. “Self-Organization of Meiotic Recombination Initiation: General Principles and Molecular Pathways.” *Annual Review of Genetics* 48: 187–214. <https://doi.org/10.1002/ar.20849.3D>.

Lapunzina, Pablo, and David Monk. 2011. "The Consequences of Uniparental Disomy and Copy Number Neutral Loss-of-Heterozygosity during Human Development and Cancer." *Biology of the Cell* 103 (7): 303–17. <https://doi.org/10.1042/BC20110013>.

Laureau, Raphaëlle, Sophie Loeillet, Francisco Salinas, Anders Bergström, Patricia Legoix-Né, Gianni Liti, and Alain Nicolas. 2016. "Extensive Recombination of a Yeast Diploid Hybrid through Meiotic Reversion." *PLoS Genetics* 12 (2): 1–30. <https://doi.org/10.1371/journal.pgen.1005781>.

Lea, D. E., and C. A. Coulson. 1949. "The Distribution of the Numbers of Mutants in Bacterial Populations." *Journal of Genetics* 49 (3): 264–85. <https://doi.org/10.1007/BF02986080>.

Li, C Y, D W Yandell, and J B Little. 1994. "Elevated Frequency of Microsatellite Mutations in TK6 Human Lymphoblast Clones Selected for Mutations at the Thymidine Kinase Locus." *Mol Cell Biol* 14 (7): 4373–79. <https://doi.org/10.1128/MCB.14.7.4373>.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25 (16): 2078–79. <https://doi.org/10.1093/bioinformatics/btp352>.

Liu, Pengfei, Bo Yuan, Claudia M B Carvalho, Arthur Wuster, Klaudia Walter, Tomasz Gambin, Zechen Chong, et al. 2017. "An Organismal CNV Mutator Phenotype Restricted to Early Human Development." *Cell* 168 (5): 830–42. <https://doi.org/10.1016/j.cell.2017.01.037>.An.

Magwene, Paul M., Ömür Kayıkçı, Joshua A. Granek, Jennifer M. Reininga, Zackary Scholl, and Debra Murray. 2011. "Outcrossing, Mitotic Recombination, and Life-History Trade-Offs Shape Genome Evolution in *Saccharomyces Cerevisiae*." *Proceedings of the National Academy of Sciences* 108 (5): 1987–92. <https://doi.org/10.1073/pnas.1012544108>.

Magwene, Paul M. 2014. "Revisiting Mortimer's Genome Renewal Hypothesis: Heterozygosity, Homothallism, and the Potential for Adaptation in Yeast." *Adv Exp Med Biol* 781: 37–48. <https://doi.org/10.1007/978-94-007-7347-9>.

Mancera, Eugenio, Richard Bourgon, Alessandro Brozzi, Wolfgang Huber, and M Steinmetz. 2008. "High-Resolution Mapping of Meiotic Crossovers and Noncrossovers in Yeast" 454 (7203): 479–85. <https://doi.org/10.1038/nature07135>.High-resolution.

McBride, H J, Y Yu, and D J Stillman. 1999. "Distinct Regions of the Swi5 and Ace2 Transcription Factors Are Required for Specific Gene Activation." *J Biol Chem* 274 (30): 21029–36. <https://doi.org/10.1074/jbc.274.30.21029>.

- McKenna, Aaron, Matthew Hanna, Erin Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernsky, Kiran Garimella, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data." *Genome Research* 20: 254–60. <https://doi.org/10.1101/gr.107524.110.20>.
- McMurray, Michael a, and Daniel E Gottschling. 2003. "An Age-Induced Switch to a Hyper-Recombinational State." *Science (New York, N.Y.)* 301 (5641): 1908–11. <https://doi.org/10.1126/science.1087706>.
- Minet, Michele, Anna-Maria Grossenbacher, and Pierre Thuriaux. 1980. "The Origin of a Centromere Effect on Mitotic Recombination" 60: 53–60.
- Morrison, a, J B Bell, T a Kunkel, and a Sugino. 1991. "Eukaryotic DNA Polymerase Amino Acid Sequence Required for 3'----5' Exonuclease Activity." *Proceedings of the National Academy of Sciences of the United States of America* 88 (November): 9473–77. <https://doi.org/10.1073/pnas.88.21.9473>.
- Pâques, F, and J E Haber. 1999. "Multiple Pathways of Recombination Induced by Double-Strand Breaks in *Saccharomyces Cerevisiae*." *Microbiology and Molecular Biology Reviews : MMBR* 63 (2): 349–404. <https://doi.org/10.1128/MMBR.63.2.349-404.1999>.
- Putnam, Christopher D, and Richard D Kolodner. 2017. *Pathways and Mechanisms That Prevent Genome Instability In Saccharomyces Cerevisiae. Genetics*. Vol. 206. <https://doi.org/10.1534/genetics.112.145805>.
- Putnam, Christopher D, Anjana Srivatsan, Rajul V Nene, Sandra L Martinez, Sarah P Clotfelter, Sara N Bell, Steven B Somach, et al. 2016. "A Genetic Network That Suppresses Genome Rearrangements in *Saccharomyces Cerevisiae* and Contains Defects in Cancers." *Nature Communications*, no. 7. <https://doi.org/10.1038/pj.2016.37>.
- Raj, Arjun, and Alexander van Oudenaarden. 2008. "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences." *Cell* 135 (2): 216–26. <https://doi.org/10.1016/j.cell.2008.09.050>.
- Serero, Alexandre, Claire Jubin, Sophie Loeillet, Patricia Legoix-Né, and Alain G. Nicolas. 2014. "Mutational Landscape of Yeast Mutator Strains." *Proceedings of the National Academy of Sciences* 111 (5): 1897–1902. <https://doi.org/10.1073/pnas.1314423111>.
- Sikorski, R. S., and P. Hieter. 1989. "A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in *Saccharomyces Cerevisiae*." *Genetics* 122 (1): 19–27. <https://doi.org/10.1093/genetics/122.1.19> [pii].

St Charles, Jordan, and Thomas D Petes. 2013. "High-Resolution Mapping of Spontaneous Mitotic Recombination Hotspots on the 1.1 Mb Arm of Yeast Chromosome IV." *PLoS Genetics* 9 (4): e1003434. <https://doi.org/10.1371/journal.pgen.1003434>.

Symington, L. S., R. Rothstein, and M. Lisby. 2014. "Mechanisms and Regulation of Mitotic Recombination in *Saccharomyces Cerevisiae*." *Genetics* 198 (3): 795–835. <https://doi.org/10.1534/genetics.114.166140>.

Tran, H T, J D Keen, M Krickler, M a Resnick, and D a Gordenin. 1997. "Hypermutability of Homonucleotide Runs in Mismatch Repair and DNA Polymerase Proofreading Yeast Mutants." *Molecular and Cellular Biology* 17 (5): 2859–65. <https://doi.org/10.1128/MCB.17.5.2859>.

Veatch, Joshua R., Michael A. McMurray, Zara W. Nelson, and Daniel E. Gottschling. 2009. "Mitochondrial Dysfunction Leads to Nuclear Genome Instability via an Iron-Sulfur Cluster Defect." *Cell* 137 (7): 1247–58. <https://doi.org/10.1016/j.cell.2009.04.014>.

Voth, Warren P, Aileen E Olsen, Mohammed Sbia, Karen H Freedman, and David J Stillman. 2005. "ACE2, CBK1, and BUD4 in Budding and Cell Separation." *Eukaryot. Cell* 4 (6): 1018–28. <https://doi.org/10.1128/EC.4.6.1018>.

Weiss, Eric L. 2012. "Mitotic Exit and Separation of Mother and Daughter Cells." *Genetics* 192 (4): 1165–1202. <https://doi.org/10.1534/genetics.112.145516>.

Zhang, Hengshan, Ane F B Zeidler, Wei Song, Christopher M. Puccia, Ewa Malc, Patricia W. Greenwell, Piotr A. Mieczkowski, Thomas D. Petes, and Juan Lucas Argueso. 2013. "Gene Copy-Number Variation in Haploid and Diploid Strains of the Yeast *Saccharomyces Cerevisiae*." *Genetics* 193 (3): 785–801. <https://doi.org/10.1534/genetics.112.146522>.

## Characterization of phenotypic consequences of heterozygosity in yeast

### Summary

A positive correlation between genomic heterozygosity and phenotype has been described for many species, a phenomenon known as heterosis. In the budding yeast *Saccharomyces cerevisiae*, abundant genomic heterozygosity is frequently found in wild strains isolated from clinical and industrial environments. However, this genomic configuration can be quickly lost in yeast through mitotic recombination, raising the question of how heterozygosity is preserved over time. Here, we hypothesize that natural selection against the negative consequences of loss-of-heterozygosity may contribute to the long-term maintenance of this genome configuration in natural hybrid yeasts. We investigated this possibility in the bioethanol strain PE-2/JAY270, whose genome is characterized by abundant structural and nucleotide polymorphisms between most pairs of homologous chromosomes. This strain is widely adopted in distilleries for its ability to thrive under the harsh biotic and abiotic stresses. To explore how changes in genomic heterozygosity influence some of the desirable traits of PE-2/JAY270, we employed two approaches, inbreeding and uniparental disomy (UPD) to reduce the

---

<sup>2</sup> A version of this chapter will be submitted for publication. Authors that contributed to this work are: Sampaio, NMV; Watson, RA and Argueso, JL.

Contributions to this research are as follows:  
Strain construction: **NMVS**, RAW. Phenotypic tests: **NMVS**, RAW. Flow cytometry: **NMVS**. Whole genome sequencing: **NMVS**, JLA. Data analysis: **NMVS**, RAW, JLA. Manuscript preparation: **NMVS**, RAW, JLA.

abundance and distribution of heterozygous alleles in its genome. The new unique combinations of homozygous alleles in each inbred strain resulted in a wide phenotypic variation under at least two assays, heat stress tolerance and growth kinetics. Genome-wide association analyses allowed the identification of broad genomic regions where genetic polymorphisms potentially impacted those two traits. Interestingly, there was little to no overlap between the loci associated with each trait. Secondly, we employed an approach to induce bidirectional UPD of three different pairs of chromosomes (Chr4, Chr14 or Chr15), while heterozygosity was conserved elsewhere in the genome. In each case UPD was associated to some level of phenotypic alteration. Our results suggest that heterozygosity underlies at least two traits in PE-2/JAY270 - heat tolerance and competitive growth and this relationship may have a significant contribution in supporting the maintenance of PE-2/JAY270's complex genome architecture over time. In addition, the experimental pipeline employed here may contribute for the identification of alleles of interest for industrial and clinical applications in this and other wild yeast strains.

## **Introduction**

One of the positive impacts of genetic diversity is manifested in many species in the form of the heterosis phenomenon, in which highly heterozygous hybrids display increased vigor phenotype in comparison to their homozygous parents (Hochholdinger and Hoecker 2007; Chen 2013a). Accordingly, genetic breeding programs have traditionally taken advantage of heterosis to generate more robust and productive plant crops and livestock (Bittante, Gallo, and Montobbio 1993; Gama et al. 2013;

Schiermiester et al. 2015; Guo et al. 2014; Technow et al. 2014; Huang et al. 2015; Birchler, Yao, and Chudalayandi 2006; J.A. Birchler, H Yao, S. Chudalayandi, D. Vaiman 2010). In contrast, loss of genetic diversity, which results mainly from matings between related individuals, can often lead to negative effects. This phenomenon, known as inbreeding depression, is the conceptual opposite of heterosis and is usually associated with decreased vigor in plants and animals as well as recessive genetic disorders in humans (Charlesworth et al. 2009; Li et al. 2001; Hoffman et al. 2014; McQuillan et al. 2012; Fareed and Afzal 2014).

Three main non-mutually exclusive models have been proposed to explain how widespread genomic heterozygosity may lead to the superior traits seen in hybrids. The dominance model proposes that slightly deleterious recessive alleles present in each of the homozygous parents are complemented in the hybrid by favorable dominant alleles, thus cumulative complementation at multiple loci results in increased performance (Kaepler 2012; Jones 1917; Xiao et al. 1995). The overdominance model suggests the existence of synergistic intra-locus inter-allele interactions that produce positive outcomes in the heterozygous state. Finally, the epistatic model explains heterosis by interactions between alleles of different loci, whose combined effects exceed the individual contributions of each locus (Kaepler 2012; Wolf and Hallauer 1997; Melchinger et al. 2007; Minvielle 1987). The overdominance and epistatic models can accommodate cases where the hybrid performance is greater than the sum of the parents, a possibility not supported by the dominance model (Kaepler 2012; Shull 1911; Krieger, Lippman, and Zamir 2010). In addition, genome-wide effects of polyploidy, gene dosage and epigenetics have also been associated with manifestations

of heterosis (Schnable and Springer 2013; Chen 2013a; Groszmann et al. 2013; Ng, Lu, and Chen 2012).

Most specific examples of the models above, and our general understanding of how genomic heterozygosity influences phenotype, stem from studies conducted on complex eukaryotes, particularly crop plants. Despite steady progress, a refined analysis of the mechanisms underlying heterosis and inbreeding depression has proven to be quite challenging, in part due to the inherent genomic complexity of these subject organisms. It is therefore possible that analogous studies using simpler and genetically tractable model organisms, including yeasts, might offer a powerful alternative to help accelerate the pace of discovery in this field (Fry, Heinsohn, and Mackay 1998; Wang et al. 2015; Chen 2013b; R Shapira et al. 2014; Plech, de Visser, and Korona 2014; Rachel Shapira and David 2016).

In the budding yeast *Saccharomyces cerevisiae*, abundant heterozygosity appears to be prevalent in wild strains isolated, for example, from clinical and industrial settings (Magwene et al. 2011; Peter et al. 2018). One of the first heterozygous wild strains to have its genome characterized was PE-2/JAY270 (referred to here simply as JAY270) (Argueso et al. 2009). This strain was originally isolated as an aggressive wild contaminant of sugarcane-based batch-fed fermentations, but given its desirable growth and yield traits, it was selected for commercial propagation and distribution, and has been widely adopted by bioethanol distilleries as a primary inoculum for nearly two decades (Basso et al. 2008; Della-Bianca et al. 2013). The industrial environment where JAY270 thrives represents an interesting model for studying the dynamics of microbial populations. During each batch of fermentation, cells are exposed to significant and

variable abiotic and biotic stresses, including high osmotic pressure that transitions to ethanol toxicity, oxidative and heat stresses, and steady introduction of wild bacterial and fungal contaminants (Basso et al. 2008; Amorim et al. 2011). In addition, a peculiar feature of this system is that this complex microbial population is recycled twice daily from one batch to the next for up to eight consecutive months during the sugarcane harvest season. The combination of these factors creates a highly competitive environment, in which only the most adapted yeast strains can persist over time. JAY270's defining characteristic is its extraordinary ability to out-compete external contaminants, dominating the microbial population in the distillery and thus ensuring stable and predictable operational conditions (Basso et al. 2008).

While its natural origins are unknown, JAY270 appears to be a wild hybrid that resulted from the mating of two diverged parent haploid strains. It is heterothallic (*i.e.*, its meiotic spores are unable to self-mate), and has a complex diploid genomic architecture, marked by abundant structural and single nucleotide polymorphisms between most pairs of homologous chromosomes (Argueso et al. 2009). Notably, this heterozygous genomic architecture is also a feature of other bioethanol strains (*e.g.*, CAT-1, BG-1) that, like JAY270, were isolated as robust contaminants at sugarcane distilleries (Babrzadeh et al. 2012; Carvalho-Netto et al. 2013; Della-Bianca et al. 2013; Coutouné et al. 2017). The maintenance of such genomic configuration during prolonged clonal propagation is particularly challenging in unicellular organisms like yeast because heterozygosity can be lost irreversibly through allelic mitotic recombination between homologous chromosomes (Symington, Rothstein, and Lisby 2014). These mutational events can lead to loss-of-heterozygosity (LOH) tracts, ranging

from a few base pairs to hundreds of kilobases of homozygous segments for either parental haplotype, which become fixed in clonal descendants.

Our group recently mapped the distribution of heterozygous loci and also measured the specific rates of LOH at four regions of the JAY270 genome (Fig. 1.1; and (Rodrigues-Prause, Sampaio et al. 2018)). Notably, we found that dozens of LOH tracts are scattered throughout the genome, indicating the occurrence of mitotic LOH in JAY270's clonal ancestors prior to isolation. We also determined that LOH occurs in this strain at a rate of approximately 1 event every 2000 cell divisions ( $5 \times 10^{-4}$ ). While this rate is very high compared to other mitotic mutational mechanisms (*i.e.* nucleotide substitutions  $\sim 10^{-8}$ ), it is nonetheless similar to the LOH rates measured in conventional laboratory *S. cerevisiae* strains. Therefore, the JAY270 genome is not inherently unstable in relative terms. Given that LOH events occur at a high rate in yeast genomes and that JAY270 had been clonally propagated at industrial scale for more than 10 years prior to isolation (and perhaps longer in wild environments), we reasoned that most of its genomic heterozygosity should have already been eroded away. The fact that a substantial portion ( $\sim 60\%$ ) of the JAY270 genome still contains heteroalleles suggests that selective forces might have acted to disfavor cells carrying LOH spanning important loci. In contrast, loss of heteroalleles that do not confer a substantial adaptive advantage may have been tolerated relatively well.

If the scenario outlined above is correct, then the present distribution of heterozygosity observed in JAY270 should correspond to the genomic regions where beneficial heteroalleles reside. This possibility brings about several intriguing questions which we set out to explore in this study: Why is it that the JAY270 genome is still fairly

heterozygous when LOH had virtually unlimited opportunity to accumulate (high rate plus industrial scale propagation)? Are the regions that remain heterozygous important for JAY270's desirable industrial traits? If so, what are the phenotypic consequences of reducing the overall level of heterozygosity and altering the distribution of heteroalleles in the JAY270 genome? To begin to answer these questions, we employed two different experimental approaches to reduce heterozygosity in the JAY270 genome. First, we used controlled inbreeding to generate a collection of experimental strains, each harboring a unique combination of homozygous alleles distributed genome-wide. In addition, we generated strains in which bidirectional LOH was confined to three different single chromosomes, while preserving heterozygosity elsewhere. We compared the phenotypes of those strains to their fully heterozygous parent (JAY270) under different culture conditions and identified broad genomic regions where genetic polymorphisms significantly impacted competitive growth and heat stress tolerance. Our results suggest that JAY270's heterozygosity may have been maintained over time by the action of selective forces, and provide a new perspective on how the interplay between heterozygosity, phenotype, and mitotic recombination may have contributed to shape the genomic architecture of this and other wild hybrid *S. cerevisiae* strains.

## **Results**

### **Controlled reduction of heterozygosity in the JAY270 genome through inbreeding**

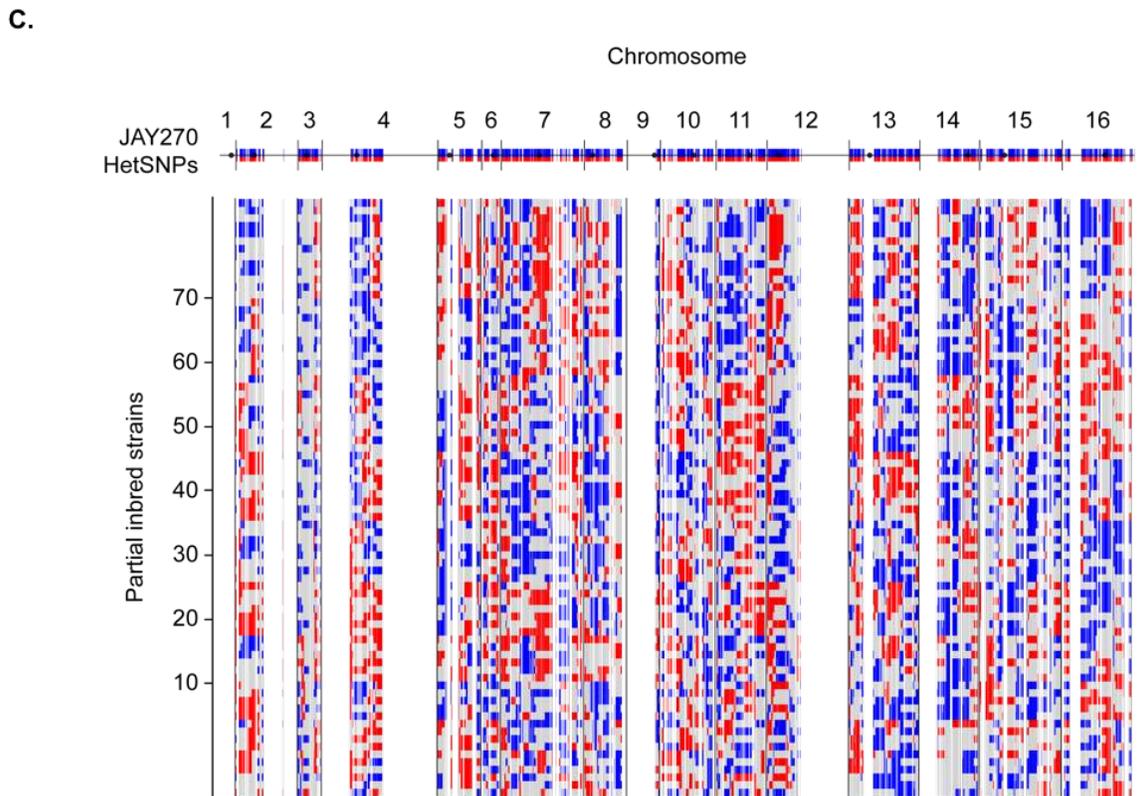
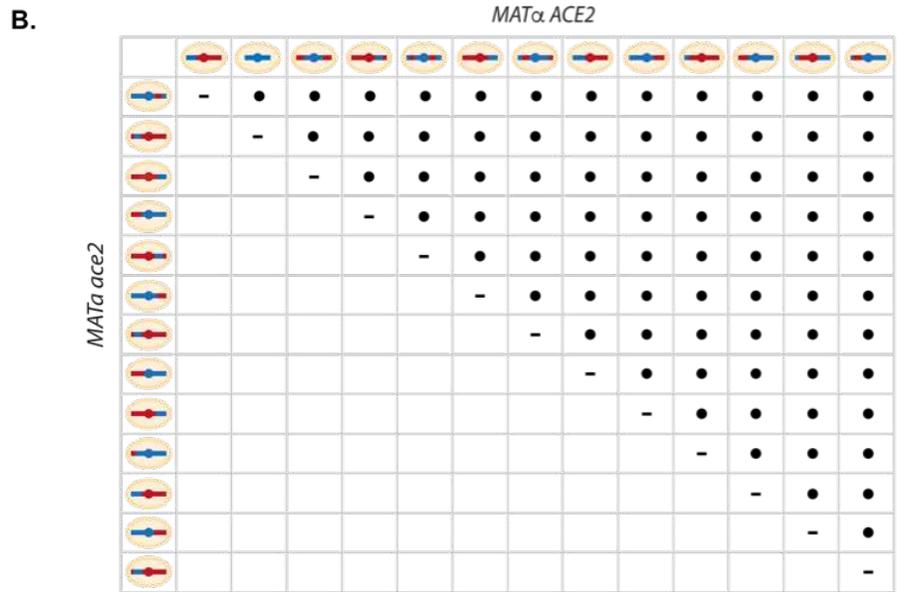
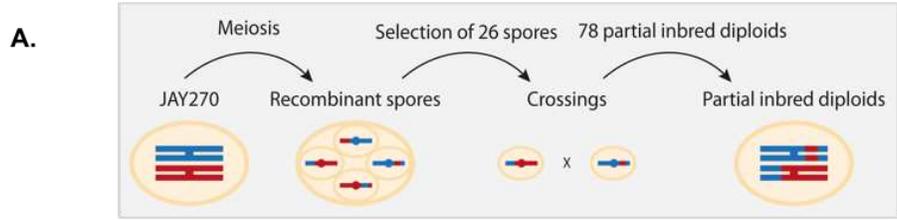
In order to investigate the interplay between JAY270 genomic heterozygosity and its phenotypes, we explored how changes in the abundance and distribution of

heterozygous alleles would affect the traits of the strain. Recently, we reported a draft phased map of ~12,000 heterozygous single nucleotide polymorphisms (HetSNPs) unevenly distributed across JAY270's genome (Fig. 1.1 and (Rodrigues-Prause, Sampaio et al. 2018)). The phased haplotype information available for each chromosome pair was arbitrarily classified as either maternal (M; red in all figures) or paternal (P; blue in all figures) origin.

Our primary strategy to create JAY270 derivatives containing reduced heterozygosity was based on inbreeding. Our group has previously isolated and whole-genome sequenced 52 haploid spore clones originated from thirteen sets of JAY270 four-spored tetrads (Rodrigues-Prause, Sampaio et al. 2018). It has been estimated that each meiotic cell division in *S. cerevisiae* produces about 90 crossovers distributed across the whole genome (Mancera et al. 2008; Chakraborty et al. 2018). These events result in the formation of recombinant chromatids that are sorted into haploid spores, each containing approximately half maternal and half paternal alleles (Fig. 3.1A). In order to maximize the genotypic variation of the haploids used in our crossings, we selected for mating only one *MATa* and one *MAT $\alpha$*  spores from each of the thirteen sequenced tetrad sets. This ensured that all inbred diploids were formed by joining recombinant haplotypes generated from independent meiotic crossover events. An additional criterion for selection of the parent spores was based on their genotype at the *ACE2* locus. We recently showed that JAY270 is heterozygous for a frameshift mutation at *ACE2* and diploid derivatives homozygous for the mutant allele display a cell-cell aggregation phenotype that could impair the analysis of inbred diploids (Rodrigues-Prause, Sampaio et al. 2018). Thus, 13 *MAT $\alpha$  ACE2* and 13 *MATa ace2-A7* spores

were crossed in inter-tetrad pairwise combinations as detailed in Table S3.1, resulting in a collection of 78 inbred diploid strains directly derived from JAY270 (Fig. 3.1B).

Because only one generation of inbreeding was carried out, the genome of each inbred strain in the collection was predicted to be a quarter homozygous for paternal alleles, a quarter homozygous for maternal alleles, and half heterozygous. Importantly, since each haploid parent inherited a unique combination of maternal and paternal alleles, no two inbred diploids were heterozygous for the same half HetSNPs. Based on the whole-genome sequence information of all parental haploids, we derived precise genotype maps for each inbred diploid. These maps list all loci that remained heterozygous (M/P), and the loci that became homozygous for either allele (M/M or P/P), and reveal the genetic variation present in our collection (Fig. 3.1C).



**Figure 0.1. Panel of inbred diploids derived from JAY270.**

A) Extensive polymorphisms in JAY270's genome are represented by a pair of homologous chromosomes colored in blue (paternal haplotype or P) or red (maternal haplotype or M). JAY270 cells were induced to undergo meiosis in sporulation media, resulting in the formation of tetrads containing four recombinant haploid spores. 52 recombinant haploids isolated from thirteen four-spored tetrads were dissected and had their whole genomes sequenced. One spore from each mating type (*MATa* or *MATα*) was selected from each tetrad and used for pairwise crossings. This setup resulted in a collection of 78 inbred diploid strains derived from inter-tetrad matings. B) Chart showing a schematic representation of the matings between the recombinant haploids. Inter-Tetrad matings (•) resulted in diploids with approximately 50% of the original heteroalleles present in the parental genome, 25% of homozygous paternal alleles and 25% of homozygous maternal alleles. Intra-Tetrad crossings (-) resulted in diploids with 0% to 100% heterozygosity and were excluded from downstream analysis. C) Genome-wide representation of the genotypes of the 78 inbred diploids. The top blue and red line corresponds to the linear representation of the distribution of 12,023 HetSNPs in the JAY270 genome (Fig. S2.1). Each line below shows the genotype of one inbred strain at each one of the HetSNPs markers. Gray = heterozygous M/P. Blue = homozygous P/P. Red = homozygous M/M. Main results in this figure were generated by: NMVS.

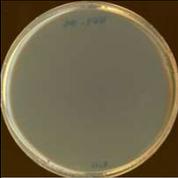
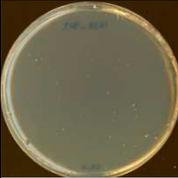
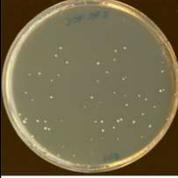
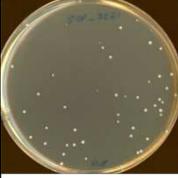
We also analyzed the genotype maps to determine the overall level of hetero- and homozygosity in each of the inbred diploids (Fig. S3.1A). The average inbred was heterozygous M/P for 51% for the JAY270 HetSNPs, within a range of ~40% to ~62% for the least and most heterozygous inbreds (Fig. S3.1B). The average of M/M and P/P homozygosity was well balanced (~26% and ~22%, respectively) and consistent with the levels predicted for a single generation of inbreeding.

**Characterization of phenotypic variation in the inbred diploid collection**

We next explored how the reduced heterozygosity in each inbred diploid affected different traits in comparison to their fully heterozygous parent (JAY270). A plate spotting assay format was used as an initial screen for growth phenotypes under a variety of individual stress conditions, some of which are known to be present in the sugarcane fermentation industrial environment (detailed information available in Table

S3.3). No significant changes in cell viability or growth characteristics were observed when cells were plated and grown in the presence of 7% or 11% v/v ethanol, 30 mM furfural (a byproduct of lignocellulose biomass fermentation), 0.75 mM of menadione (an inducer of oxidative stress), or 100 and 150 J/m<sup>2</sup> ultraviolet light exposure and 0.01% methyl methanesulfonate (DNA damage). The abilities to metabolize galactose and the non-fermentable carbon sources ethanol and glycerol were also uniform across all inbred strains. Mild phenotypic variation was observed when cells were grown on raffinose as the sole carbon source, or in the presence of 100 mM of hydroxyurea, an inducer of DNA replication stress (data not shown). Finally, a pronounced variation in tolerance to high temperature stress (39°C) was observed for many strains in the inbred collection (Fig. S3.2).

The wide range in the distribution of this phenotype during the screening phase made it suitable for a subsequent detailed phenotypic characterization and genotype association analysis. We categorized the inbred strains into five phenotypic groups using a colony-size scoring system (Fig. 3.2). JAY270 displayed an intermediate heat tolerance phenotype (score 3), characterized by a good viability, but substantial variation in colony diameter, ranging from small to medium sized colonies. Most inbred strains (36 of 78) displayed a similar phenotype. The remaining inbred strains displayed heat tolerance patterns that were either lower or higher than JAY270. At the extremes were nine that formed either micro-colonies or showed no growth at all when incubated at 39°C (score 1), and fifteen that formed uniformly large colonies and were classified as the most heat tolerant strains (score 5; Fig. 3.2).

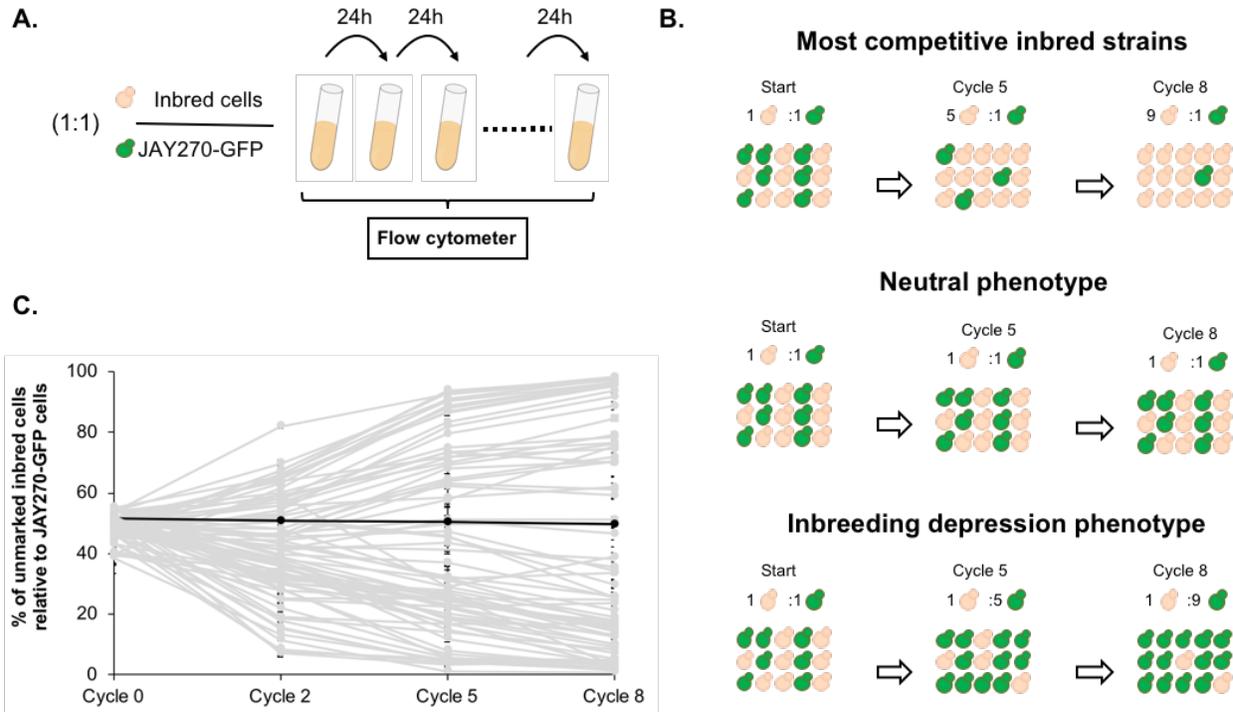
Phenotype classification	Representative growth at 39°C	Representative growth at 30°C
<b>Score 1</b> (JAY1664)  Completely clear plates or containing microscopic colonies.		
<b>Score 2</b> (JAY1628)  Small, yet easily visible, colonies.		
<b>Score 3</b> (JAY270)  Heterogeneous colony size, ranging from medium to small sizes.		
<b>Score 4</b> (JAY1658)  Medium to large colonies, with most having medium size. Narrow variation in colony size.		
<b>Score 5</b> (JAY1610)  Large colonies. Homogeneous colony size.		

**Figure 0.2. Inbred strains show heterogeneous levels of tolerance to high temperature stress.**

A scoring system was used to assign different levels of tolerance to each inbred strain. The criteria for each score used and the strains chosen as representative examples for each phenotypic category are described in column 1. Representative growth of strains assigned to different scores after incubation at 39°C and 30°C is depicted, respectively, on the second and third columns of the image. Main results in this figure were generated by: NMVS.

In addition to the discrete phenotypes examined through the plating assays above, we also investigated a more subtle and continuous variation in mitotic growth kinetics. JAY270 is known to grow very robustly, and that is likely a key factor contributing to its ability to outcompete most wild yeast contaminants in the sugarcane

fermentation process. Thus, we also sought to explore the effects of heterozygosity on JAY270's growth vigor phenotype. To do so, we quantified relative growth kinetics through a cumulative co-culture competition assay. Each inbred strain was co-cultured with a GFP-marked JAY270 derivative (JAY270-GFP) under ideal yeast growth conditions (YPD liquid rich medium at 30°C). The co-cultures were started with an approximately equal inoculum of the two competitors ( $\sim 2.5 \times 10^6$  cells each) and were incubated for 24 hours, a little past nutrient depletion and population saturation ( $\sim 15$  hours). At the end of each daily cycle, 1% volume of each co-culture was transferred to fresh liquid medium to allow continued growth. The ratio of GFP-negative (inbreds) to GFP-positive (JAY270) present in the co-cultures was measured periodically with a flow cytometer, and used as a parameter to estimate the growth kinetics of each of the inbred diploids relative to JAY270. Inbred strains with intact growth kinetics should have steady  $\sim 1:1$  ratios over time, whereas ratio deviations up or down would indicate a phenotypic alteration (Fig. 3.3A-B).



**Figure 0.3. Phenotypic assessment through growth competition assays.**

A) Co-cultures were started with an approximate 1:1 ratio between a GFP-marked JAY270 derivative and each inbred strain competitor. Every 24 hours (one cycle of competition), 1% volume of each culture was transferred to fresh media through 8 consecutive days of co-cultures. The ratio of inbred to JAY270-GFP cells was monitored periodically using a flow cytometer. B) Schematic representation of inbred strains displaying a positive, neutral or negative growth kinetics profile in comparison to the parent JAY270. C) Competitive growth profiles of the 78 partial inbred diploids. Each line shows the growth profile of an individual inbred strain (average of three replicates) in co-culture with JAY270-GFP. The percentage of inbred cells in the culture at each cycle analyzed is shown in the y-axis. Black line represents a control competition between JAY270 and JAY270-GFP. Main results in this figure were generated by: NMVS.

Besides the genotype of the inbreds, another factor that may cause the GFP ratio to deviate is the emergence of beneficial *de novo* mutations within the culture. However, this effect should be delayed until the newly formed mutants become numerous enough to be detected. In order to determine the period of time during which the GFP ratio can be confidently attributed solely to the initial genotype of the inbreds, we performed

control co-culture competitions of each of four independently derived GFP-marked JAY270 clones versus the original unmarked JAY270 strain. We carried out a total of twelve co-cultures (four GFP-marked clones, three replicates each) with daily 1% volume transfer cycles to fresh media for 22 consecutive days, and the GFP ratio was measured at 7 day intervals. The GFP ratios in all 12 independent co-cultures remained steady at ~1:1 by the end of the first week (cycle 8). By the end of the second and third weeks, some of the ratios had diverged up or down, presumably through emergence of mutations in the GFP- or GFP+ strains (Fig. S3.3D). Therefore, we limited our experimental competitions of the inbred diploids versus JAY270-GFP to a maximum of 8 daily transfer cycles, in order to insulate the measured GFP ratios from the effect of *de novo* mutations. These control experiments also showed that integration of the GFP cassette into the JAY270 genome did not by itself have an effect on growth kinetics. Additional JAY270-GFP versus JAY270 control co-culture competitions were included every time a new experimental evaluation of the inbred collection was performed (39 replicates), and in no cases a significant deviation in the GFP ratio was observed before transfer cycle 8.

The results of the competitive growth profiles of the inbred collection were characterized collectively by a “fan out” shape, highlighting a wide range phenotypic variation between inbred strains (Fig. 3.3C). Approximately 20% of the inbred diploids displayed growth kinetics that were substantially different to JAY270, not only slower but also faster in many cases. Of this group at the extremes of the competitiveness range, 16 displayed a strong reduction in growth kinetics and comprised less than 10% of the total cell population by the last cycle of co-culture; while 13 inbreds showed a

substantial improvement in growth, outcompeting JAY270 to reach more than 90% of the cells in the co-cultures. Importantly, all inbreds, regardless of the neutral, positive or negative relative growth kinetics profiles, followed a steady unidirectional trend from the early cycles until the end of co-culture. This result is consistent with their phenotype being a function of their initial genotype, and not due to the random appearance of *de novo* mutations during the experiment. In addition, there was very little variation between the independent replicates of each inbred co-culture, further disfavoring a potential role for *de novo* mutations on the observed phenotypes.

It is important to note that all 78 inbreds, even those with the poorest performance in the co-culture competition, grew apparently normally and indistinguishably from JAY270 on solid/agar rich medium at 30°C. This shows that the cumulative liquid co-culture competition assay was able to reliably and consistently uncover very subtle relative differences in growth kinetics that would most likely be missed by conventional methods, such as optical density measurements from exponentially growing pure cultures. Specifically, we estimate that the most extreme competition phenotypes among the inbreds, reaching <10% or >90% of the total co-culture cell population by transfer cycle 8, should have a rate of cell division only ~3% longer or shorter than JAY270, respectively. Most inbreds that had intermediate phenotypes, should have changes in growth rate of only ~1% or less. Thus, the co-culture competition assay offered an opportunity to reliably measure even minor phenotypic changes that were determined by the different genotype combinations represented in the inbred collection.

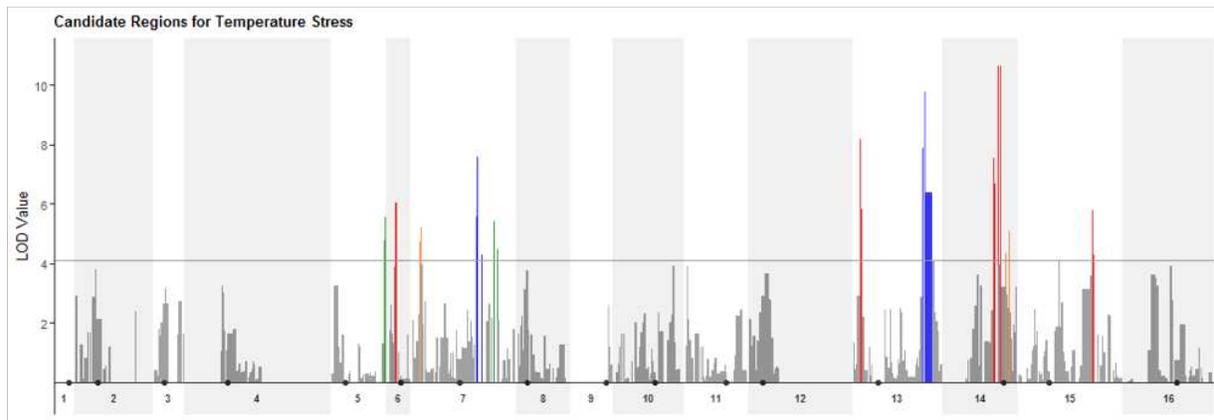
Another noteworthy aspect of this analysis was that for both phenotypes measured (heat stress tolerance and relative growth kinetics), we identified multiple inbred strains that actually outperformed their fully heterozygous parent. This result may appear to be counter-intuitive at first glance, since inbreds are known to generally display negative phenotypic consequences. However, we believe our result is not unexpected, in light of the fact that the environment in these laboratory assays was quite different from the one where the JAY270 strain was isolated from. If it had been possible for us to exactly reproduce the biotic and abiotic challenges found in sugarcane bioethanol distilleries, then we would expect most inbreds in the collection to perform poorly, and some to be comparable to JAY270, but never superior.

We asked whether the superior phenotypes in the two specific assays were somehow related, perhaps as result of a general improvement in the vitality of the inbred strains. Instead, we found that there was no correlation between the inbreds that ranked at the top and the bottom of the distributions for each phenotype (Fig. S3.4). The most competitive strains had as much overlap with the most and the least heat tolerant strains, and vice versa. This result suggests that even though the heteroalleles present in the JAY270 genome have the potential to be re-assorted to improve performance for one trait, this will often result in a decrease in performance for a different trait. Therefore, the heterozygous genomic configuration present in JAY270 may actually represent a finely-tuned compromise that enables it to respond well to a variety of adverse conditions, resulting in its high overall fitness in the complex and dynamic bioethanol industrial environment.

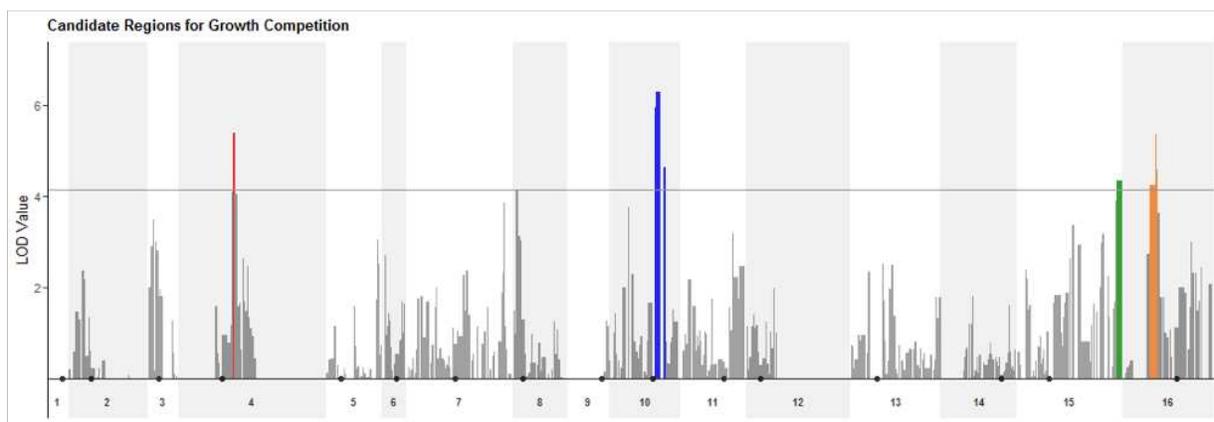
## Identification of genomic regions associated with phenotypic variation

We next performed a genome-wide association analysis to identify possible causal relationships between the specific genotypes at JAY270 HetSNPs and the phenotypic variation in heat stress tolerance and growth kinetics characterized among inbred strains. We used the genotype maps of all inbreds (Fig. 3.1C) to determine the frequencies of homozygous M/M and P/P, and heterozygous M/P genotypes for all HetSNP markers. Because the inbred diploids in the collection were necessarily heterozygous at the Chr3 *MAT* locus (*MATa/MAT $\alpha$* ), and were designed to also be heterozygous at the Chr12 *ACE2* locus (*ACE2/ace2-A7*), we excluded from the association analyses all markers located within ~50kb and ~75kb upstream and downstream of these loci, respectively. This resulted in a list of 11,774 HetSNPs that were included in the analyses. Then, for each marker we calculated the average phenotype observed among strains homozygous M/M, homozygous P/P and heterozygous. The statistical significance threshold for the identification of putative loci associated to each trait was established by randomized phenotype by genotype permutation tests (five independent runs of 10,000 iterations for each trait) at the  $p < 0.05$  significance level. For heat tolerance, loci with LOD > 4.11 were considered to have a significant association (Fig. 3.4A), and for competitive growth the significant threshold was LOD > 4.13 (Fig. 3.4B).

A.



B.



■ Additive Maternal ■ Additive Paternal ■ Dominant Maternal ■ Dominant Paternal ● Centromere

**Figure 0.4. Genome-wide association analyses to identify potential loci underlying heat stress tolerance.**

(A) and relative growth kinetics (B). Genome-wide association plots of LOD scores for each HetSNPs. For heat tolerance, loci with  $LOD > 4.11$  were considered to have a significant association, and for competitive growth the significant threshold was  $LOD > 4.13$ . The color code depicted at the bottom of the figure indicates the quantitative inheritance model that best fit the loci showing significant association. Main results in this figure were generated by: RAW and NMVS.

Although our inbred diploid population size was limited, this analysis was sufficient to reveal some genomic segments that may have an important contribution to the traits analyzed. Additional regions that make modest contributions to these traits

likely also exist, but their reliable detection would require a substantial expansion of the size of the inbred collection, along with a higher throughput phenotyping platform. In total, thirteen regions from six different chromosomes showed association to heat tolerance (Chr5, Chr6, Chr7, Chr13, Chr14, Chr15), and four regions from four different chromosomes to competitive growth (Chr4, Chr10, Chr15 and Chr16). Importantly, none of the significant association regions overlapped between the two phenotypes. This suggested that loci present at different genomic regions are important for superior phenotypes in these two different traits, consistently with the observation that there was very little overlap between the inbred strains ranked in the upper or lower tiers of two phenotypes (Fig. S3.4).

For each region showing significant association, we then evaluated which quantitative inheritance model better fit the traits (Fig. 3.4A-B, Table S3.6). Most regions were consistent with an additive variance model in which the heterozygote has an intermediate phenotype. We also found regions with likely dominance, but cases of overdominance were not detected. A two-dimensional scan of the genome was also performed, but no significant pairwise epistatic interactions were detected (data not shown).

### **Phenotypic consequences of chromosome-scale LOH**

The analysis of inbred diploids described above allowed us to investigate the possible relationship between genome-wide heterozygosity and the phenotypes of JAY270. Overall hybrid vigor is likely the result of complex interactions between multiple genes. In the inbred collection approach we removed roughly half of the heterozygosity

present in the JAY270, thus a large fraction of the genome was affected in each strain. We next decided to take an independent and more conservative approach in which fewer heterozygous loci could be removed at a time, and asked if phenotypic consequences would still be observed.

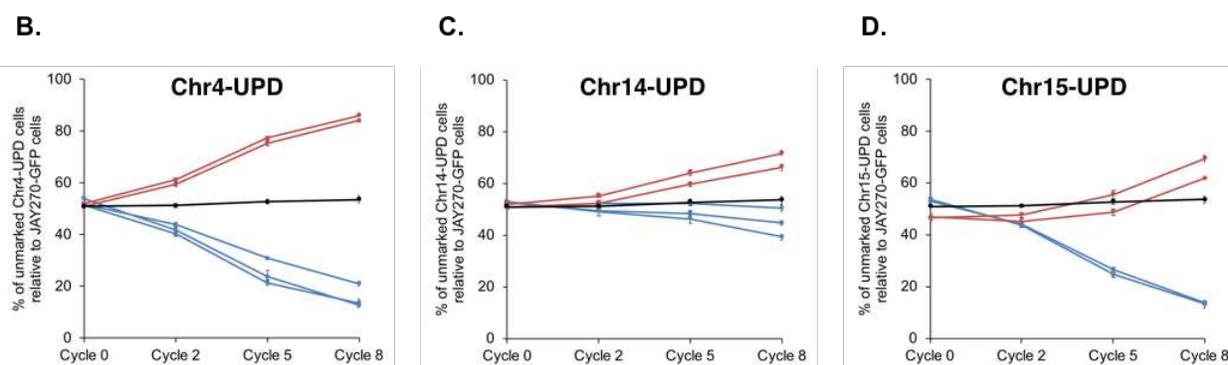
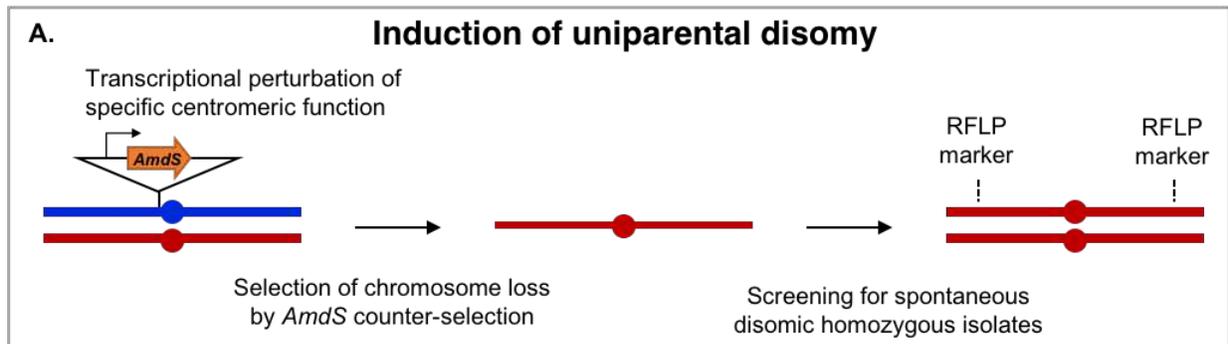
To do so, we conducted a proof-of-concept experiment in which we adapted a procedure to induce uniparental disomy (UPD); *i.e.*, homozygosis for individual whole chromosomes), while fully preserving heterozygosis in the remaining 15 chromosome pairs. Our strategy took advantage of previous demonstrations that driving transcription through centromeric regions leads to perturbation of the function of centromeres, and can be used to induce targeted chromosome loss, resulting in  $2n - 1$  monosomic diploid cells (Reid et al. 2008; Hill and Bloom 1987). Reid et al (2008) applied this strategy to map mutations to individual chromosomes in a *ura3* auxotrophic diploid strain background by inducing transcription of a *pGAL1-URA3* cassette integrated at near centromeric regions, and then applying selection for 5-FOA resistance to recover clones that had lost the targeted chromosome. A key part of the strategy is that monosomic diploid *S. cerevisiae* strains tend to rapidly endoduplicate the remaining homolog, which results in reestablishment of the normal  $2n$  DNA through UPD. Here, we adapted this approach for use in diploid prototrophic strains by integrating into JAY270 a hemizygous copy of the heterologous forward and counter selectable marker *AmdS* (Solis-Escalante et al. 2013) immediately upstream of the centromeric regions of interest. We modified the *AmdS* cassette by removing the transcriptional terminator sequence, thus leading to constitutive transcription through the centromere. Insertion of the *AmdS* cassette adjacent to the M or P centromere was obtained and stably maintained through forward

selection for growth in media containing acetamide as the only nitrogen source. Then, fluoroacetamide resistance selection for loss of the cassette was used to isolate clones carrying chromosome loss followed by UPD (Fig. 3.5A). This approach allowed us to evaluate the phenotypic consequences of inducing homozygosity for each of the two haplotypes in one chromosome at a time. In addition, the use of the *AmdS* marker allowed us to do this work in prototrophic JAY270 cells, eliminating potential confounding effects that the introduction of uracil auxotrophy could have brought.

We focused the bidirectional UPD analysis on three pairs of chromosomes (Chr4, Chr14 and Chr15) chosen on the basis of their overall chromosome size, distribution and number of heterozygous loci, and genetic association to growth kinetics (Fig. 3.4B) detected with the inbred collection. Loss of long *S. cerevisiae* chromosomes typically results in a heavier phenotypic burden than loss of a small chromosome, thus the ability to recover UPD through endoduplication is higher. Chr4 was particularly attractive because it is a large chromosome and shows evidence of pre-existing LOH events in the terminal regions of both arms (Fig. 3.1C and Fig. S2.1). This configuration suggests that cells that underwent LOH spanning the relatively short central region (~ 418,440 Kb; ~ 600 HetSNPs) might have had a fitness disadvantage in comparison to cells that remained heterozygous. We also chose to study Chr15, because it is a large chromosome with over 1,400 HetSNPs distributed throughout its whole length and we detected two different segments that showed association to competitive growth (Fig. 3.4B). Finally, a third interesting case-study was Chr14, which is a mid-size chromosome, but still contains a relatively large number of HetSNPs (~700), yet none of them showed statistical significance in the inbred association analysis for competitive

growth (Fig. 3.4B). Accordingly, we interpreted that the HetSNPs in this chromosome may have minor or no effect on competitive growth individually, but making them all simultaneously homozygous might have a detectable phenotype.

Co-culture growth competition assays between Chr4-UPD M/M and Chr4-UPD P/P versus JAY270, supported a strong association of heteroalleles in this chromosome with phenotype (Fig. 3.5B). Homozygosis for the two Chr4 haplotypes influenced the growth kinetics significantly and in opposite directions. Chr4-UPD M/M strains showed a relatively faster growth, whereas Chr4-UPD P/P strains showed the opposite phenotype, consistently with the additive maternal inheritance of the HetSNPs in this chromosome identified through by the genome association analysis (Fig. 3.4B). Homozygosis for the two Chr15 haplotypes resulted in a less symmetrical change in growth kinetics, but still following a general trend predicted by our genome association analysis. Chr15-UPD M/M strains displayed a subtle but steady growth advantage, while Chr15-UPD P/P strains were outcompeted by the parent strain JAY270 at a faster pace (Fig. 3.5D). Finally, the changes in growth competition profiles of the Chr14-UPD strains were more subtle but reproducible, which corroborated the observations from our association analysis that did not point to a strong contribution of Chr14 HetSNPs for the growth phenotype (Fig. 3.4B and Fig. 3.5C). Taken together, even though each of the three UDP pairs the strains retained 90-95% of the overall HetSNPs of JAY270, these relatively small and localized erosions of heterozygosity were more than sufficient to create reproducible phenotypic alterations. This suggests that mitotic recombination events leading to LOH of comparable magnitude would most likely have phenotypic outcomes that could be acted upon by selective forces.



**Figure 0.5. Construction and relative growth kinetics profiles of Chr-UPD strain pairs.**

A) A cassette containing the counter-selectable marker *AmdS* under the transcriptional control of the *TEF1* constitutive promoter and lacking a terminator sequence was integrated immediately upstream to the centromeric regions of each homolog of Chr4, Chr14 and Chr15. The transcription of this gene perturbs centromere function and disrupts chromosome segregation during mitosis. Cells that have lost the *AmdS* marker were selected for in media containing fluoroacetamide. Spontaneous endoduplication of the remaining homolog results in strains containing UDP, in this case represented as the maternal (red) homolog. Confirmation of the loss of one haplotype was obtained by RFLP analysis at both chromosome arms and confirmation of endoduplication was obtained by tetrad dissection and spore viability analysis. B-D) Growth kinetics profiles of Chr-UPD strain pairs. Each line shows the relative growth profile of an individual Chr-UPD strain in co-culture with JAY270-GFP. The percentage of Chr-UPD cells relative to JAY270-GFP cells at the 0, 2, 5 and 8<sup>th</sup> cycle of culture is shown in the y-axis. Red and blue lines represent strains containing two copies of the maternal (M/M) or paternal (P/P) haplotypes, respectively, of each chromosome pair analyzed (Chr4, Chr14 and Chr15). Black line represents a control competition between JAY270 wild type and JAY270 labeled with GFP. Main results in this figure were generated by: NMVS and RAW.

## Discussion

### Heterozygosity and fitness in yeast hybrids

The genetic basis of heterosis in yeast has been investigated before in a few studies. For example, Plech *et al.* created a collection of 253 hybrid *S. cerevisiae* strains by self- and inter-mating haploid derivatives of isolates from different geographical and ecological settings. They compared the maximum growth rate under different growth conditions between self-diploid and inter-mating hybrids, they found that heterosis was prevalent, primarily in the hybrids derived from matings between domesticated yeast strains (*e.g.* clinical and industrial). Heterosis appeared to be less pronounced among hybrids generated from matings between haploids isolated from natural habitats (Plech, de Visser, and Korona 2014). Another study quantified heterosis in a similar collection of 120 *S. cerevisiae* artificial hybrids tested in five growth conditions. A combination of dominance, overdominance and epistatic models was found to drive heterosis in the strains analyzed (R Shapira *et al.* 2014). While these studies provided important insights about the influence that heterozygosity has on yeast fitness, the hybrid strains analyzed were artificially created by mating highly diverged haploid backgrounds isolated from different environments. As a result, the diploid strains formed through these extreme outcrossings are highly heterozygous (>60,000 SNPs), however, they are not actually observed outside of the laboratory setting, therefore their allelic combinations were never subjected to natural selection forces.

Here, we investigated how heterozygosity contributes to the traits of the natural hybrid yeast strain JAY270, a wild isolate initially identified as a highly adapted contaminant of bioethanol industrial fermentations. Another important distinction of our

approach is that the genetic variation in our strain panel was obtained through reduction (rather than creation) of heterozygosity through inbreeding and UPD. The choice of parent strain and genetic manipulation approaches allowed us to examine heteroalleles that were vulnerable to erosion through mitotic recombination (LOH) and prolonged propagation in ethanol distilleries, yet were maintained in the JAY270 genome, presumably due to the detrimental phenotypic consequences associated with their loss. Using a similar approach, Granek *et al.* successfully mapped 13 genes linked to biofilm formation by coupling inbred matings between spores derived from a highly heterozygous *S. cerevisiae* clinical isolate and bulk segregant analysis (Granek et al. 2013). These combined results show that exploiting the cryptic genetic and phenotypic variation of naturally heterozygous strains can be a powerful tool for functional studies, including the genetic basis of traits important to fungal pathogenicity or biotechnology applications.

### **Cryptic phenotypic variation in JAY270**

Our results show that the JAY270 genome harbors alleles that can support a wide phenotypic plasticity for competitive growth and heat stress tolerance (and possibly additional traits) (Fig. 3.3C, Fig. 3.2). This plasticity may be accessible to JAY270 through meiotic recombination followed by inbreeding or mitotic recombination events (Magwene 2014), and the environmental selection pressures to which the cells are exposed to will determine the success or failure of the new unmasked allelic combinations. The adaptive contributions of these genomic changes have been nicely characterized in inter- and intra- species yeast hybrids grown in chemostats over

several generations under different growth conditions (Smukowski Heil et al. 2017). Mitotic recombination leading to LOH was shown to be a major driver of adaptation in those hybrids. Interestingly, when clones carrying an LOH event that conferred superior fitness in a specific growth condition were tested in an alternate condition, their fitness was reduced. This is consistent with our findings that reduction in heterozygosity, throughout the genome as a result of inbreeding or confined to individual chromosomes, can have very distinct effects for different traits.

Our genome association analyses pointed to broad genomic regions scattered throughout the genome that were associated with heat stress tolerance and relative growth kinetics, consistently with the quantitative nature of these traits and the cumulative effects of inter-allelic interactions. Additionally, we observed a prevalence of dominance inheritance in most regions associated with phenotype in both of our assays, which was also previously found to be the prevalent mode of inheritance in artificial hybrids displaying a heterotic phenotype (R Shapira et al. 2014). Interestingly, we also found that inbreeding was just as often associated with superior as it was with inferior performance in the two traits analyzed. While the existence of inbred strains that exceed the performance of their fully heterozygous parent may seem counter-intuitive, it is important to note that the phenotypic assays we used did not reproduce the conditions that shaped the JAY270 genetic background. The challenges posed by the complex and dynamic sugarcane fermentation environment might be better met by a heterozygous genomic configuration that allows JAY270 to be a well-rounded generalist. Our observations were consistent with a model in which the erosion of heterozygosity in the JAY270 genome is curtailed by natural selection forces. While homozygosity to a

specific chromosomal region can lead to faster growth, it may also decrease tolerance to elevated temperatures. We speculate that the net result of this genomic change would be disadvantageous in JAY270's natural habitat, which would explain the persistence of cells with a heterozygous genotype. However, in cases where the beneficial effect of homozygosity can be narrowed down to specific loci, it may be possible to use targeted allele engineering approaches for strain improvement. This would minimize the negative effects of losing heterozygosity in the neighboring loci.

### **UPD strategy for phenotypic testing of confined LOH**

The phenotypic analysis of JAY270-derived strain pairs carrying bidirectional UPD engineered for specific chromosomes provided important clues about the extent to which heterozygosity influences phenotype in this wild strain. Chr4-UPD M/M displayed improved growth kinetics in comparison to the parent JAY270, while the Chr4-UPD P/P clones showed signs of inbreeding depression (Fig. 3.5B), in accordance to the additive maternal inheritance found for the significant HetSNPs in this chromosome. It is entirely possible, and likely, that other phenotypes could have diverged in opposite directions between M/M and P/P Chr4-UPD. This observation supports the hypothesis that the ancestral LOH events spanning the terminal segments of this chromosome likely did not impact or even provide a growth advantage to cells in the industrial fermentation environment. In contrast, cells that underwent LOH events spanning the central HetSNPs in Chr4 may have faced a selective disadvantage, falling behind cells that remained heterozygous at those segments. A similar profile was observed in the Chr15-UPD strains, highlighting again that even localized reduction of heterozygosity levels

can significantly impact the JAY270's phenotype (Fig. 3.5D). Importantly, these results also support the efficacy of our approach in facilitating the discovery of genomic segments associated with traits of interest, which was also corroborated by the phenotypic profiles of Chr14-UPD strains. Chr14 harbors more than 700 HetSNPs, however, none of these sites showed a significant association to competitive growth in our analysis (Fig. 3.4B). Accordingly, we interpreted that the HetSNPs in this chromosome should have a small effect on competitive growth, at least in the experimental conditions employed here (YPD 2%; 8 cycles of fermentation at 30°C). As expected, a modest, but symmetric change in growth vigor was observed in the Chr14-UPD strain pair, indicating that our analysis of the inbred collection was likely able to reveal the regions of the genome harboring the strongest heteroalleles contributing to this trait (Fig. 3.5C). Importantly, although the consequences of LOH on this chromosome were not as dramatic as observed for Chr4-UPD and Chr15-UPD strains, they were substantial by the end of 8 cycles of competition. This shows evidence that the cumulative effect of numerous small effect alleles may also contribute to the maintenance of genomic heterozygosity in the long term.

Overall, the approach we used for inducing targeted UPD provided new insights into the relationship between heterozygosity and phenotype in JAY270 and validated the results from our initial genomic association analyses. The use of the heterologous *AmdS* counter-selectable marker for induction of chromosome loss eliminated the need to introduce auxotrophic markers that could influence the phenotypic analyses and, therefore, represents a valuable resource for mapping of genetic traits to specific chromosomes in prototrophic industrial yeast strains. Combining this approach to

recently developed methods for targeted recombination events to chromosomal regions of interest (Sadhu et al. 2016) could significantly accelerate genome association studies by reducing the number of individuals required for preliminary mapping of the most significant contributors to any trait of interest at the chromosome-scale resolution.

## **Materials and Methods**

### **Growth media**

Yeast cells were grown in YPD (20 g/L glucose, 20 g/L peptone, 10 g/L yeast extract, 20 g/L bacteriological agar for solid media), unless otherwise noted.

Transformants carrying the *GFP-kanMX* cassette were selected in YPD plates supplemented with 400 mg/L of geneticin. Selection of *AmdS* positive (*Amds+*) clones was performed in acetamide media (20 g/L glucose, 6.6 g/L potassium sulfate, 1.7 g/L YNB without aminoacids, 0.6 g/L acetamide, 20 g/L bacteriological agar).

Fluoroacetamide media was used for *AmdS* counter-selection (20 g/L glucose, 5 g/L ammonium sulfate, 1.7 g/L YNB without aminoacids, 1.4 g/L complete drop-out mix, 2.3 g/L fluoroacetamide, 20 g/L bacteriological agar). Spot assays for phenotypic screening of the inbred collection was performed in different types of media, including: 2% YPGE (20 g/L peptone, 10 g/L yeast extract, 30 ml/L glycerol; 30 ml/L 100% ethanol, 20 g/L bacteriological agar), 2% YP Galactose (20 g/L galactose, 20 g/L peptone, 10 g/L yeast extract, 20 g/L bacteriological agar), 2% YP Raffinose (20 g/L raffinose, 20 g/L peptone, 10 g/L yeast extract, 20 g/L bacteriological agar).

## **Yeast genetic backgrounds and microbiology procedures**

All *Saccharomyces cerevisiae* strains used in this study descended from the JAY270 background (Table S3.1), a heterothallic diploid derived from the industrial bioethanol strain PE-2 (Argueso et al. 2009). Standard procedures for yeast culture, transformation, crossing and sporulation were followed (Ausubel, F. M.; Brent 1998).

### **Construction of a collection of partial inbred diploids derived from JAY270**

As part of a previous published work from our group, JAY270 cells were induced to sporulate and 14 full tetrads were dissected, resulting in 56 recombinant spores (Table S3.1; (Rodrigues-Prause et al. 2018)). The genome of all 56 haploids isolated were sequenced using an Illumina short read whole genome sequencing platform. Genome sequencing data associated with this study is available in the Sequence Read Archive (SRA) database under study number SRP082524. The sequencing reads were used to build a draft phased map of heterozygous SNPs (HetSNPs) in JAY270 as described in (Rodrigues-Prause et al. 2018). The whole genome information from each haploid parent strain also allowed the generation of genotype maps for each partial inbred diploid that provided precise information of genomic positions that remained heterozygous and positions that became homozygous for either allele (Fig. 3.1C). Two spores were selected per each tetrad to generate the collection of partial inbred strains. We selected one *MAT<sub>a</sub> ace2-A7* and one *MAT<sub>α</sub> ACE2* spore per tetrad. As described previously, JAY270 is heterozygous for a frameshift mutation in the *ACE2* gene (*ACE2/ace2-A7*) and diploid derivatives homozygous for the mutation display an aggregated growth phenotype (Rodrigues-Prause et al. 2018) that could impair the

phenotypic evaluation *ace2-A7/ace2-A7* inbreds through flow cytometry. For this reason, 13 *MAT $\alpha$  ACE2* and 13 *MAT $\alpha$  ace2-A7* spores were crossed in pairwise combinations as detailed in Table S3.1, resulting in 78 partial inbred diploid strains.

### **Construction of a GFP-tagged JAY270 derivative**

A *GFP-KanMX* cassette with homology to a non-coding region located 365 bp upstream to the centromere 5 (*CEN5*, genomic coordinate = 151522) was built. A *GFP* cassette was amplified from pFA6a-TEF2P-GFP-ADH1-NATMX4, kindly provided by Dr. Maitreya Dunham's laboratory, using the primers JAO1385 and JAO1386. The *kanMX4* cassette was amplified from pFA6-KanMX4 using primers JAO1387 and JAO466. Both cassettes were fused by double-joining PCR and transformed into JAY270. Four transformants were selected, purified and tested in 22-cycles of co-culture with the wildtype JAY270 strain, one of which (JAY2208) was used for the co-culture competitions against the inbred and UDP strains.

### **Phenotypic assessment of the inbred collection**

#### Phenotypic screenings through plate spotting assay

Three cultures of JAY270 and of each inbred strain were grown to saturation at 30°C in 96-well plates containing 200  $\mu$ l of YPD. Cultures were diluted by immersing a 96-pin replicator in the resuspended saturated cultures and subsequently in a 96-well plate containing 100  $\mu$ L of distilled water. Diluted cells were pinned in different types of plates and allowed to grow under different conditions as detailed in Table S3.3.

### High temperature stress assay

Cells were refreshed from the - 80°C freezer and incubated at 30°C in YPD plates for 24 hours. Cells were inoculated into 5 ml liquid YPD, and incubated for 24 hours at 30°C in a rotating drum. Saturated cultures were diluted 1000-fold and 40 µL were plated in 4 YPD plates, two of which were incubated for 48 hours at 30°C and two were incubated for 96 hours at 41°C. Variation in growth tolerance to high temperature between strains was assessed by a colony size scoring system. Growth of representative strains assigned to each category is shown Fig. 3.2. This experiment was repeated independently three times for the whole collection of partial inbred strains.

### Flow cytometry-based competitive growth fitness assay

Yeast cells were refreshed from the -80C biofreezer and incubated at 30C in YPD plates for 24 hours. Cells were inoculated into 5 ml liquid YPD, and grown until saturation for 24 hours at 30°C in a rotating drum. Equal volumes of each inbred culture and the JAY270-GFP labeled culture were mixed and used to inoculate three assay tubes containing 5ml of fresh liquid YPD, establishing “Cycle 0” of the competition assay. An aliquot of each mixture was also run through the flow cytometer to determine the starting (pre-culture) ratio of inbred (GFP-) to JAY270 (GFP+) cells. Cultures were incubated at 30C in a rotating drum and every 24 hours (one cycle of competition) 1% of the co-culture volume (50 µl) was transferred to 50 ml of fresh YPD medium. Each experimental co-culture competition was performed in triplicate. The ratio of inbred to wild type cells was assessed at the beginning of cycle 0 and at the end of cycles 2, 5 and 8 using a Cyan ADP7 color flow cytometer coupled to a HyperCyt Rapid Sampler

system for 96-well plate-based assays. 96 well-plates for flow cytometry readings were prepared by diluting 10  $\mu$ l of each culture in 190  $\mu$ l of 1% PBS buffer. A PBS-only well was placed after each triplicate and a triplicates of a control competition between unmarked JAY270 and JAY270-GFP were included every time a new experiment was initiated. Flow cytometry parameters were optimized by applying a series of gatings that excluded from the analysis cell debris (Fig. S3.3A) and cell agglomerates (Fig. S3.3B), resulting in a final cell count that was gated into FITC- and FITC+ populations based on their fluorescence signals (Fig. S3.3C).

### **Identification of allele combinations enriched in high and low fitness strains**

#### Genotype calling of inbred diploids

A previously described phased map of 12,023 heterozygous SNPs (HetSNPs) in JAY270's genome (Rodrigues-Prause et al. 2018) was used for calling the genotype of the recombinant haploid strains that originated the collection of inbreds. The phased JAY270 HetSNP haplotypes were arbitrarily designated as maternal (M) or paternal (P) in order to facilitate genotyping analysis.

CLC genomics workbench software was used for mapping sequencing reads from each parental recombinant haploid on to the S288c reference and detecting SNPs across their whole genome. The nucleotides present at each of the 12,023 loci in the JAY270 HetSNP list were determined for each haploid. When no SNPs were detected at those positions, the reference nucleotide genotype was called, while the alternative nucleotide was called when the alternative SNP was detected at a frequency higher

than 0.95. After the genotypes were determined they were converted to the respective haplotype designations as M or P.

In order to deduce the diploid genotype of each partial inbred strain, we examined the genotype of their respective parents for each of the 12,023 HetSNPs. Heterozygous M/P loci were called whenever both haploid parents presented distinct nucleotides at a specific position. Whenever both parents presented the same nucleotide at a specific position, the locus was designated either homozygous M/M or homozygous P/P.

#### Statistical analysis of genotype/phenotype association

Analysis was done using R version 3.4.0 and the R/qtl package version 1.42-8 (Broman et al. 2003; R Core Team 2017). As inputs for the QTL (quantitative trait loci) analysis, quantitative data from two independent assays (competitive growth and high temperature stress) were used as the phenotype data, and genotypes at 11,774 HetSNPs across the genome were used as the genotype data from 78 partial inbred strains. Markers ~50kb and ~75kb upstream and downstream of the *MAT* and *ACE2* loci (respectively) were excluded from the data set since heterozygosity was forced in those regions by the haploid parent selection criteria described above. Each phenotype was analyzed separately using a one-dimensional scan of the genome. Standard interval mapping was used by applying the expectation maximum algorithm to the data set to determine the log<sub>10</sub> likelihood ratio (LOD) scores for each marker position. Five independent 10,000 permutation tests were run (took median value of the 5 runs) to determine the null distribution of our data and the genome-wide LOD threshold value.

Using the 95<sup>th</sup> percentile of the distribution of maximum LOD scores generated from the permutation tests, this resulted in genome-wide LOD thresholds of LOD > 4.11 for heat stress tolerance and LOD > 4.13 for co-culture competition.

To determine the mode of action of the QTL (additive, dominant/recessive, or overdominance) two different methods were used, focusing on the significant regions in the genome and the marker with the highest LOD score for the region. First, each significant region of the genome was visually inspected using effect plots to show the mean phenotype values for each genotype at each locus of interest. Second, Tukey pairwise comparisons were used to determine which genotypes had significantly different mean phenotype values from each other at an alpha level of 0.05 at each locus of interest. Plots showing magnitude of LOD value vs. whole genome position were generated using the ggplot2 package version 2.2.1 (Wickham 2009).

A two-dimensional scan of the genome was also performed, but no evidence was found for interactions between QTL (data not shown). From this, conclusions were drawn from results of the 1-dimensional scan only.

### **Construction of UPD strains**

Destabilization of centromere function was achieved by the insertion of the counter-selectable gene *AmdS* (Solis-Escalante et al. 2013) at the consensus centromeric region of each chromosome analyzed. Cassettes targeting different integration sites (~100 bp or ~5 bp upstream to the targeted centromere), as well as including or excluding a terminator sequence, were tested. Clones showing uniparental disomy were more frequently achieved when cassettes that excluded the terminator

region of *AmdS* were integrated immediately upstream to the consensus centromere sequence (data not shown). All cassettes were amplified from pCfB2399 (Stovicek et al. 2015) (gift from Irina Borodina; addgene plasmid # 67550) and targeted *CEN4*, *CEN14* and *CEN15*. Transformants were selected and purified in acetamide media. The integration of the cassettes was confirmed by PCR that amplified the left and right junctions between the *AmdS* cassette and the centromeric region. PCR products were designed to span at least one HetSNP that was used to determine in which homolog the integration occurred using Sanger sequencing. At least two independent clones containing the insertion in each homolog were selected. Cells were grown in YPD plates for 24 hours to allow for loss of the *AmdS*-marked homolog, diluted in 200  $\mu$ l of water and plated in the counter-selection fluoroacetamide media. At least four types of cells should be able to grow in this selective media: (1) cell with inactivation of the *AmdS* gene through point mutation; (2) cells that acquired loss-of-heterozygosity (LOH) tracts spanning *AmdS* as a result of a DNA repair mechanism, such as gene conversion or mitotic recombination leading to LOH, (3) cells that lost the whole homolog containing *AmdS* and persisted as monosomics; and (4), our targeted UDP class, cells that lost the whole homolog containing *AmdS*, undergoing a transient monosomic state followed by endoduplication of the *AmdS*- homolog. Two sequential tests were performed to screen for true UDP clones. First, candidates were genotyped using restriction fragment length polymorphism (RFLP) analysis at three genomic positions along the chromosome (HetSNPs near the left and right ends of the chromosome arms, and near the centromeric region). Clones genotyped as homozygous at all three markers could be either monosomic or homozygous disomic for the chromosome of interest. To

distinguish between these cell types, candidates were sporulated and tetrads were dissected. Monosomic clones should generate tetrads with two viable and two inviable spores, whereas UDP clones should generate tetrads with four viable spores. Candidates that were homozygous for all three RFLP markers and produced tetrads with four viable spores were selected for phenotypic tests.

## REFERENCES

- Amorim, H.V, M.L. Lopes, J.V. de Castro Oliveira, M. S. Buckeridge, and G.H. Goldman. 2011. "Scientific Challenges of Bioethanol Production in Brazil." *Applied Microbiology and Biotechnology* 91 (5): 1267–75.
- Argueso, Juan Lucas, Marcelo F. Carazzolle, Piotr A. Mieczkowski, Fabiana M. Duarte, Osmar V C Netto, Silvia K. Missawa, Felipe Galzerani, et al. 2009. "Genome Structure of a *Saccharomyces Cerevisiae* Strain Widely Used in Bioethanol Production." *Genome Research* 19 (12): 2258–70. <https://doi.org/10.1101/gr.091777.109>.
- Ausubel, F. M.; Brent, R. et al. 1998. *Current Protocols in Molecular Biology.Pdf*. New York: John Wiley & Sons Inc.
- Babrzadeh, F., R. Jalili, C. Wang, S. Shokralla, S. Pierce, A. Robinson-Mosher, P. Nyren, et al. 2012. "Whole-Genome Sequencing of the Efficient Industrial Fuel-Ethanol Fermentative *Saccharomyces Cerevisiae* Strain CAT-1." *Molecular Genetics and Genomics : MGG* 287 (6): 485–94.
- Basso, L.C., H.V. Amorim, A.J. Oliveira, and M.L. Lopes. 2008. "Yeast Selection for Fuel Ethanol Production in Brazil." *FEMS Yeast Research* 8 (7): 1155–63.
- Birchler, James a, Hong Yao, and Sivanandan Chudalayandi. 2006. "Unraveling the Genetic Basis of Hybrid Vigor." *Proceedings of the National Academy of Sciences of the United States of America* 103 (35): 12957–58. <https://doi.org/10.1073/pnas.0605627103>.
- Bittante, Giovanni, Luigi Gallo, and Paolo Montobbio. 1993. "Estimated Breed Additive Effects and Direct Heterosis for Growth and Carcass Traits of Heavy Pigs." *Livestock Production Science* 34: 101–14.
- Broman, Karl W., Hao Wu, Śaunak Sen, and Gary A. Churchill. 2003. "R/Qtl: QTL Mapping in Experimental Crosses." *Bioinformatics* 19 (7): 889–90. <https://doi.org/10.1093/bioinformatics/btg112>.
- Carvalho-Netto, Osmar V., Marcelo F. Carazzolle, Aline Rodrigues, Welbe O. Bragança, Gustavo G L Costa, Juan Lucas Argueso, and Gonçalo A G Pereira. 2013. "A Simple and Effective Set of PCR-Based Molecular Markers for the Monitoring of the *Saccharomyces Cerevisiae* Cell Population during Bioethanol Fermentation." *Journal of Biotechnology* 168 (4): 701–9. <https://doi.org/10.1016/j.jbiotec.2013.08.025>.

- Chakraborty, Parijat, Ajith V Pankajam, Abhishek Dutta, and Koodali T. Nishant. 2018. "Genome Wide Analysis of Meiotic Recombination in Yeast : For a Few SNPs More." *IUBMB Life*. <https://doi.org/10.1002/iub.1877>.
- Charlesworth, Deborah, Deborah Charlesworth, John H Willis, and John H Willis. 2009. "The Genetics of Inbreeding Depression." *Nature Reviews. Genetics* 10: 783–96. <https://doi.org/10.1038/nrg2664>.
- Chen, Z. Jeffrey. 2013. "Genomic and Epigenetic Insights into the Molecular Bases of Heterosis." *Nature Reviews Genetics* 14 (7): 471–82. <https://doi.org/10.1038/nrg3503>.
- Coutouné, Natalia, Aline Tieppo, Nogueira Mulato, Diego Mauricio Riaño-pachón, and Juliana Velasco de Castro Oliveira. 2017. "Draft Genome Sequence of *Saccharomyces Cerevisiae* Barra Grande (BG-1), a Brazilian Industrial Bioethanol- Producing Strain" 5 (13): 1–2.
- Della-Bianca, Bianca Eli, Thiago Olitta Basso, Boris Ugarte Stambuk, Luiz Carlos Basso, and Andreas Karoly Gombert. 2013. "What Do We Know about the Yeast Strains from the Brazilian Fuel Ethanol Industry?" *Applied Microbiology and Biotechnology* 97 (3): 979–91. <https://doi.org/10.1007/s00253-012-4631-x>.
- Fareed, Mohd, and Mohammad Afzal. 2014. "Estimating the Inbreeding Depression on Cognitive Behavior: A Population Based Study of Child Cohort." *PLoS ONE* 9 (10): e109585. <https://doi.org/10.1371/journal.pone.0109585>.
- Fry, James D., Stefanie L. Heinsohn, and Trudy F.C. Mackay. 1998. "Heterosis for Viability, Fecundity, and Male Fertility in *Drosophila Melanogaster*: Comparison of Mutational and Standing Variation." *Genetics* 148 (3): 1171–88.
- Gama, L. T., M. C. Bressan, E. C. Rodrigues, L. V. Rossato, O. C. Moreira, S. P. Alves, and R. J B Bessa. 2013. "Heterosis for Meat Quality and Fatty Acid Profiles in Crosses among *Bos Indicus* and *Bos Taurus* Finished on Pasture or Grain." *Meat Science* 93 (1): 98–104. <https://doi.org/10.1016/j.meatsci.2012.08.005>.
- Granek, Joshua A., Debra Murray, Ömür Kayrkçi, and Paul M. Magwene. 2013. "The Genetic Architecture of Biofilm Formation in a Clinical Isolate of *Saccharomyces Cerevisiae*." *Genetics* 193 (2): 587–600. <https://doi.org/10.1534/genetics.112.142067>.
- Groszmann, Michael, Ian K. Greaves, Ryo Fujimoto, W. James Peacock, and Elizabeth S. Dennis. 2013. "The Role of Epigenetics in Hybrid Vigour." *Trends in Genetics* 29 (12): 684–90. <https://doi.org/10.1016/j.tig.2013.07.004>.
- Guo, Tingting, Ning Yang, Hao Tong, and Qingchun Pan. 2014. "Genetic Basis of Grain Yield Heterosis in an "immortalized F2" Maize Population." *Theor Appl Genet*, 2149–58. <https://doi.org/10.1007/s00122-014-2368-x>.

Hill, A, and K Bloom. 1987. "Genetic Manipulation of Centromere Function." *Molecular and Cellular Biology* 7 (7): 2397–2405. <https://doi.org/10.1128/MCB.7.7.2397.Updated>.

Hochholdinger, Frank, and Nadine Hoecker. 2007. "Towards the Molecular Basis of Heterosis." *Trends in Plant Science* 12 (9): 427–32. <https://doi.org/10.1016/j.tplants.2007.08.005>.

Hoffman, Joseph I, Fraser Simpson, Patrice David, Jolianne M Rijks, Thijs Kuiken, Michael a S Thorne, Robert C Lacy, and Kanchon K Dasmahapatra. 2014. "High-Throughput Sequencing Reveals Inbreeding Depression in a Natural Population." *Proceedings of the National Academy of Sciences of the United States of America* 111 (10): 3775–80. <https://doi.org/10.1073/pnas.1318945111>.

Huang, Xuehui, Shihua Yang, Junyi Gong, Yan Zhao, Qi Feng, Hao Gong, Wenjun Li, et al. 2015. "Genomic Analysis of Hybrid Rice Varieties Reveals Numerous Superior Alleles That Contribute to Heterosis." *Nature Communications*, 1–9. <https://doi.org/10.1038/ncomms7258>.

J.A. Birchler, H Yao, S. Chudalayandi, D. Vaiman, R.A.Veitia. 2010. "Heterosis." *The Plant Cell* 22 (July): 2105–12. <https://doi.org/10.1105/tpc.110.076133>.

Jones, D. F.. 1917. "Dominance of Linked Factors as a Means of Accounting for Heterosis," 310–12.

Kaeppler, Shawn. 2012. "Heterosis: Many Genes, Many Mechanisms—End the Search for an Undiscovered Unifying Theory." *ISRN Botany* 2012: 1–12. <https://doi.org/10.5402/2012/682824>.

Krieger, Uri, Zachary B Lippman, and Dani Zamir. 2010. "The Flowering Gene SINGLE FLOWER TRUSS Drives Heterosis for Yield in Tomato." *Nature Genetics* 42 (5): 459–63. <https://doi.org/10.1038/ng.550>.

Li, Zhi Kang, L. J. Luo, H. W. Mei, D. L. Wang, Q. Y. Shu, R. Tabien, D. B. Zhong, et al. 2001. "Overdominant Epistatic Loci Are the Primary Genetic Basis of Inbreeding Depression and Heterosis in Rice. I. Biomass and Grain Yield." *Genetics* 158 (4): 1737–53.

Magwene, Paul M., Ömür Kayıkçı, Joshua A. Granek, Jennifer M. Reininga, Zackary Scholl, and Debra Murray. 2011. "Outcrossing, Mitotic Recombination, and Life-History Trade-Offs Shape Genome Evolution in *Saccharomyces Cerevisiae*." *Proceedings of the National Academy of Sciences* 108 (5): 1987–92. <https://doi.org/10.1073/pnas.1012544108>.

Magwene, Paul M. 2014. "Revisiting Mortimer's Genome Renewal Hypothesis: Heterozygosity, Homothallism, and the Potential for Adaptation in Yeast." *Adv Exp Med Biol* 781: 37–48. <https://doi.org/10.1007/978-94-007-7347-9>.

Mancera, Eugenio, Richard Bourgon, Alessandro Brozzi, Wolfgang Huber, and M Steinmetz. 2008. "High-Resolution Mapping of Meiotic Crossovers and Noncrossovers in Yeast" 454 (7203): 479–85. <https://doi.org/10.1038/nature07135.High-resolution>.

McQuillan, Ruth, Niina Eklund, Nicola Pirastu, Maris Kuningas, Brian P. McEvoy, Tõnu Esko, Tanguy Corre, et al. 2012. "Evidence of Inbreeding Depression on Human Height." *PLoS Genetics* 8 (7). <https://doi.org/10.1371/journal.pgen.1002655>.

Melchinger, a. E., H. F. Utz, H. P. Piepho, Z. B. Zeng, and C. C. Schön. 2007. "The Role of Epistasis in the Manifestation of Heterosis: A Systems-Oriented Approach." *Genetics* 177 (November): 1815–25. <https://doi.org/10.1534/genetics.107.077537>.

Minvielle, Francis. 1987. "Dominance Is Not Necessary for Heterosis: A Two-Locus Model." *Genetical Research*. <https://doi.org/10.1017/S0016672300027142>.

Ng, Danny W K, Jie Lu, and Z. Jeffrey Chen. 2012. "Big Roles for Small RNAs in Polyploidy, Hybrid Vigor, and Hybrid Incompatibility." *Current Opinion in Plant Biology* 15 (2): 154–61. <https://doi.org/10.1016/j.pbi.2012.01.007>.

Peter, Jackson, Matteo De Chiara, Anne Friedrich, Jia-xing Yue, David Pflieger, Anders Bergström, Anastasie Sigwalt, et al. 2018. "Genome Evolution across 1,011 *Saccharomyces Cerevisiae* Isolates." *Nature* 556. <https://doi.org/10.1038/s41586-018-0030-5>.

Plech, Marcin, J Arjan G M de Visser, and Ryszard Korona. 2014. "Heterosis Is Prevalent among Domesticated but Not Wild Strains of *Saccharomyces Cerevisiae*." *G3 (Bethesda, Md.)* 4 (February): 315–23. <https://doi.org/10.1534/g3.113.009381>.

R Core Team 2017, R. n.d. "R Core Team (2017). R: A Language and Environment for Statistical Computing." *R Foundation for Statistical Computing, Vienna, Austria*. <https://doi.org/10.1016/j.cell.2010.09.012>.

Reid, Robert J D, Ivana Sunjevaric, Warren P. Voth, Samantha Ciccone, Wendy Du, Aileen E. Olsen, David J. Stillman, and Rodney Rothstein. 2008. "Chromosome-Scale Genetic Mapping Using a Set of 16 Conditionally Stable *Saccharomyces Cerevisiae* Chromosomes." *Genetics* 180 (4): 1799–1808. <https://doi.org/10.1534/genetics.108.087999>.

Rodrigues-Prause, Aline, Nadia M V Sampaio, Theodore M Gurol, Guadalupe M Aguirre, Hailey N C Sedam, Mary J Chapman, Ewa P Malc, et al. 2018. "Characterization of a Yeast Colony Morphology Phenotypic Transition Caused by

Mitotic Recombination Leading to Loss-of-Heterozygosity.” *G3 Genes Genomes Genetics*.

Sadhu, Meru J., Joshua S. Bloom, Laura Day, and Leonid Kruglyak. 2016. “CRISPR-Directed Mitotic Recombination Enables Genetic Mapping without Crosses.” *Science* 352 (6289): 1113–16. <https://doi.org/10.1126/science.aaf5124>.

Schiermiester, L N, R M Thallman, L A Kuehn, S D Kachman, and M L Spangler. 2015. “Estimation of Breed-Specific Heterosis Effects for Birth , Weaning , and Yearling Weight in Cattle,” 46–52. <https://doi.org/10.2527/jas2014-8493>.

Schnable, Patrick S, and Nathan M Springer. 2013. “Progress toward Understanding Heterosis in Crop Plants.” *Annual Review of Plant Biology* 64: 71–88. <https://doi.org/10.1146/annurev-arplant-042110-103827>.

Shapira, R, T Levy, S Shaked, E Fridman, and L David. 2014. “Extensive Heterosis in Growth of Yeast Hybrids Is Explained by a Combination of Genetic Models.” *Heredity*, no. October 2013: 1–11. <https://doi.org/10.1038/hdy.2014.33>.

Shapira, Rachel, and Lior David. 2016. “Genes with a Combination of Over-Dominant and Epistatic Effects Underlie Heterosis in Growth of *Saccharomyces Cerevisiae* at High Temperature.” *Frontiers in Genetics* 7 (MAY): 1–14. <https://doi.org/10.3389/fgene.2016.00072>.

Shull, G.H. 1911. “The Genotypes of Maize” 45 (532): 234–52.

Smukowski Heil, Caiti S., Christopher G. DeSevo, Dave A. Pai, Cheryl M. Tucker, Margaret L. Hoang, and Maitreya J. Dunham. 2017. “Loss of Heterozygosity Drives Adaptation in Hybrid Yeast.” *Molecular Biology and Evolution* 34 (7): 1596–1612. <https://doi.org/10.1093/molbev/msx098>.

Solis-Escalante, Daniel, Niels G A Kuijpers, Nadine Bongaerts, Irina Bolat, Lizanne Bosman, Jack T. Pronk, Jean Marc Daran, and Pascale Daran-Lapujade. 2013. “AmdSYM, A New Dominant Recyclable Marker Cassette for *Saccharomyces Cerevisiae*.” *FEMS Yeast Research* 13 (1): 126–39. <https://doi.org/10.1111/1567-1364.12024>.

Stovicek, Vratislav, Gheorghe M. Borja, Jochen Forster, and Irina Borodina. 2015. “EasyClone 2.0: Expanded Toolkit of Integrative Vectors for Stable Gene Expression in Industrial *Saccharomyces Cerevisiae* Strains.” *Journal of Industrial Microbiology and Biotechnology* 42 (11): 1519–31. <https://doi.org/10.1007/s10295-015-1684-8>.

Symington, L. S., R. Rothstein, and M. Lisby. 2014. “Mechanisms and Regulation of Mitotic Recombination in *Saccharomyces Cerevisiae*.” *Genetics* 198 (3): 795–835. <https://doi.org/10.1534/genetics.114.166140>.

Technow, Frank, Tobias a. Schrag, Wolfgang Schipprack, Eva Bauer, Henner Simianer, and Albrecht E. Melchinger. 2014. "Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize." *Genetics* 197 (August): 1343–55. <https://doi.org/10.1534/genetics.114.165860>.

Wang, Li, Ian K. Greaves, Michael Groszmann, Li Min Wu, Elizabeth S. Dennis, and W. James Peacock. 2015. "Hybrid Mimics and Hybrid Vigor in *Arabidopsis*." *Proceedings of the National Academy of Sciences* 112 (35): E4959–67. <https://doi.org/10.1073/pnas.1514190112>.

Wickham, H. 2009. "Ggplot2 - Elegant Graphics for Data Analysis." *Journal of Statistical Software*. <https://doi.org/10.18637/jss.v077.b02>.

Wolf, D P, and Arnel R. Hallauer. 1997. "Triple Test Cross Analysis to Detect Epistasis in Maize." *Crop Science* 37: 763–70.

Xiao, J., J. Li, L. Yuan, and S. D. Tanksley. 1995. "Dominance Is the Major Genetic Basis of Heterosis in Rice as Revealed by QTL Analysis Using Molecular Markers." *Genetics* 140: 745–54.

## CHAPTER IV

### **Conclusions and future directions**

My doctoral work consisted of two parallel tracks aimed at (1) characterizing a novel phenomenon of genome instability in yeast that can rapidly erode heterozygosity through mitotic recombination, and (2) determining how genomic heterozygosity plays a role in conferring traits of interest to the industrial yeast strain JAY270. Below, I summarize and discuss the main observations and conclusions of my studies, their limitations, and some of the prospective avenues of research that my work has opened.

#### **Systemic genomic instability and resulting rapid karyotype evolution**

It is now evident that mutations are not always acquired sequentially as a result of independent lesions to the genome. Recent whole-genome sequencing studies, particularly ones leveraging single-cell methods, have identified different classes of mutational phenomena in which multiple chromosomal alterations and point mutation “showers” appear to accumulate in an “all-at-once” fashion. The investigation of the molecular mechanisms underlying such mutational bursts and the reconstruction of cell lineages represent significant technical challenges in the field. Even though some of these challenges have started to be overcome with the development of spatially-resolved single cell sequencing (Casasent et al. 2018), which allows the application of more robust phylogenetic inference methods to determine cell lineages, the

establishment of simpler, yet powerful, model systems would greatly facilitate research in the field.

In Chapter II of this dissertation, we showed evidence that multiple events of LOH accumulate in short windows of time in the yeast *S. cerevisiae*, a phenomenon that we named “systemic genomic instability”. We demonstrated that selection of one LOH event helps to predict the existence of additional unselected LOH events distributed throughout the genome, which indicates a subset of cells in the population might have undergone episodes of genomic instability that resulted in multiple coincident LOH events. This observation was validated by quantification of double LOH events in three different yeast backgrounds, which revealed that two coincident LOH events occur at rates 30 to 100 times higher than would be expected if they had emerged independently of each other. We further demonstrated that this process is truly mitotic and not a result of initiation of meiotic recombination followed by return-to-growth. Our work provides evidence for the occurrence of short episodes of systemic genomic instability in yeast, which we believe could represent a powerful model system for the characterization of the underlying universal mechanisms of such mutational bursts. Together, our initial results and the development of this experimental platform have set the stage for exciting future investigations. What is the duration of the episodes of systemic genomic instability? Are mutational bursts associated to mutational signatures other than LOH, for example, point mutations, CNVs and aneuploidies? What are the mechanisms underlying mutational bursts? Is the instability transient or prolonged? Can these results be recapitulated in human cell lines? Although challenges related to single-cell analyses and pedigree and phylogenetic inference still remain, we envision our yeast model

system offers several opportunities to test hypotheses that can later be validated in human cell lines and tumor samples. Interesting areas to pursue include, for instance, the application of the Mother Enrichment Program (MEP)(Lindstrom and Gottschling 2009) to test a potential relationship between ageing and mitSGL, application of recently released technologies (ex.: Milo<sup>TM</sup> Single-Cell Westerns) that allow protein quantification of thousands of single cells to test a potential effect of transient defects in gene expression in mitSGL and the development of assays that will allow the identification of potential bursts of point mutations, CNVs, chromosome losses and gains, etc. Another interesting area to explore is whether this phenomenon would also be observed in human cell lines. A suitable system for this analysis would be the TK6 human B-lymphoblastoid cell line. This cell line is heterozygous on chromosome 17q at the counter-selectable thymidine kinase locus (*Tk<sup>+</sup>/Tk<sup>-</sup>* genotype), which allows for selection of one primary LOH events at that region (Revollo et al. 2016). Preliminary evidence in the literature suggests that clones carrying selected LOH spanning the TK locus also show unselected events distributed genome-wide at a frequency that is higher than expected for a model of gradual accumulation of mutations (Li, Yandell, and Little 1994). These observations were based on the interrogation of only in a few (~12) microsatellite markers distributed throughout the human genome and a much more comprehensive characterization of this phenomenon could be achieved with the latest genome sequencing technologies. The work described in this dissertation and future developments that will help answering the relevant questions described above will allow us to contribute discoveries that may have important implications for cancer dynamics and genome evolution in general.

## **Effects of heterozygosity on phenotype and its influence in genome configuration**

Although it has been known for decades that some species show a positive correlation between heterozygosity and phenotype, this relationship is still not well characterized in yeast. To our knowledge, the work presented in Chapter III of this dissertation, is one of the first to investigate the role that genomic heterozygosity plays on the phenotype of a wild and naturally heterozygous yeast strain. Two different approaches were used to reduce the levels of heterozygosity in the bioethanol strain JAY270. Initially, we used inbreeding to generate a collection of 78 partial inbred strains showing an approximate 50% reduction of the heterozygous alleles present in the parent strain JAY270. We observed a wide phenotypic variation in the competitive growth and heat tolerance of inbred strains, supporting a substantial role of heterozygosity in JAY270's genome and in its industrial traits. To validate these results, we developed an approach that allowed the construction of strains containing blocks of homozygosity restricted to a single chromosome, specifically Chr4, Chr14 and Chr15. Again, the growth vigor and heat tolerance phenotypes of all strains with chromosome-specific uniparental disomy differed significantly from the parent JAY270. Together, these results suggested that heterozygosity does contribute to JAY270's desirable traits, which consequently may explain the long-term maintenance of this genome configuration in its natural niche.

Although size of the inbred collection limited the resolution of our genome mapping analysis, the work presented here pointed to candidate genomic regions that are likely contain genes important for the phenotypes tested. Our results have laid the groundwork for next research projects aimed at further interrogating these candidate

regions and identifying the specific SNPs underlying the phenotypes observed. This could be achieved through reciprocal hemizyosity analysis(Steinmetz et al. 2002) or recently developed CRISPR-based techniques that allow genetic mapping refining to target regions(Sadhu et al. 2016). The specific genes identified could then be manipulated to improve microbial performance in industrial application. Beyond potential biotechnological applications, our results provide new insights into how the interplay between heterozygosity and fitness may contribute in shaping the genome of wild *S. cerevisiae* strains.

## REFERENCES

Casasent, Anna K., Aislyn Schalck, Ruli Gao, Emi Sei, Annalyssa Long, William Pangburn, Tod Casasent, Funda Meric-Bernstam, Mary E. Edgerton, and Nicholas E. Navin. 2018. "Multiclonal Invasion in Breast Tumors Identified by Topographic Single Cell Sequencing." *Cell* 172 (1–2): 205–217.e12. <https://doi.org/10.1016/j.cell.2017.12.007>.

Li, C Y, D W Yandell, and J B Little. 1994. "Elevated Frequency of Microsatellite Mutations in TK6 Human Lymphoblast Clones Selected for Mutations at the Thymidine Kinase Locus." *Mol Cell Biol* 14 (7): 4373–79. <https://doi.org/10.1128/MCB.14.7.4373>.

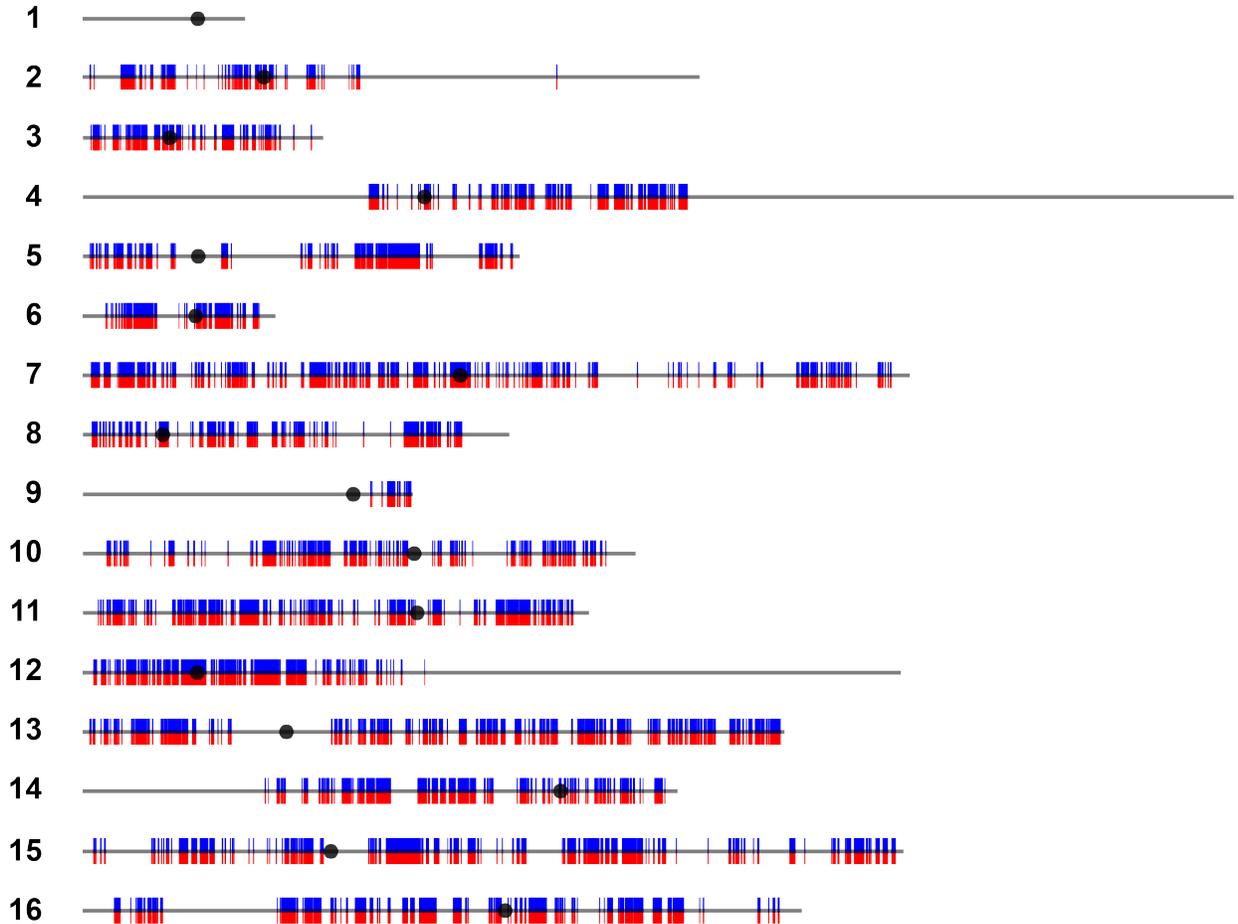
Lindstrom, Derek L., and Daniel E. Gottschling. 2009. "The Mother Enrichment Program: A Genetic System for Facile Replicative Life Span Analysis in *Saccharomyces Cerevisiae*." *Genetics* 183 (2): 413–22. <https://doi.org/10.1534/genetics.109.106229>.

Revollo, Javier, Dayton M Petibone, Page Mckinzie, Bridgett Knox, Suzanne M Morris, Baitang Ning, and Vasily N Dobrovolsky. 2016. "Whole Genome and Normalized mRNA Sequencing Reveal Genetic Status of TK6, WTK1, and NH32 Human B-Lymphoblastoid Cell Lines" 795: 60–69.

Sadhu, Meru J., Joshua S. Bloom, Laura Day, and Leonid Kruglyak. 2016. "CRISPR-Directed Mitotic Recombination Enables Genetic Mapping without Crosses." *Science* 352 (6289): 1113–16. <https://doi.org/10.1126/science.aaf5124>.

Steinmetz, Lars M, Himanshu Sinha, Dan R Richards, Jamie I Spiegelman, Peter J Oefner, John H McCusker, and Ronald W Davis. 2002. "Dissecting the Architecture of a Quantitative Trait Locus in Yeast." *Nature* 416 (6878): 326–30. <https://doi.org/10.1038/416326a>.

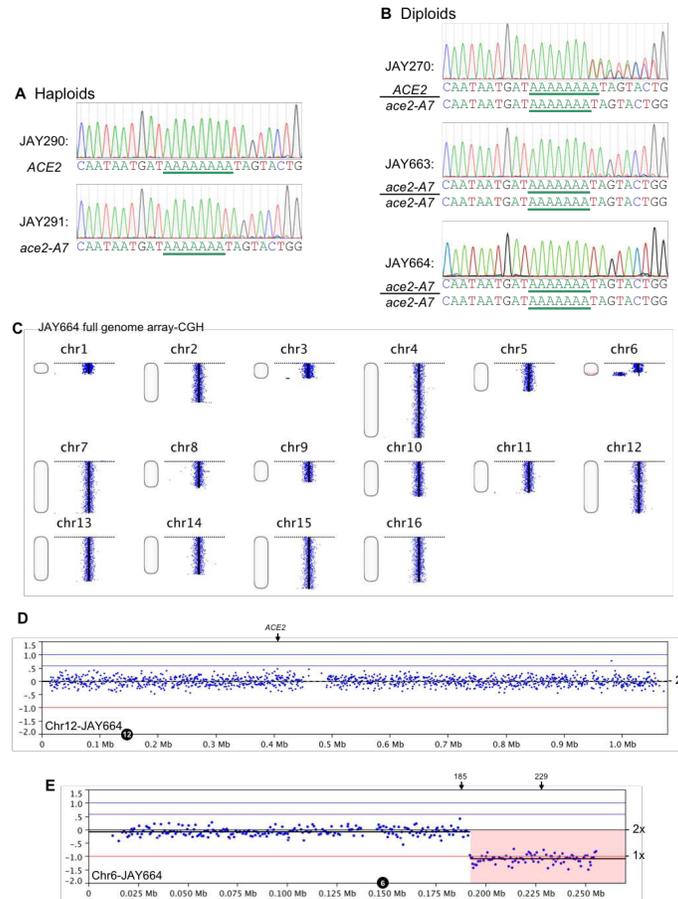
## APPENDIX A: SUPPLEMENTAL FIGURES



**Figure S2.1. Draft map of heterozygosity in the JAY270 genome.**

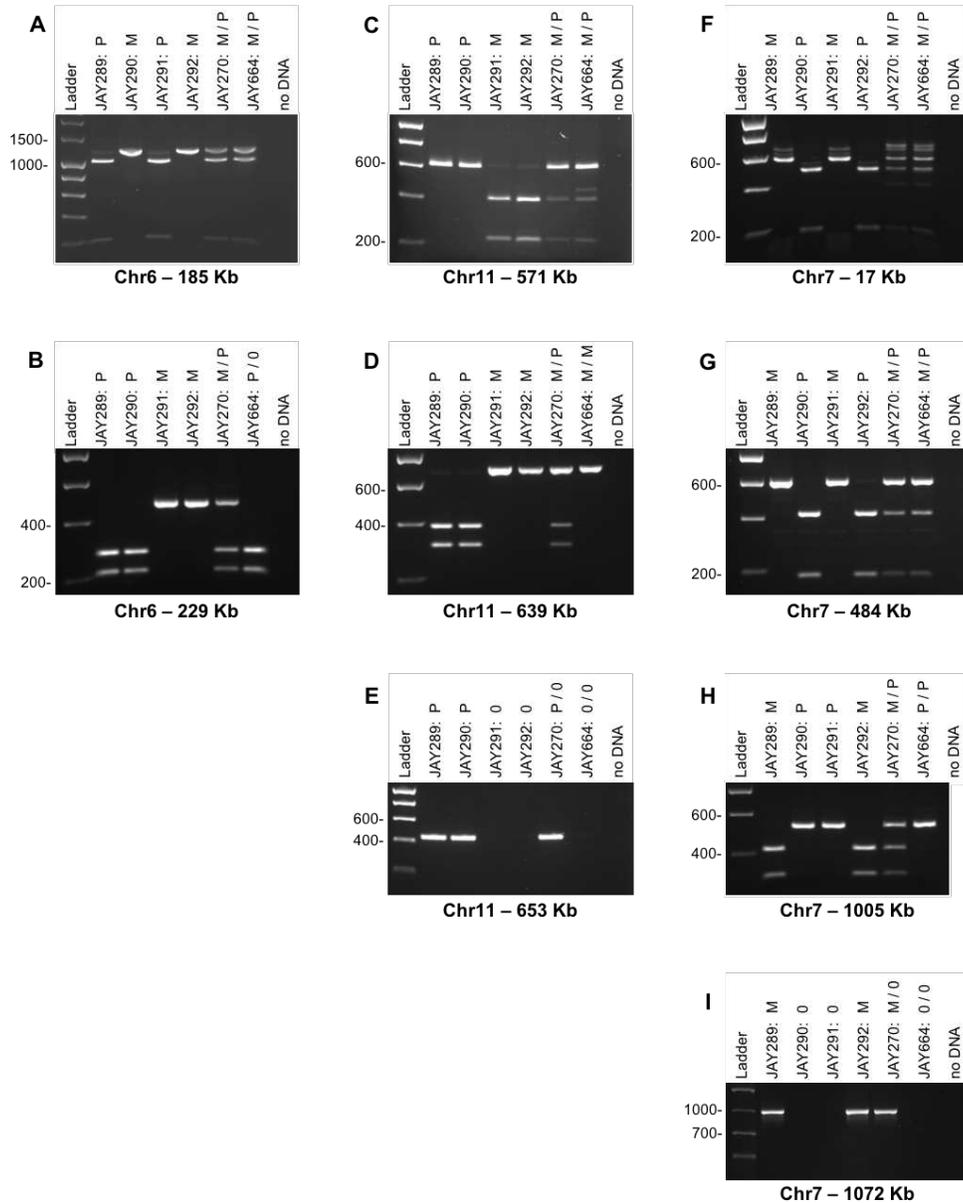
The distribution of 12,023 phased heterozygous single nucleotide polymorphisms is depicted (HetSNPs; double colored markers, arbitrarily defined as paternal [blue] or maternal [red] haplotypes). Each horizontal line represents a *S. cerevisiae* chromosome. Black circles indicate the positions of the centromeres. The HetSNPs are unevenly distributed across the genome, with several regions showing evidence of ancestral LOH events that likely occurred in the JAY270 lineage. Plots were generated to scale in Python 2.7 using the matplotlib package and a custom script. For size reference, Chr1 is 230 Kb.





**Figure S2.3. Sanger DNA sequencing analysis of *ACE2* locus and genome-wide copy number in JAY664.**

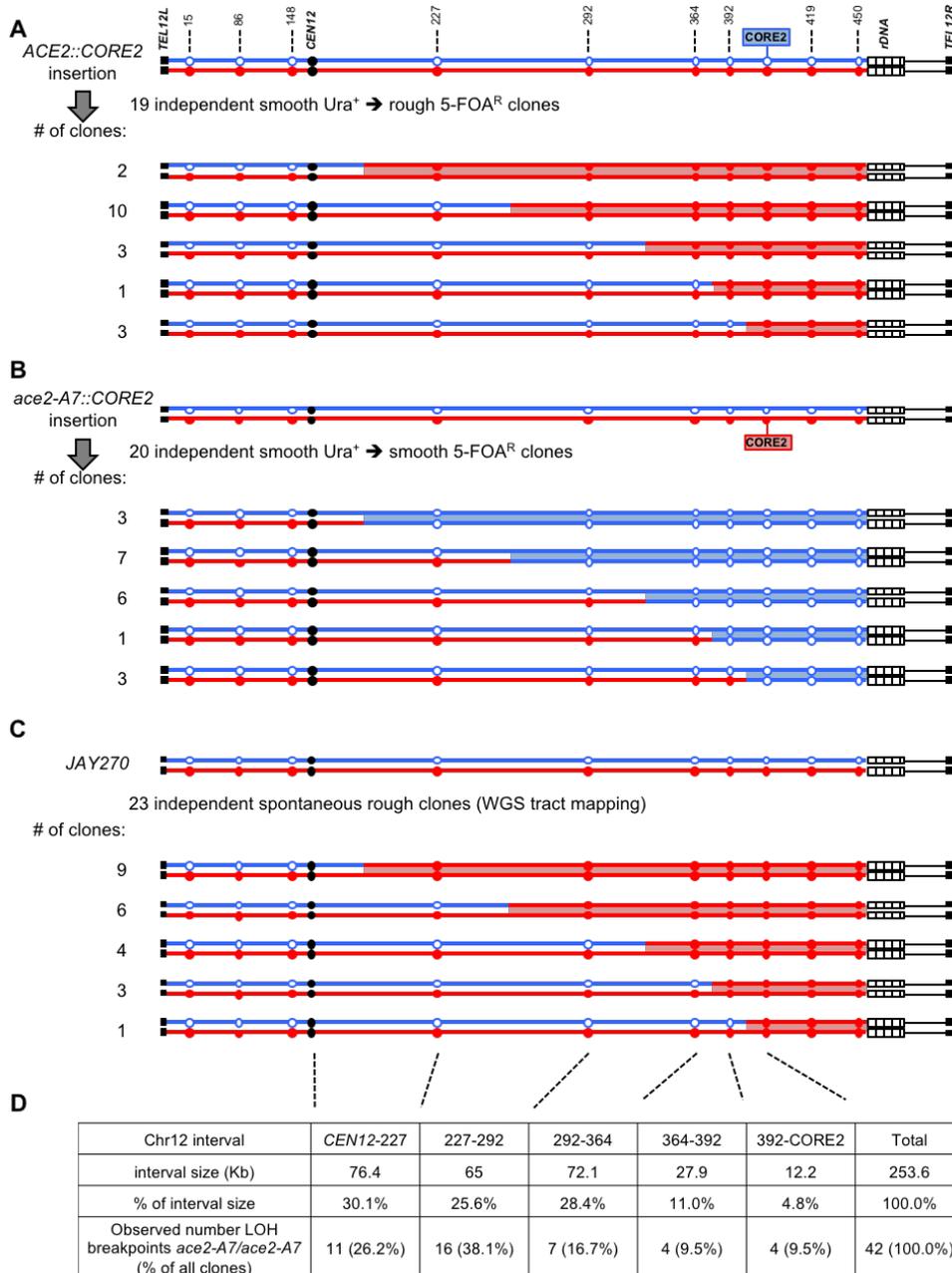
Panel **A** shows segments of PCR-Sanger sequencing chromatograms for the *ACE2* locus from haploids with the wild type *ACE2* (JAY290) and mutant *ace2-A7* (JAY291) alleles. Note the difference in the A-homopolymer runs, with eight consecutive peaks in *ACE2* and seven consecutive peaks in *ace2-A7*. Panel **B** shows PCR-Sanger chromatograms from diploids (primer extension was from left to right in all cases). The JAY270 chromatogram is a mixture of the two alleles (heterozygous), while the rough colony clones JAY663 and JAY664 only had the *ace2-A7* pattern. The inferred DNA sequences in **A** and **B** are shown below the chromatograms. Panel **C** displays the array-CGH genome-wide plot of copy number for the JAY664 clone relative to its JAY270 parent strain, as described in Fig. 4 in the main text. Each of the 16 *S. cerevisiae* chromosomes has its own plot, shown vertically. Probe signals clustered near the center indicate neutral copy number (2x), while clustering to the right indicates amplifications and clustering to the left indicates deletions. Panels **D** and **E** show single chromosome plots for Chr12 and Chr6, respectively. No copy number alterations were detected for Chr12, including at the *ACE2* locus (indicated). Chr6 had a deletion on the right arm (pink shading). The positions of two PCR markers shown in Fig. S4 are indicated.



**Figure S2.4. PCR genotyping of LOH and CNV in JAY664.**

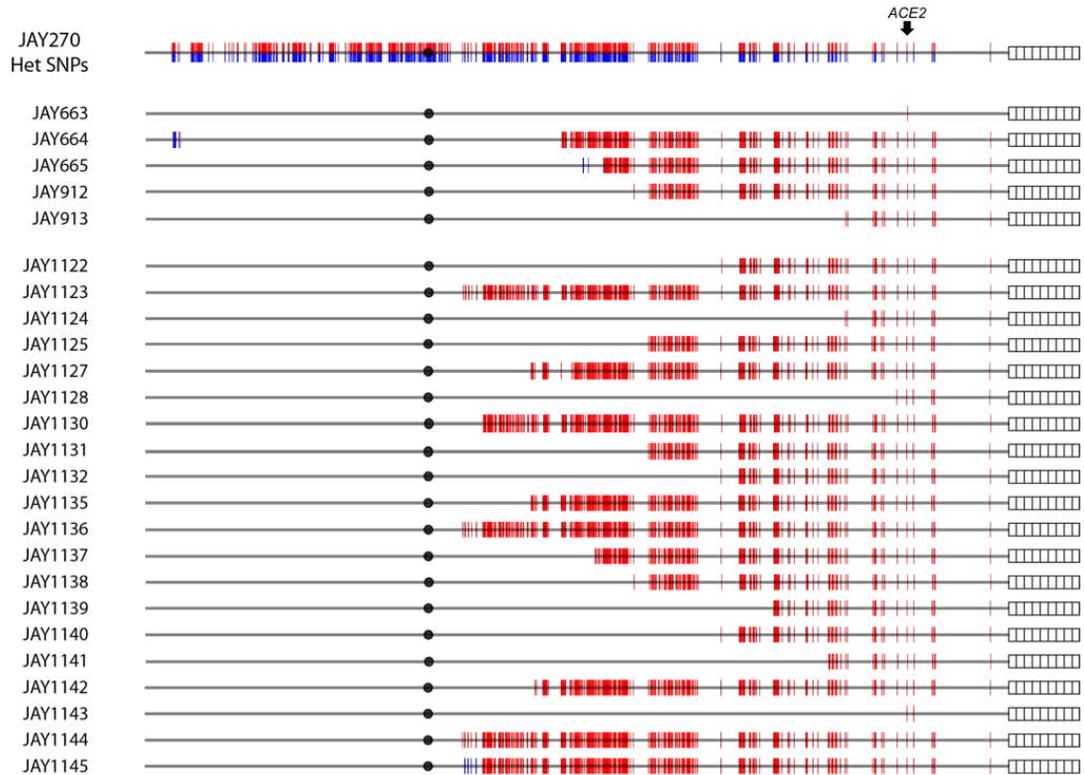
Panels **A-I** show agarose gel electrophoresis images of PCR-RFLP genotyping of HetSNP markers, and PCR detection of hemizygous regions of Chr6 (Fig. S3E), Chr11 (Fig. 4B) and Chr7 (Fig. 4C). The lanes in all gels contain four sibling haploids from one JAY270 tetrad showing the segregation of each marker, the hetero- or hemizygous JAY270 diploid, the rough colony isolate JAY664, and a no DNA template negative control. The inferred genotypes are indicated for each haploid strain as M for maternal, P for paternal, or 0 for absence of a sequence that is hemizygous in the JAY270 diploid. For diploids, the genotypes are indicated as M/P for heterozygous, M/0 or P/0 for hemizygous specifying which homolog has the sequence detected by PCR (M or P) and which homolog does not (0). A 0/0 genotype indicates cases where JAY664 lost hemizygosity for the sequence detected by PCR. For the right arm of Chr6 (**A-B**) JAY664

is heterozygous at a region proximal to a gross deletion breakpoint detected by array CGH, and is hemizygous for the distal region. For the right arm of Chr11 (**C-E**) JAY664 is heterozygous for a HetSNP proximal to an LOH breakpoint, homozygous for a HetSNP distal to the LOH breakpoint, and completely lost a DNA sequence further down the chromosome that is hemizygous in JAY270. An analogous event (proximal LOH and distal CNV) is shown for the right arm of Chr7 (**G-I**). Note that Chr7 in JAY664 is heterozygous at the left arm near the left telomere (**F**), even though an amplification event was detected by array-CGH for the distal left tip of Chr7 (Fig. 4C). This amplification event is a reflection of the LOH event at the left end of Chr12 where the amplified probes are found as a hemizygous sequence in JAY270.



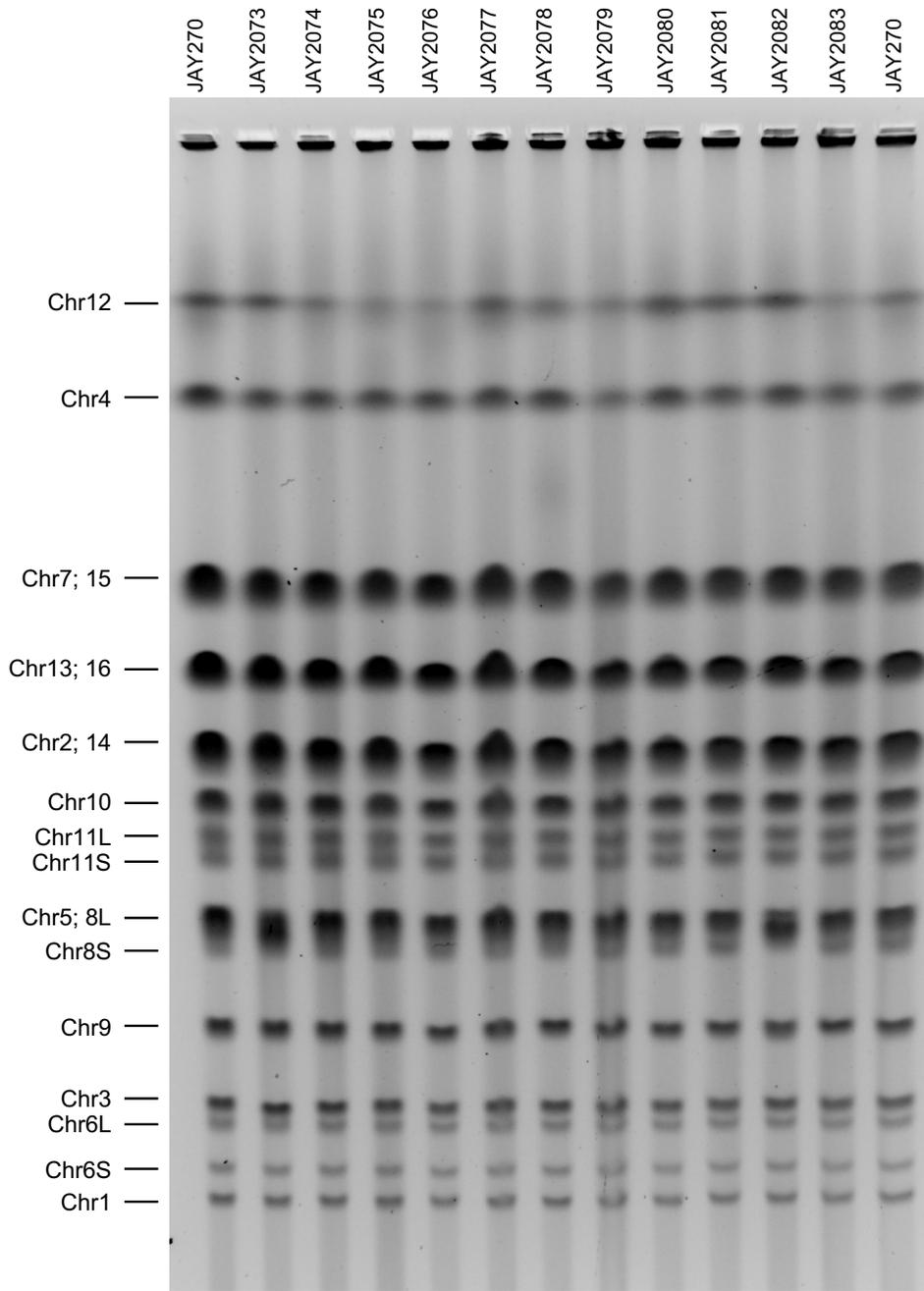
**Figure S2.5. Analysis of Chr12 LOH in selected smooth and rough clones.** Panels **A** and **B** show the patterns of LOH found in independent 5-FOA resistant clones derived from JAY270 with hemizygous insertions of the CORE2 cassette (*κIURA3-scURA3-KanMX4*) either adjacent to the *ACE2* allele in the paternal Chr12 homolog (**A**) or adjacent to the *ace2-A7* allele in the maternal Chr12 homolog (**B**). The clones selected in **A** were *ace2-A7/ace2-A7* and had rough colonies, whereas the clones selected in **B** were *ACE2/ACE2* and had smooth colonies. The genotypes at the nine HetSNP loci at the indicated positions were determined by PCR-RFLP (Table S3). The clones showing continuous LOH tracts were grouped according to the breakpoint interval between HetSNPs, and the number of clones in each class is indicated to the

left. Panel **C** shows a representation of the breakpoint distribution among rough colony clones analyzed by WGS (detailed in Fig. S6). These clones were arranged in classes according to the PCR-RFLP HetSNPs marker positions used in **A** and **B**. The table in panel **D** shows the analysis of absolute interval size, relative size compared to the *CEN12* to *CORE2* distance, and the number and frequency of rough colony *ace2-A7/ace2-A7* (combined from **A** and **C**) for each Chr12 interval. Note that only clones with continuous tracts are shown in **A-C**. Four clones with complex tracts were omitted from this analysis. One of the clones analyzed by PCR-RFLP was discontinuous, and three clones had a gene conversion pattern not associated with a crossover (one from PCR-RFLP, and two from WGS [JAY663 and JAY1143]).



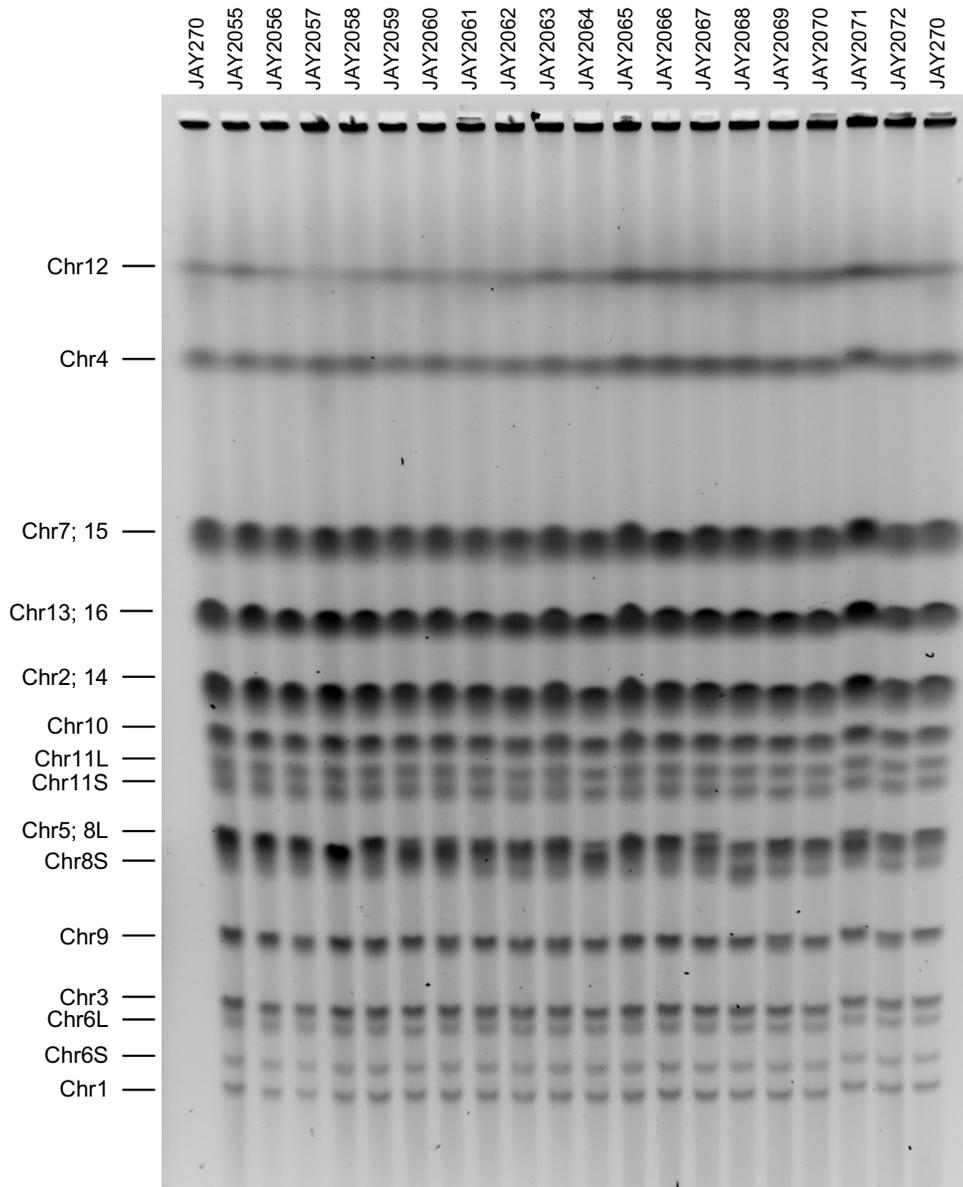
**Figure S2.6. Detailed maps of Chr12 LOH in the 25 sequenced spontaneous rough colony isolates.**

The top horizontal line is a depiction of Chr12 in JAY270 with HetSNPs represented as paternal (blue) and maternal (red) markers, and the position of the *ACE2* locus is shown. Each horizontal line below corresponds to Chr12 in each of the 25 spontaneous rough colony isolates sequenced. Only the markers that were homozygous P/P or M/M are shown (heterozygous markers were omitted to emphasize visualization of LOH tracts). As expected from selection for the rough colony morphology, all clones were homozygous for the maternal *ace2-A7* allele (red). In 3 cases, the tracts had complex discontinuities and some even showed limited LOH for the paternal allele near the breakpoint. Note the unselected LOH tract on the left arm in JAY664. A detailed view of the genotype for each HetSNP is available in Table S5. Plots were generated to scale in Python 2.7 using the matplotlib package and a custom script. For size reference, *CEN12* is at position 151 Kb, and the right-most marker is at position 450 Kb.



Independent smooth clones randomly picked after  
~57 cell generations

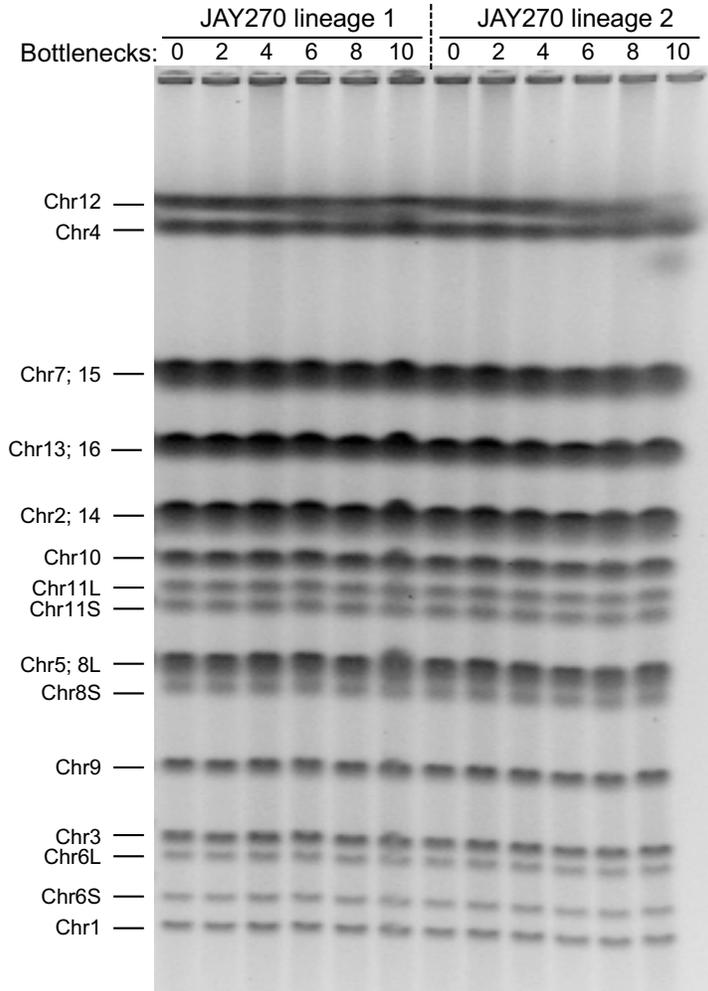
Figure S2.7., page 1 of 2.



Independent smooth clones randomly picked after  
~57 cell generations

**Figure S2.7. PFGE profiles of 29 independent smooth derivatives of JAY270.  
continued, page 2 of 2**

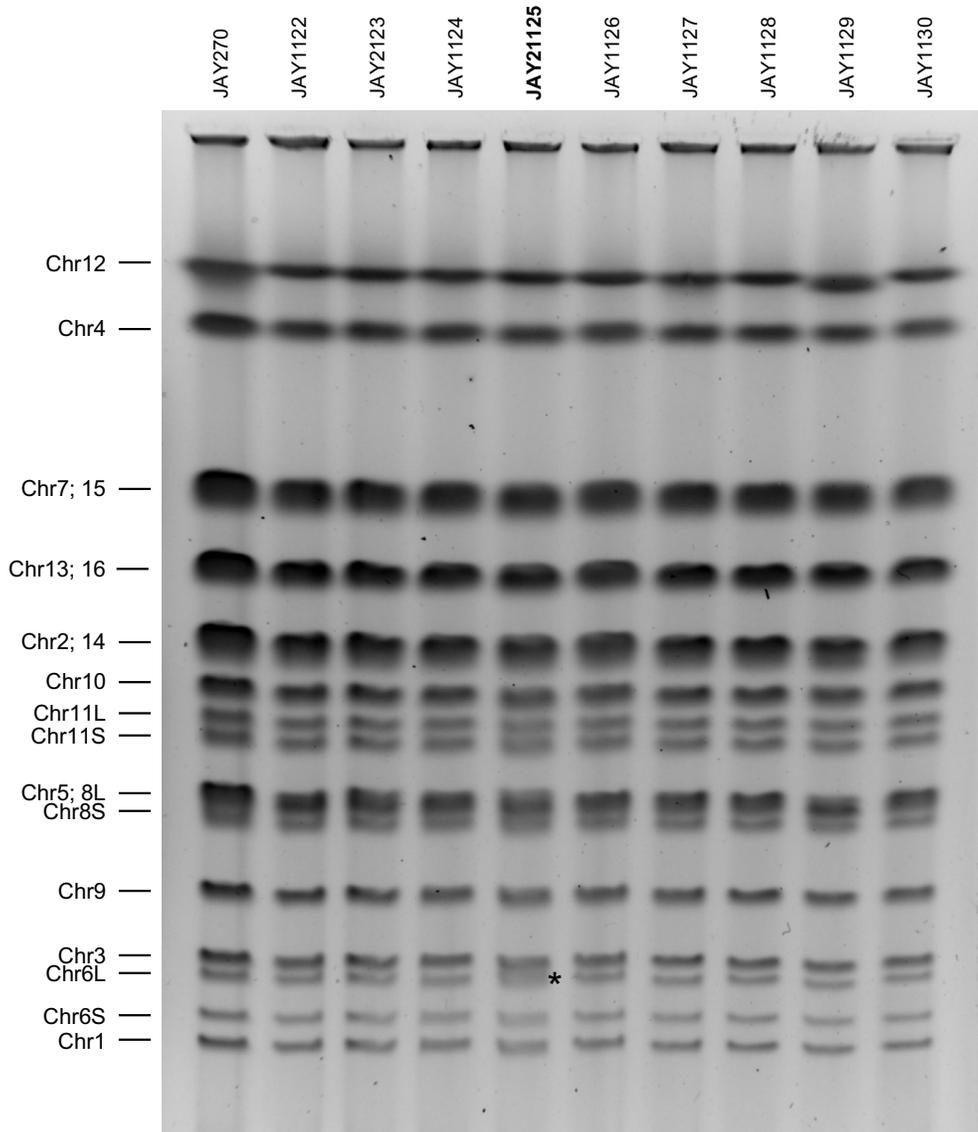
All clones were isolated after five transfer cycles in liquid culture without bottlenecks (~57 generations; Table S4; Methods). No gross chromosomal rearrangements were detected in any of the 29 smooth clones. Size polymorphisms in chromosomes 8 and 12 were not taken into account, as they are frequent due to contractions and expansions of the *CUP1* and rDNA repeats, respectively.



Two independent lineages of smooth clones passaged for 10 consecutive single-cell bottlenecks (~220 cell generations each lineage)

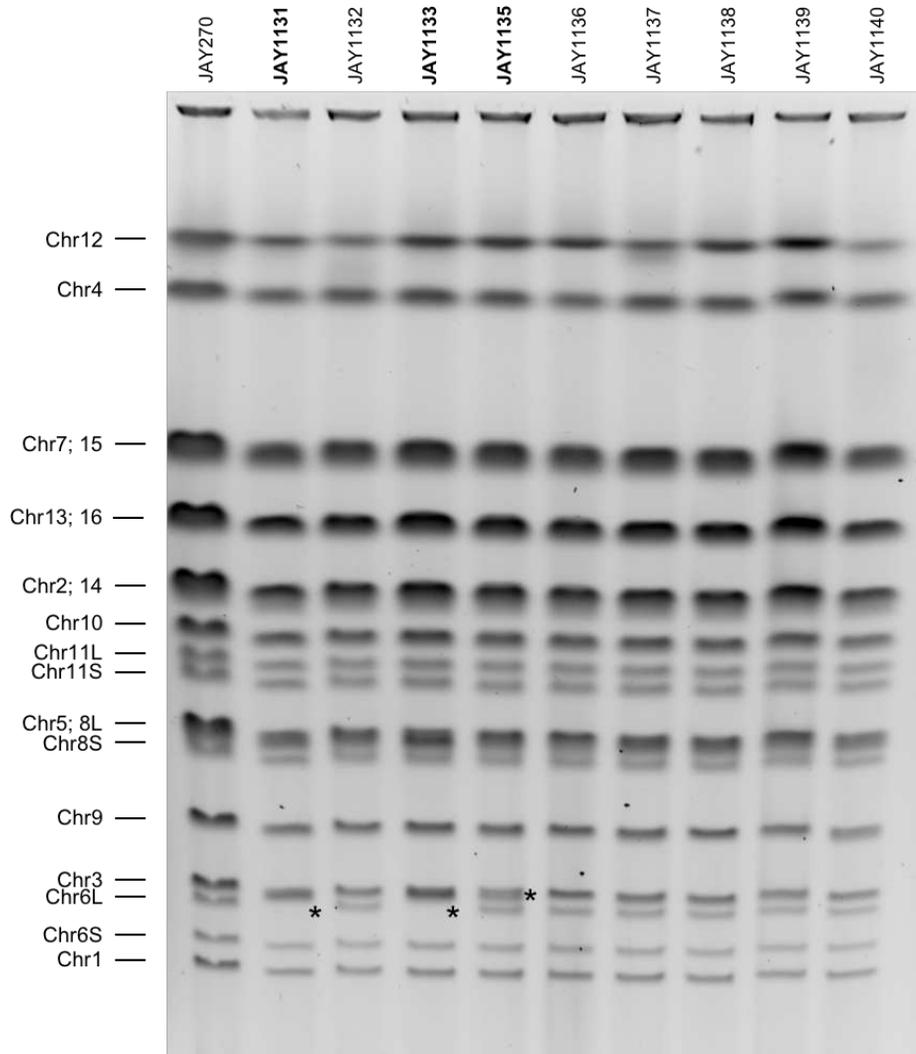
**Figure S2.8. PFGE profiles of smooth clones obtained through two bottlenecking lineages of JAY270.**

The gel shows the karyotypes of clones obtained along two independently passaged lineages starting from JAY270 (Table S4; Methods). The passage at which the intermediate clones were frozen is indicated at the top of each lane. Bottleneck 0 is JAY270 itself, and bottleneck 10 is the final clone from each lineage, which grew from an estimated ~220 cell generations. All clones were smooth, and no visible chromosome size polymorphisms were identified. Size polymorphisms in chromosomes 8 and 12 were not taken into account, as they are frequent due to contractions and expansions of the *CUP1* and rDNA repeats, respectively.



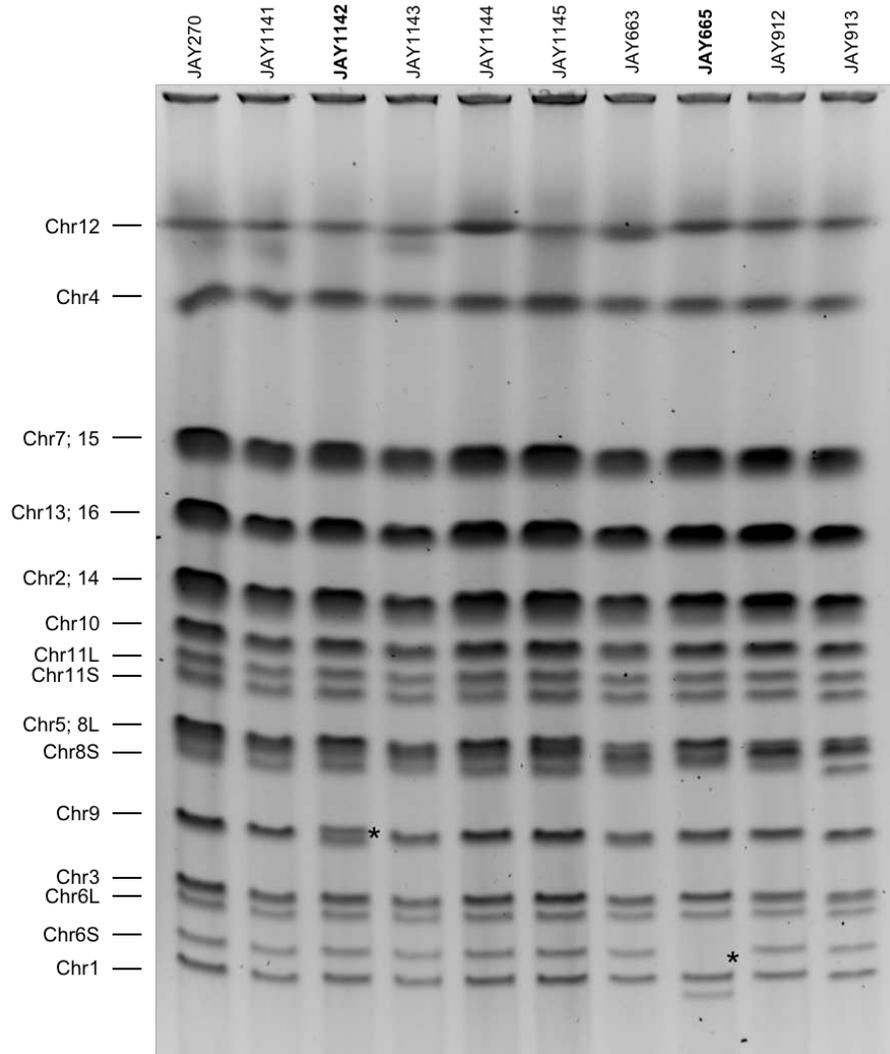
Independent rough clones selected at the first time they were detected. Detection of the first rough clone ended the passage for each specific culture. Most rough clones were identified after ~43 or less cell generations.

**Figure S2.9. page 1 of 3.**



Independent rough clones selected at the first time they were detected. Detection of the first rough clone ended the passage for each specific culture. Most rough clones were identified after ~57 or less cell generations.

**Figure S2.9. continued, page 2 of 3.**

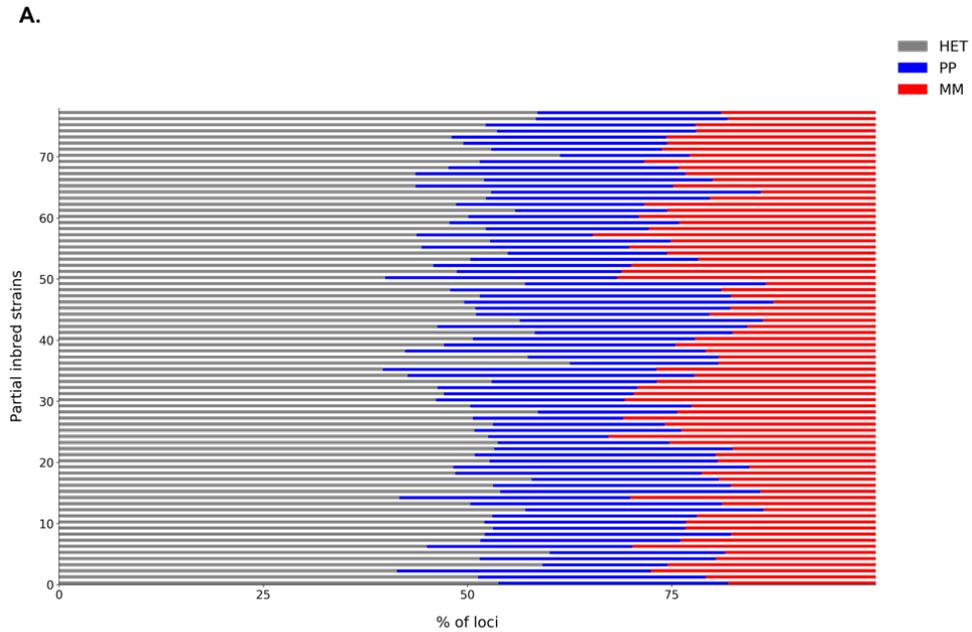


Independent rough clones selected at the first time they were detected. Detection of the first rough clone ended the passage for each specific culture. Most rough clones were identified after ~57 or less cell generations.

**Figure S2.9. PFGE profile of 27 independent rough colony derivatives of JAY270. continued, page 3 of 3.**

Rough colony clones were isolated during cycles of liquid growth without bottlenecks (most clones isolated at ~57 or less cell generations; Table S4; Methods). Size polymorphisms in chromosomes 8 and 12 were not taken into account, as they are frequent due to contractions and expansions of the *CUP1* and rDNA repeats, respectively. Clones in which unselected chromosomal rearrangements were detected are highlighted in bold and their altered chromosomes are marked with an asterisk symbol (\*). For consistency of the dataset, clones JAY1126, JAY1129 and JAY1133 were discarded from the overall analysis because their whole genome sequencing failed to yield usable data.

### S3



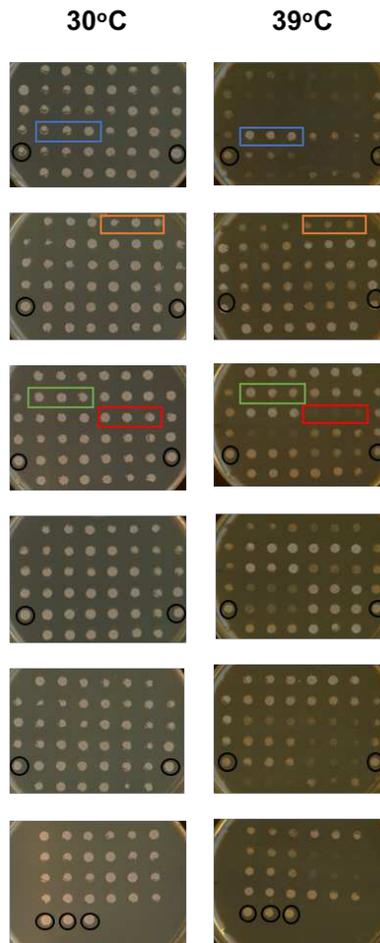
**B.**

	Genotype frequency (%)		
	MP	PP	MM
<b>Maximum</b>	62.52	37.95	34.65
<b>Minimum</b>	39.64	14.72	12.57
<b>Median</b>	51.49	25.98	22.68
<b>Average</b>	51.02	26.11	22.87

**Figure S3.1. Genotype distribution in the inbred collection.**

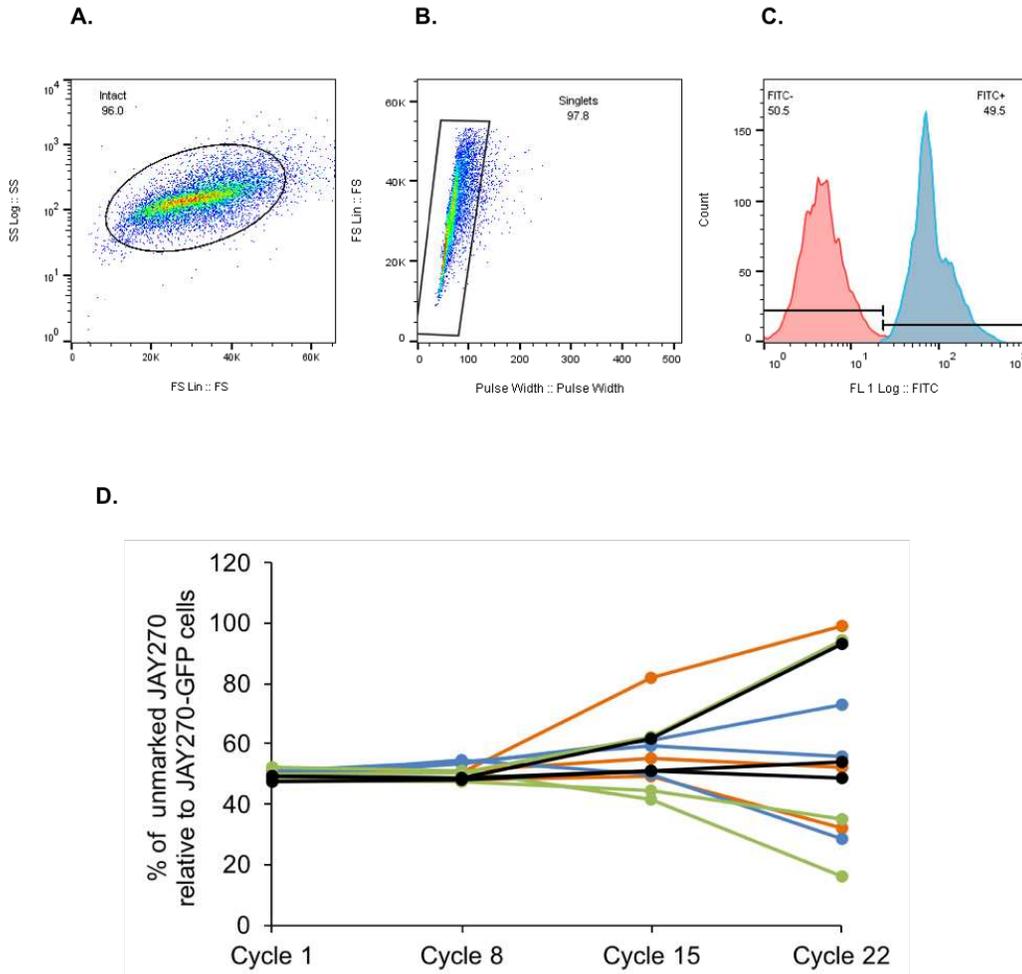
A) The genotypes of the inbred collection from Figure 1C are plotted as cumulative frequencies of genotypes for all HetSNP markers. Gray = heterozygous M/P. Blue = homozygous P/P. Red = homozygous M/M. B) Maximum, minimum, median and average frequency of genotypes that are heterozygous (M/P) and homozygous (M/M and P/P) found in the inbred strains in the collection.

### Temperature of incubation



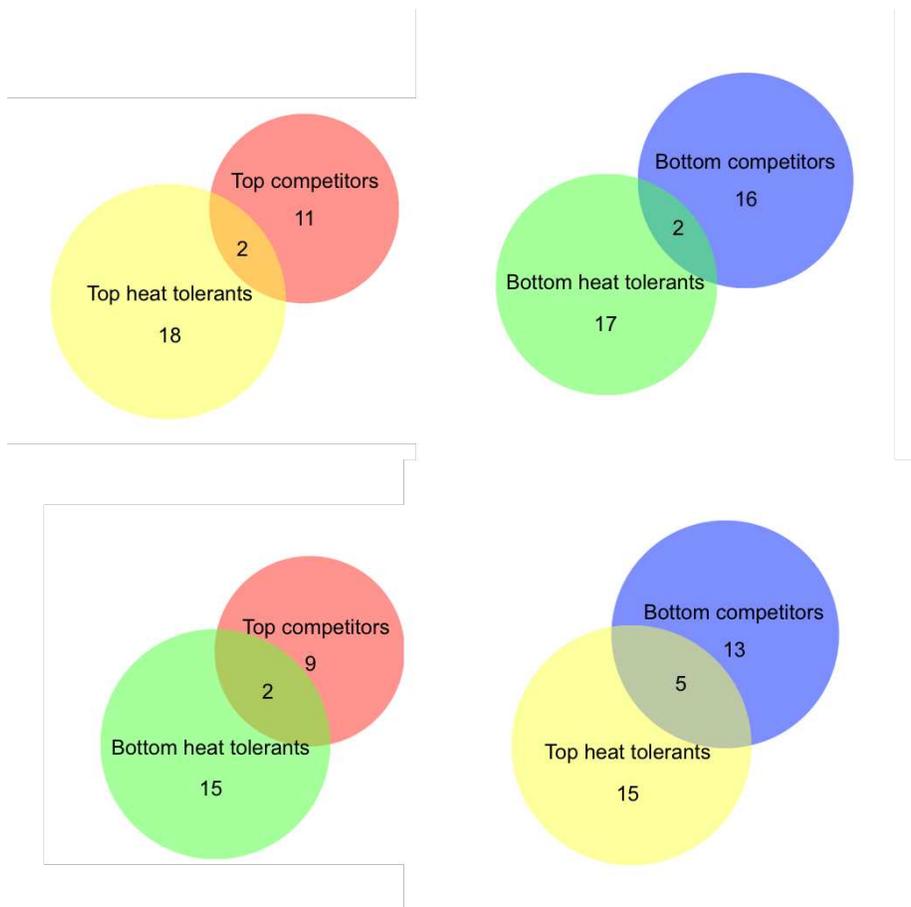
**Figure S3.2. Inbred strains show uniform growth at 30°C and wide variation in tolerance to high temperature stress at 39°C.**

A plate spotting assay was used for rapid phenotypic screening of the strain collection. Column to the left shows uniform growth among strains incubated for 24 hours at 30°C and column to the right shows variable growth at 39°C after four days of incubation. Black circles highlight JAY270, which shows an intermediate level of heat tolerance. Rectangles highlight the growth of representative strains shown in Figure 3.2 with different levels of temperature tolerance. Red = JAY1664, orange = JAY1628, green = JAY1658, blue = JAY1610.



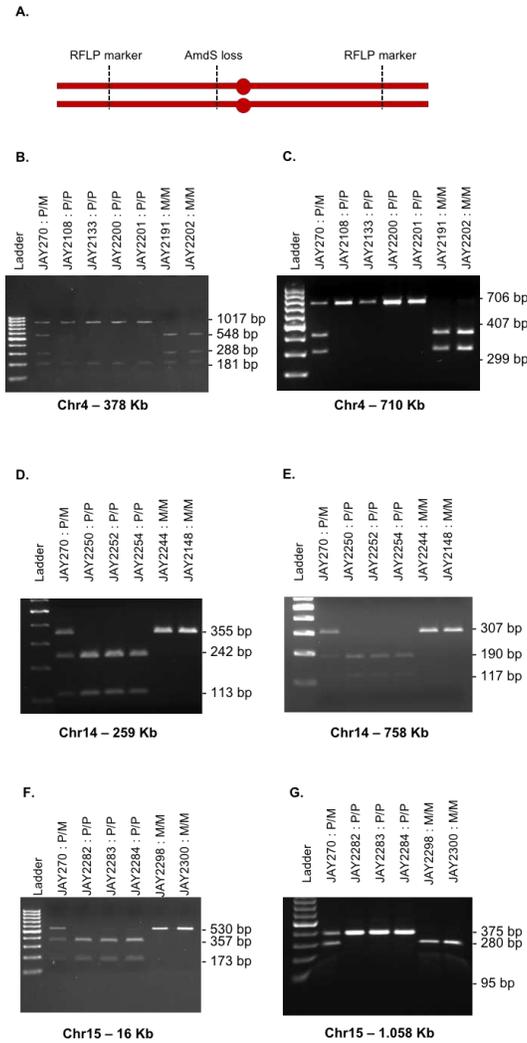
**Figure S3.3. Flow cytometry data analysis.**

A) False events corresponding to cell debris or other material were eliminated by gating intact cells in a side scatter by forward scatter plot. B) Potential cell aggregates were subsequently eliminated from our event counting by gating single cells in a forward scatter by pulse width plot. C) Final counts of GFP+ (JAY270 parent strain) and GFP- cells (partial inbred diploid strains) in each culture were determined by gating the FITC+ and FITC- population in a count by FITC signal plot. D) Competition assay between four independent GFP-labeled JAY270 derivatives and the unlabeled JAY270 parent over 22 cycles of co-culture. Each of the four GFP-tagged controls were co-cultured with the original JAY270 parent in triplicates shown in orange, blue, green and black curves. Phenotype of GFP-tagged strains was maintained until cycle 8 of competition. After that point the phenotypic variation between triplicates increased significantly likely due to the appearance of *de novo* mutations.



**Figure S3.4. Venn Diagrams.**

Venn diagrams displaying pairwise overlapping comparisons of inbred strains that displayed the most significant changes in the heat tolerance and growth kinetics phenotypes. Circles represent the group of inbred strains that showed the greatest improvement in growth kinetics (top competitors, red) and in heat tolerance (top heat tolerant, yellow) and the group of inbred strains that displayed substantial decrease in growth kinetics (bottom competitors, blue) and heat tolerance (bottom heat tolerant, green). Venn diagrams were generated using the BioVenn web application (Hulsen, de Vlieg, and Alkema 2008).



### Figure S3.5. Genotype confirmation of UPD strains.

A) To determine whether the UPD strains contained one fully homozygous chromosome, the genotypes at two HetSNP loci located near the last HetSNP chromosome arm were determined by PCR-RFLP. In addition, loss of the *AmdS* marker at the centromere region indicated LOH at the central chromosomal region. The possibility of LOH due to whole chromosome loss was eliminated by tetrad analysis and confirmation of four viable spores for all strains. Panels **B-G** show agarose gel electrophoresis images of PCR-RFLP genotyping of HetSNP markers. Panels B, D and F show genotypes of RFLP markers located at the left arm of chromosomes 4, 14 and 15, respectively. Panels C, E and G show genotypes of RFLP markers located at the right arm of chromosomes 4, 14 and 15, respectively. The approximate coordinates of each RFLP marker tested are indicated below each image. The corresponding genotype at each position is indicated for each UPD strain as P/P for homozygous paternal, M/M for homozygous maternal and P/M for heterozygous.

## APPENDIX B: SUPPLEMENTAL TABLES

**Table S2.1.** Yeast strains used in this study.

Strain	Relevant genotype <sup>1</sup>	Genetic background	Description	Source
<b>Diploids:</b>				
JAY270	<i>MATa/MAT<math>\alpha</math>, ACE2/ace2-A7</i>	JAY270	Representative single colony isolate from the PE-2 bioethanol production strain	Argueso <i>et al.</i> , 2009
JAY663	<i>MATa/MAT<math>\alpha</math>, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate	This study
JAY664	<i>MATa/MAT<math>\alpha</math>, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate	This study
JAY665	<i>MATa/MAT<math>\alpha</math>, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate	This study
JAY912	<i>MATa/MAT<math>\alpha</math>, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate	This study
JAY913	<i>MATa/MAT<math>\alpha</math>, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate	This study
JAY585	<i>MATa/MAT<math>\alpha</math>, ACE2/ace2-A7, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0</i>	JAY270	Ura <sup>-</sup> derivative of JAY270, also known as FGY50	Our strain collection
JAY1101, 1102, 1103	<i>MATa/MAT<math>\alpha</math>, ACE2::CORE2/ace2-A7, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0</i>	JAY270	Hemizygous CORE2 insertion at Chr12-P	This study
JAY1099, 1100, 1104	<i>MATa/MAT<math>\alpha</math>, ACE2/ace2-A7::CORE2, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0</i>	JAY270	Hemizygous CORE2 insertion at Chr12-M	This study
JAY865, 866	<i>MATa/MAT<math>\alpha</math>, ACE2/ace2-A7, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0, SSF2::CORE2/SSF2</i>	JAY270	Hemizygous CORE2 insertion at Chr4	This study
JAY868	<i>MATa/MAT<math>\alpha</math>, ACE2/ace2-A7, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0, ADH6::CORE2/ADH6</i>	JAY270	Hemizygous CORE2 insertion at Chr13	This study
JAY859, 860	<i>MATa/MAT<math>\alpha</math>, ACE2/ACE2, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0, ADH6::CORE2/ADH6</i>	CG379*	Hemizygous CORE2 insertion at Chr13	This study
JAY861, 862	<i>MATa/MAT<math>\alpha</math>, ACE2/ACE2, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0, SSF2::CORE2/SSF2</i>	CG379*	Hemizygous CORE2 insertion at Chr4	This study
JAY1105, 1106	<i>MATa/MAT<math>\alpha</math>, ACE2::CORE2/ACE2, ura3<math>\Delta</math>0/ura3<math>\Delta</math>0</i>	CG379*	Hemizygous CORE2 insertion at Chr12	This study
JAY1567, 1568	<i>MATa/MAT<math>\alpha</math>, SSF2/SSF2::CORE2, CAN1/can1D::Nat</i>	CG379*	Hemizygous CORE2 insertion at Chr4 and hemizygous <i>CAN1</i> at Chr5	This study
JAY1569, 1570	<i>MATa/MAT<math>\alpha</math>, ADH6/ADH6::CORE2, CAN1/can1D::Nat</i>	CG379*	Hemizygous CORE2 insertion at Chr13 and hemizygous <i>CAN1</i> at Chr5	This study
JAY1804, 1805	<i>MATa/MAT<math>\alpha</math>, SSF2/SSF2::CORE2, CAN1/can1D::Nat</i>	JAY270	Hemizygous CORE2 insertion at Chr4 and hemizygous <i>CAN1</i> at Chr5 in FGY50 strain	This study
JAY1812	<i>MATa/MAT<math>\alpha</math>, ADH6/ADH6::CORE2, CAN1/can1D::Nat</i>	JAY270	Hemizygous CORE2 insertion at Chr13 and hemizygous <i>CAN1</i> at Chr5 in FGY50 strain	This study
JAY1808	<i>MATa/<math>\Delta</math>MAT<math>\alpha</math>, SSF2/SSF2::CORE2, CAN1/can1D::Nat</i>	CG379	JAY1567 with deletion of <i>MAT<math>\alpha</math></i> allele.	This study
JAY1809	<i>MATa/<math>\Delta</math>MAT<math>\alpha</math>, ADH6/ADH6::CORE2, CAN1/can1D::Nat</i>	CG379	JAY1569 with deletion of <i>MAT<math>\alpha</math></i> allele.	This study

**Table S2.1. (continued).** Yeast strains used in this study.

Strain	Relevant genotype <sup>1</sup>	Genetic background	Description	Source
JAY1810, 1811	<i>MATa/ΔMATα, MAL13::CORE2/MAL13, CAN1/can1D::Nat</i>	CG379	JAY1571 with deletion of MATα allele.	This study
JAY1122	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1123	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1124	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1125	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1126	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1127	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1128	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1129	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1130	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1131	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1132	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1133	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1135	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1136	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1137	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1138	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1139	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1140	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1141	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1142	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1143	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1144	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY1145	<i>MATa/MATα, ace2-A7/ace2-A7</i>	JAY270	Spontaneous rough colony isolate from JAY270	This study
JAY2055	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2056	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2057	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2058	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2059	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2060	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2061	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2062	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2063	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2064	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2065	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2066	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2067	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study
JAY2068	<i>MATa/MATα, WT</i>	JAY270	Smooth colony isolate from JAY270	This study

**Table S2.1. (continued).** Yeast strains used in this study.

Strain	Relevant genotype <sup>1</sup>	Genetic background	Description	Source
JAY2069	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2070	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2071	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2072	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2073	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2074	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2075	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2076	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2077	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2078	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2079	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2080	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2081	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2082	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY2083	<i>MATa/MAT<math>\alpha</math></i> , WT	JAY270	Smooth colony isolate from JAY270	This study
JAY289	<i>MATa</i> , <i>ACE2</i>	JAY270	Sibling spores from a JAY270 tetrad	Argueso <i>et al.</i> , 2009
JAY290	<i>MAT<math>\alpha</math></i> , <i>ACE2</i>	JAY270		
JAY291	<i>MATa</i> , <i>ace2-A7</i>	JAY270		
JAY292	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270		
JAY293	<i>MAT<math>\alpha</math></i> , <i>ACE2</i>	JAY270	Sibling spores from a JAY270 tetrad	Our strain collection
JAY294	<i>MAT<math>\alpha</math></i> , <i>ACE2</i>	JAY270		
JAY295	<i>MATa</i> , <i>ace2-A7</i>	JAY270		
JAY296	<i>MATa</i> , <i>ace2-A7</i>	JAY270		
JAY325	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270	Sibling spores from a JAY270 tetrad	Our strain collection
JAY326	<i>MATa</i> , <i>ace2-A7</i>	JAY270		
JAY327	<i>MATa</i> , <i>ACE2</i>	JAY270		
JAY328	<i>MAT<math>\alpha</math></i> , <i>ACE2</i>	JAY270		
JAY2176	<i>MATa</i> , <i>ace2-A7</i>	JAY270	Sibling spores from a JAY663 tetrad	This study
JAY2177	<i>MATa</i> , <i>ace2-A7</i>	JAY270		
JAY2178	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270		
JAY2179	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270		
JAY2180	<i>MATa</i> , <i>ace2-A7</i>	JAY270	Sibling spores from a JAY663 tetrad	This study
JAY2181	<i>MATa</i> , <i>ace2-A7</i>	JAY270		
JAY2182	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270		
JAY2183	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270		
JAY2184	<i>MATa</i> , <i>ace2-A7</i>	JAY270	Sibling spores from a JAY663 tetrad	This study
JAY2185	<i>MATa</i> , <i>ace2-A7</i>	JAY270		
JAY2186	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270		
JAY2187	<i>MAT<math>\alpha</math></i> , <i>ace2-A7</i>	JAY270		
JAY1039	<i>MAT<math>\alpha</math></i> , <i>ACE2</i> , <i>ura3-154</i> , <i>CEN5::HphMX4</i>	JAY270	<i>ACE2</i> allele replacement in JAY292	This study
JAY1051	<i>MATa</i> , <i>ACE2</i> , <i>ura3-W256G</i> , <i>CEN5::HphMX4</i>	JAY270	<i>ACE2</i> allele replacement in JAY291	This study

**Table S2.2.** Oligonucleotides used in this study.

Name	5'-3' sequence <sup>1</sup>	Description
JAO904	ACACTCAAGATGAGGAGTAT	ACE2 primer Fwd, pAR1 construction
JAO905	TCGCATGAATACGGTCTATC	ACE2 primer Rev, pAR1 construction
JAO906	TACTTCAATAAATGGTTTAC	ACE2 primer Fwd, sequencing of A8/A7 tract
JAO907	TGATATTGTCGAGACCGTGG	ACE2 primer Rev, sequencing of A8/A7 tract
JAO912	TAGCACAAAGATATGAACT	Chr12 409,200 distal SNP sequencing Fwd
JAO913	TATATCATCTACAATGAC	Chr12 409,200 distal SNP sequencing Rev
JAO944	ATACTGGAGAGTGTGGG	Chr12 400,451 proximal SNP sequencing Fwd
JAO945	TCTAGCGACCAAATGGCC	Chr12 400,451 proximal SNP sequencing Rev
JAO1077	CCATTCTTTATCCGCATTC	Chr12 14,795 SNP HhaI RFLP Fwd
JAO1078	CACTTGCTACGCTCCTTA	Chr12 14,795 SNP HhaI RFLP Rev
JAO1079	GCCTCGTCACAGTTTTTT	Chr12 85,625 SNP HhaI RFLP Fwd
JAO1080	CGTTTGGATTTGCCCTTT	Chr12 85,625 SNP HhaI RFLP Rev
JAO1081	TCCTCGCTGTTTCGTTTT	Chr12 148,252 SNP Apol RFLP Fwd
JAO1082	AGGGGTACTGGTATTGTC	Chr12 148,252 SNP Apol RFLP Rev
JAO1083	CGGAATGAGACACTGTTA	Chr12 227,328 SNP HinfI RFLP Fwd
JAO1084	GTTGAAGTGAAGAAGGGA	Chr12 227,328 SNP HinfI RFLP Rev
JAO1085	CGACAAACCGTGGTACAA	Chr12 292,284 SNP XbaI RFLP Fwd
JAO1086	CGAACATTCTCACTCCAT	Chr12 292,284 SNP XbaI RFLP Rev
JAO1087	CACAGTCAAGAGAAAGAACA	Chr12 364,388 SNP HhaI RFLP Fwd
JAO1088	GCACACACAAAAGGAACATA	Chr12 364,388 SNP HhaI RFLP Rev
JAO1089	CACTGAACACCAACATCT	Chr12 392,257 SNP HhaI RFLP Fwd
JAO1090	TCCTCAACAAGCAAGCAA	Chr12 392,257 SNP HhaI RFLP Rev
JAO1091	CCTACGTAAAAAGATGACC	Chr12 419,261 SNP HhaI RFLP Fwd
JAO1092	CATTACCACAACAGATCC	Chr12 419,261 SNP HhaI RFLP Rev
JAO1093	CGTGTGGATGATCTGATT	Chr12 450,044 SNP Apol RFLP Fwd
JAO1094	GAAAAGAAATGACTACGGTG	Chr12 450,044 SNP Apol RFLP Rev
JAO14	AGGAGGGTATTCTGGGCCCTCCATG	Inside Mx4 region Fwd
JAO15	ATGCGAAGTTAAGTGGCAGAAAG	Inside Mx4 region Rev
JAO1073	GGCAATGTACCCTAAAGGTTGTG	Proximal of ACE2::CORE2 insertion Fwd
JAO1074	<b>CTTTAGGGTTATGTCCTTATAAACGATGACTATTGCCTTTTTTGGCCCTTAAGACTACAGTGTACGTAATC</b> <i>GTCGTTTTTCGACACTGG</i>	Insertion of CORE2 proximal of ACE2 Fwd
JAO1075	<b>GAACATTTATCTATGCATGATATTAACATAATAAATAATAGTAACAATAATATAATACATTTATTTCTTTAC</b> <i>AGTTGATCCATTGTGTGC</i>	Insertion of CORE2 proximal of ACE2 Rev
JAO957	AGCGTACCAAAAAGAGAAT	Inside KIURA3 Fwd
JAO1076	GATCAACAAGAACTTACATCTCCC	Distal of ACE2::CORE2 insertion Rev
JAO501	GTCCGATACCCTATGAACGTG	Proximal of ADH6::CORE2 insertion Fwd
JAO502	<b>TTGTTAGTGTATTGATATGTGTTTCTTTTCCACCTTAAAGGTGCTTAGCAAGGAG</b> <i>CCTTACCATTAAGTTGATC</i>	Insertion of CORE2 proximal of Chr13 ADH6 Fwd
JAO503	<b>TTTTTATGATTATAAGGTACTATTAAATATTTACAACCTCGTACAGTTCTC</b> <i>GAGCTCGTTTTTCGACACTGG</i>	Insertion of CORE2 proximal of Chr13 ADH6 Rev
JAO504	GGTCTGTATATAGGAGTGCTG	Distal of ADH6::CORE2 insertion Rev
JAO505	GTGACTTATTCAGTGAAGTAG	Proximal of SSF2::CORE2 insertion Fwd
JAO506	<b>CTCCGTACGTAACATCACTATCCATATAGTAGCCATGACTCCGATGGAC</b> <i>CCTTACCATTAAGTTGATC</i>	Insertion of CORE2 distal of Chr4 SSF2 Fwd
JAO507	<b>TTGAGGTGTTCCCTCACCTATGAATAAACAGACACTTCTGTTCTTTAA</b> <i>GAGCTCGTTTTTCGACACTGG</i>	Insertion of CORE2 distal of Chr4 SSF2 Rev
JAO508	TTTGTCTTTCCATGATGCCG	Distal of SSF2::CORE2 insertion Rev
JAO611	AGGAGGCAAGATATTATTGTC	Proximal of MAL13::CORE2 insertion Fwd

**Table S2.2. (continued).** Oligonucleotides used in this study.

Name	5'-3' sequence <sup>1</sup>	Description
JAO612	<b>AGTTACTAGATACGGACATATCTCTAGGA</b> ACTATGAAGGCTG <i>GAGCTCGTTTTTCGACACT</i>	Insertion of CORE2 distal of Chr7 <i>MAL13</i> Fwd
JAO613	<b>TTAGGAAGGAAATGAATTAAGCTACGCAGAAAGGACATCTCTT</b> <i>CCTTACCATTAAGTGA</i>	Insertion of CORE2 distal of Chr7 <i>MAL13</i> Rev
JAO614	CCTGCCCAAGGCGAGGTGCAG	Distal of <i>MAL13::CORE2</i> insertion Rev
JAO1591	TGCCAAGATGCGAGTATGT	Chr6 R 185,546 SNP <i>KpnI</i> RFLP marker Fwd
JAO1592	TGTTGGGAAAATGCTTGCT	Chr6 R 185,546 SNP <i>KpnI</i> RFLP marker Rev
JAO1115	AGCAGATGCTCAAACGCGGCGCTGA	Chr6 R 229,586 SNP <i>NdeI</i> RFLP marker Fwd
JAO1116	GCGTTAGCCACTTCATTAGATCG	Chr6 R 229,586 SNP <i>NdeI</i> RFLP marker Rev
JAO1134	CAATGTTGGCCAAACCGGGTAACATG	Chr11 571,243 SNP <i>EcoRV</i> RFLP marker Fwd
JAO1135	GATGAAGGCAATGTCACAAAGTCTCG	Chr11 571,243 SNP <i>EcoRV</i> RFLP marker Rev
JAO1117	CTCTTACTCTGTGAGTAGTTG	Chr11 639,322 SNP <i>Asel</i> RFLP marker Fwd
JAO1118	TAAATACTATGTGCCAGCATACC	Chr11 639,322 SNP <i>Asel</i> RFLP marker Rev
JAO1119	TGTGCTAGGCGAGAATATATCGAG	Chr11 653 Kb hemizygous marker Fwd
JAO1120	CCACCGAAATGACTGGCTTGC	Chr11 653 Kb hemizygous marker Rev
JAO1593	GTGGACGAGAAAACCGTGTGA	Chr7 17,017 SNP <i>BglII</i> RFLP marker Fwd
JAO1594	TGTGTCATTCACATGCGCATAT	Chr7 17,017 SNP <i>BglII</i> RFLP marker Rev
JAO1599	TCTCATCTTCTTTCCCGT	Chr7 484,135 SNP <i>HindIII</i> RFLP marker Fwd
JAO1600	GATTTTCATCCTAGCTGC	Chr7 484,135 SNP <i>HindIII</i> RFLP marker Rev
JAO1140	GAGCTTCGAAACTTCTGGCAGG	Chr7 1,005,631 SNP <i>NdeI</i> RFLP marker Fwd
JAO1141	CAACCGGATTGGGCCTTAGTAAC	Chr7 1,005,631 SNP <i>NdeI</i> RFLP marker Rev
JAO611	AGGAGGCAAGATATTATTGTC	Chr7 1,072 Kb hemizygous marker Fwd
JAO614	CCTGCCCAAGGCGAGGTGCAG	Chr7 1,072 Kb hemizygous marker Rev
JAO271	<b>GCGAAATGGCGTGAAATGTGATCAAAGGTAATAAAACGTCATAT</b> <i>AATTAAGGCGCCAGATCTG</i>	To replace <i>CAN1</i> with <i>NAT</i> Fwd
JAO272	<b>ATCGAAAGTTTATTTCAGAGTTCTTCAGACTTCTTAACCTCTGTA</b> <i>GCATAGGCCACTAGTGGAT</i>	To replace <i>CAN1</i> with <i>NAT</i> Rev
JAO1438	GATTATAGTAAGCTCATTGATCC	Upstream to <i>CAN1</i> Fwd
JAO1439	GAACAGAGTAAACCGAATCAGG	Downstream to <i>CAN1</i> Rev
JAO1440	<b>GCGAGATAAACTGGTATTCTTCATTAGATTCTCTAGGCCCTTGGTATCTAGATATGGGT</b> <i>TTCGTACGCTGCAGGTCGAC</i>	To replace <i>MATα</i> allele with Hyg Fwd
JAO1441	<b>TCCCATATFCCGTGCTGCATTTTGTCCGCGTGCCATTCTTCAGCGAGCAGAGAAGACAAG</b> <i>CGAGTCAGTGAGCGAGGAAG</i>	To replace <i>MATα</i> allele with Hyg Rev
JAO1442	AAGAGGTCGCTAATTCTGGAG	<i>MAT</i> locus Fwd
JAO1371	AGAACAAGAAGGATGCTAAG	<i>MAT</i> locus Rev

1. In the case of long oligonucleotides used for PCR-based integrations, the nucleotides in the 5' end (**bold**) are targeting tails with homology to the respective chromosomal insertion sites; the 3' end nucleotides (*italicized*) correspond to primer sequences used to amplify the selectable marker sequence from the specific template plasmids.

**Table S2.3.** List of phased JAY270 HetSNPs and hemizygous sequences and respective detection methods.

Marker SGD coordinates		PCR primers	Polymorphism detection method	SGD Watson bases 1	JAY270 Watson phased bases	
Chromosome	Nucleotide				Chr12-P	Chr12-M
Chr12	14,795	JAO1077 + JAO1078	<i>HhaI</i>	C	A	C
Chr12	85,625	JAO1079 + JAO1080	<i>HhaI</i>	G	A	G
Chr12	148,252	JAO1081 + JAO1082	<i>ApoI</i>	A	A	G
Chr12	227,328	JAO1083 + JAO1084	<i>HinfI</i>	C	T	C
Chr12	292,284	JAO1085 + JAO1086	<i>XbaI</i>	T	C	T
Chr12	364,388	JAO1087 + JAO1088	<i>HhaI</i>	G	A	G
Chr12	392,257	JAO1089 + JAO1090	<i>HhaI</i>	G	A	G
Chr12	400,451	JAO944 + JAO945	Sanger seq	G	A	G
Chr12	405,711 to 405,718	JAO906 + JAO907	Sanger seq	8x T	8x T	7x T
Chr12	409,200	JAO912 + JAO913	Sanger seq	C	T	C
Chr12	419,261	JAO1091 + JAO1092	<i>HhaI</i>	T	T	C
Chr12	450,044	JAO1093 + JAO1094	<i>ApoI</i>	C	C	T
Chr06	185,546	JAO1591 + JAO1592	<i>KpnI</i>	G	G	A
Chr06	229,586	JAO1115 + JAO1116	<i>NdeI</i>	C	C	G
Chr07	17,017	JAO1593 + JAO1594	<i>BglII</i>	C	T	C
Chr07	484,135	JAO1599 + JAO1600	<i>HindIII</i>	C	C	G
Chr07	1,005,631	JAO1140 + JAO1141	<i>NdeI</i>	A	G	A
Chr07	1,072,000	JAO611 + JAO614	none	present	absent	present
Chr11	571,243	JAO1134 + JAO1135	<i>EcoRV</i>	T	A	T
Chr11	639,322	JAO1117 + JAO1118	<i>Asel</i>	C	A	C
Chr11	653,131	JAO1119 + JAO1120	none	present	present	absent

1. Nucleotides that are underlined indicate the sequences that are cut by the respective restriction endonucleases, whereas lack of underlining corresponds to no cut site.

**Table S2.4.** Smooth and rough clones isolation specifics and analysis of chromosome size polymorphisms by PFGE.

<b>Clone ID</b>	<b>Colony phenotype</b>	<b>Selected on cycle</b>	<b>Estimated generations</b>	<b>PFGE polymorphisms</b>
JAY2055	Smooth	5	57	0
JAY2056	Smooth	5	57	0
JAY2057	Smooth	5	57	0
JAY2058	Smooth	5	57	0
JAY2059	Smooth	5	57	0
JAY2060	Smooth	5	57	0
JAY2061	Smooth	5	57	0
JAY2062	Smooth	5	57	0
JAY2063	Smooth	5	57	0
JAY2064	Smooth	5	57	0
JAY2065	Smooth	5	57	0
JAY2066	Smooth	5	57	0
JAY2067	Smooth	5	57	0
JAY2068	Smooth	5	57	0
JAY2069	Smooth	5	57	0
JAY2070	Smooth	5	57	0
JAY2071	Smooth	5	57	0
JAY2072	Smooth	5	57	0
JAY2073	Smooth	5	57	0
JAY2074	Smooth	5	57	0
JAY2075	Smooth	5	57	0
JAY2076	Smooth	5	57	0
JAY2077	Smooth	5	57	0
JAY2078	Smooth	5	57	0
JAY2079	Smooth	5	57	0
JAY2080	Smooth	5	57	0
JAY2081	Smooth	5	57	0
JAY2082	Smooth	5	57	0

**Table S2.4. (continued).** Smooth and rough clones isolation specifics and analysis of chromosome size polymorphisms by PFGE.

Clone ID	Colony phenotype	Selected on cycle	Estimated generations	PFGE polymorphisms
JAY2083	Smooth	5	57	0
JAY2084	Smooth	5	57	0
Bottlenecks 1	Smooth	10	220	0
Bottlenecks 2	Smooth	10	220	0
JAY663	Rough	Unknown	Unknown	0
JAY664	Rough	Unknown	Unknown	2
JAY665	Rough	Unknown	Unknown	1
JAY912	Rough	Unknown	Unknown	0
JAY913	Rough	Unknown	Unknown	0
JAY1122	Rough	1	29	0
JAY1123	Rough	3	43	0
JAY1124	Rough	3	43	0
JAY1125	Rough	5	57	1
JAY1127	Rough	9	85	0
JAY1128	Rough	2	36	0
JAY1130	Rough	4	50	0
JAY1131	Rough	4	50	1
JAY1132	Rough	4	50	0
JAY1135	Rough	4	50	1
JAY1136	Rough	5	57	0
JAY1137	Rough	5	57	0
JAY1138	Rough	3	43	0
JAY1139	Rough	3	43	0
JAY1140	Rough	3	43	0
JAY1141	Rough	3	43	0

**Table S2.4. (continued).** Smooth and rough clones isolation specifics and analysis of chromosome size polymorphisms by PFGE.

Clone ID	Colony phenotype	Selected on cycle	Estimated generations	PFGE polymorphisms
JAY1142	Rough	4	50	1
JAY1143	Rough	1	29	0
JAY1144	Rough	2	36	0
JAY1145	Rough	2	36	0

PFGE polymorphisms indicates the number of bands with altered size in each clone, excluding changes in Chr8 and Chr12.

**Table S2.5.** Summary of WGS analysis of rough colony isolates.

Rough clone ID	Generations at rough selection	Number of Point mutations	Primary selected and Secondary LOH tracts	Tract size (Kb)	Terminal or Interstitial?	Continuous or Interrupted?	Unidirectional or Bidirectional?
JAY663	unknown	2	Primary: Chr12	4.4	I	C	U
			Secondary: Chr04	1063.9	T	I	B
			Secondary: Chr13	272.7	T	I	B
JAY664	unknown	4	Primary: Chr12	859.9 + rDNA	T	C	U
			Secondary: Chr02	20.0	I	C	U
			Secondary: Chr06	78.6	T	C	U
			Secondary: Chr07	310.1	T	C	U
			Secondary: Chr11	85.1	T	C	U
			Secondary: Chr12	21.8	T	I	B
			Secondary: Chr14	4.6	I	C	U
			Secondary: Chr15	5.8	I	C	U
JAY665	unknown	0	Primary: Chr12	845.6 + rDNA	T	I	B
			Secondary: Chr03	20.8	I	C	U
			Secondary: Chr06	60.6	T	C	U
JAY912	unknown	1	Primary: Chr12	818.9 + rDNA	T	C	U
JAY913	unknown	0	Primary: Chr12	706.6 + rDNA	T	C	U
JAY1122	29	0	Primary: Chr12	777.7 + rDNA	T	C	U
			Secondary: Chr15	5.8	I	C	U
JAY1123	43	0	Primary: Chr12	912.3 + rDNA	T	C	U
JAY1124	43	3	Primary: Chr12	706.6 + rDNA	T	C	U
JAY1125	57	1	Primary: Chr12	814.0 + rDNA	T	C	U
			Secondary: Chr15	0.5	I	C	U
JAY1127	85	0	Primary: Chr12	873.4 + rDNA	T	I	U
			Secondary: Chr13	3.4	I	C	U
JAY1128	36	0	Primary: Chr12	681.2 + rDNA	T	C	U

**Table S2.5. (continued).** Summary of WGS analysis of rough colony isolates.

Rough clone ID	Generations at rough selection	Number of Point mutations	Primary selected and Secondary LOH tracts	Tract size (Kb)	Terminal or Interstitial?	Continuous or Interrupted?	Unidirectional or Bidirectional?
JAY1130	50	0	Primary: Chr12	900.0 + rDNA	T	C	U
			Secondary: Chr04	12.1	I	I	U
			Secondary: Chr10	373.5	T	C	U
JAY1132	50	1	Primary: Chr12	777.7 + rDNA	T	C	U
JAY1131	50	0	Primary: Chr12	814.0 + rDNA	T	C	U
			Secondary: Chr16	308.8	T	C	U
JAY1135	50	0	Primary: Chr12	873.4 + rDNA	T	C	U
JAY1136	57	0	Primary: Chr12	912.3 + rDNA	T	C	U
JAY1137	57	0	Primary: Chr12	838.8 + rDNA	T	C	U
JAY1138	43	0	Primary: Chr12	818.9 + rDNA	T	C	U
JAY1139	43	0	Primary: Chr12	747.2 + rDNA	T	C	U
JAY1140	43	0	Primary: Chr12	777.7 + rDNA	T	C	U
			Secondary: Chr16	10.4	I	C	U
JAY1141	43	1	Primary: Chr12	717.2 + rDNA	T	C	U
			Secondary: Chr04	6.1	I	C	U
			Secondary: Chr06	5.0	I	C	U
			Secondary: Chr10	209.6	T	C	U
JAY1142	50	0	Primary: Chr12	870.9 + rDNA	T	C	U
			Secondary: Chr15	26.7	T	C	U
JAY1143	29	1	Primary: Chr12	11.1	I	C	U
			Secondary: Chr04	0.9	I	C	U
			Secondary: Chr04	4.1	I	C	U
			Secondary: Chr14	1.6	I	C	U
JAY1144	36	0	Primary: Chr12	912.3 + rDNA	T	C	U

**Table S2.5. (continued).** Summary of WGS analysis of rough colony isolates.

Rough clone ID	Generations at rough selection	Number of Point mutations	Primary selected and Secondary LOH tracts	Tract size (Kb)	Terminal or Interstitial?	Continuous or Interrupted?	Unidirectional or Bidirectional?
JAY1145	36	1	Primary: Chr12	908.4 + rDNA	T	I	B
			Secondary: Chr04	3.9	I	C	U
			Secondary: Chr15	254.7	T	C	U

None of the LOH tracts above crosses the centromere of the respective chromosomes.

The median size of the 15 interstitial secondary tracts was 5.0 Kb; The median size of the 12 terminal secondary tracts was 232.1 Kb.

**Table S3.1.** Yeast strains used in this study.

Strain	Mating type	Relevant genotype	Description	Source
<b>Haploids:</b>				
JAY1520	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1521	<i>MATa</i>	<i>ace2-A7</i>		
JAY1525	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1522	<i>MATa</i>	<i>ace2-A7</i>		
JAY1526	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1529	<i>MATa</i>	<i>ace2-A7</i>		
JAY1531	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1533	<i>MATa</i>	<i>ace2-A7</i>		
JAY1536	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1535	<i>MATa</i>	<i>ace2-A7</i>		
JAY1539	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1540	<i>MATa</i>	<i>ace2-A7</i>		
JAY1545	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1543	<i>MATa</i>	<i>ace2-A7</i>		
JAY1546	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1549	<i>MATa</i>	<i>ace2-A7</i>		
JAY1553	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1552	<i>MATa</i>	<i>ace2-A7</i>		
JAY1557	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	(Rodrigues-Prause et al., 2018)
JAY1554	<i>MATa</i>	<i>ace2-A7</i>		
JAY290	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	Our strain collection
JAY291	<i>MATa</i>	<i>ace2-A7</i>		
JAY293	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	Our strain collection
JAY295	<i>MATa</i>	<i>ace2-A7</i>		
JAY328	<i>MATa</i>	<i>ACE2</i>	Sibling spores from a JAY270 tetrad	Our strain collection
JAY326	<i>MATa</i>	<i>ace2-A7</i>		
<b>Diploids:</b>				
JAY270	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Sugarcane bioethanol fermentation strain	Argueso et al., 2009
JAY2208 (JAY270-GFP)	<i>MATa/MATa</i>	<i>ACE2/ace2-A7; CEN5::GFP-KanMX</i>	Hemizygous <i>GFP-KanMX</i> insertion near Chr5	This study
JAY1598	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1525	This study
JAY1600	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1526	This study
JAY1602	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1531	This study
JAY1604	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1536	This study
JAY1606	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1539	This study
JAY1608	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1545	This study
JAY1610	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1546	This study
JAY1612	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1553	This study
JAY1614	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY1557	This study
JAY1616	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY290	This study
JAY1618	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY293	This study

**Table S3.1. (continued).** Yeast strains used in this study.

<b>Strain</b>	<b>Mating type</b>	<b>Relevant genotype</b>	<b>Description</b>	<b>Source</b>
JAY1620	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1521 x JAY328	This study
JAY1622	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1526	This study
JAY1624	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1531	This study
JAY1626	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1536	This study
JAY1628	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1539	This study
JAY1630	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1545	This study
JAY1632	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1546	This study
JAY1634	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1553	This study
JAY1636	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY1557	This study
JAY1638	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY290	This study
JAY1640	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY293	This study
JAY1642	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1522 x JAY328	This study
JAY1644	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY1531	This study
JAY1646	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY1536	This study
JAY1648	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY1539	This study
JAY1650	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY1545	This study
JAY1652	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY1546	This study
JAY1654	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY1553	This study
JAY1656	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY1557	This study
JAY1658	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY290	This study
JAY1660	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY293	This study
JAY1662	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1529 x JAY328	This study
JAY1664	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY1536	This study
JAY1666	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY1539	This study
JAY1668	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY1545	This study
JAY1670	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY1546	This study
JAY1672	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY1553	This study
JAY1674	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY1557	This study
JAY1676	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY290	This study
JAY1678	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY293	This study
JAY1680	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1533 x JAY328	This study
JAY1682	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY1539	This study
JAY1684	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY1545	This study
JAY1686	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY1546	This study
JAY1688	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY1553	This study
JAY1690	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY1557	This study
JAY1692	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY290	This study
JAY1694	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY293	This study
JAY1696	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1535 x JAY328	This study
JAY1698	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1540 x JAY1545	This study
JAY1700	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1540 x JAY1546	This study
JAY1702	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1540 x JAY1553	This study
JAY1704	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1540 x JAY1557	This study

**Table S3.1. (continued).** Yeast strains used in this study.

Strain	Mating type	Relevant genotype	Description	Source
JAY1706	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1540 x JAY290	This study
JAY1708	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1540 x JAY293	This study
JAY1710	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1540 x JAY328	This study
JAY1712	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1543 x JAY1546	This study
JAY1714	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1543 x JAY1553	This study
JAY1716	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1543 x JAY1557	This study
JAY1718	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1543 x JAY290	This study
JAY1720	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1543 x JAY293	This study
JAY1722	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1543 x JAY328	This study
JAY1724	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1549 x JAY1553	This study
JAY1726	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1549 x JAY1557	This study
JAY1728	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1549 x JAY290	This study
JAY1730	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1549 x JAY293	This study
JAY1732	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1549 x JAY328	This study
JAY1734	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1552 x JAY1557	This study
JAY1736	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1552 x JAY290	This study
JAY1738	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1552 x JAY293	This study
JAY1740	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1552 x JAY328	This study
JAY1742	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1554 x JAY290	This study
JAY1744	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1554 x JAY293	This study
JAY1746	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY1554 x JAY328	This study
JAY1748	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY291 x JAY293	This study
JAY1750	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY291 x JAY328	This study
JAY1752	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i>	Inbred diploid resulting from crossing between JAY295 x JAY328	This study
JAY2108	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr4-UPD P/P	JAY270 derivative containing uniparental disomy for Chr4 (P/P)	This study
JAY2133	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr4-UPD P/P	JAY270 derivative containing uniparental disomy for Chr4 (P/P)	This study
JAY2191	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr4-UPD P/P	JAY270 derivative containing uniparental disomy for Chr4 (P/P)	This study
JAY2200	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr4-UPD P/P	JAY270 derivative containing uniparental disomy for Chr4 (P/P)	This study
JAY2201	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr4-UPD M/M	JAY270 derivative containing uniparental disomy for Chr4 (M/M)	This study
JAY2202	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr4-UPD M/M	JAY270 derivative containing uniparental disomy for Chr4 (M/M)	This study
JAY2250	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr14-UPD P/P	JAY270 derivative containing uniparental disomy for Chr14 (P/P)	This study
JAY2252	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr14-UPD P/P	JAY270 derivative containing uniparental disomy for Chr14 (P/P)	This study
JAY2254	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr14-UPD P/P	JAY270 derivative containing uniparental disomy for Chr14 (P/P)	This study
JAY2244	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr14-UPD M/M	JAY270 derivative containing uniparental disomy for Chr14 (M/M)	This study
JAY2248	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr14-UPD M/M	JAY270 derivative containing uniparental disomy for Chr14 (M/M)	This study
JAY2282	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr15-UPD P/P	JAY270 derivative containing uniparental disomy for Chr15 (P/P)	This study
JAY2283	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr15-UPD P/P	JAY270 derivative containing uniparental disomy for Chr15 (P/P)	This study
JAY2284	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr15-UPD P/P	JAY270 derivative containing uniparental disomy for Chr15 (P/P)	This study
JAY2298	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr15-UPD M/M	JAY270 derivative containing uniparental disomy for Chr15 (M/M)	This study
JAY2300	<i>MATa/MATa</i>	<i>ACE2/ace2-A7</i> ; Chr15-UPD M/M	JAY270 derivative containing uniparental disomy for Chr15 (M/M)	This study

**Table S3.2.** Oligonucleotides used in this study.

Name	5' – 3' sequence <sup>1</sup>	Description
JAO14	AGGAGGGTATTCTGGGCCTCCATG	Inside <i>Mx4</i> region. To confirm insertion of <i>Mx4</i> cassettes; Rev.
JAO15	ATGCGAAGTTAAGTGCAGCAAAAG	Inside <i>Mx4</i> region. To confirm insertion of <i>Mx4</i> cassettes; Fwd.
JAO1385	GATTACACCGTACTTCTTTCAATGCGTAAACAAGTAGAGTTGACAAA CTGCTGTCGATTCGATAC	To amplify <i>GFP</i> cassette; Fwd. Homology tail to <i>CEN5</i> region.
JAO1386	CACGTCAAGACTGTCAAGGAGGGTATTCTGGGCCTCCATGTC GTCGATGAATTCGAGCTC	To amplify <i>GFP</i> cassette; Rev. Homology tail to <i>KanMx4</i> cassette.
JAO1387	GACATGGAGGCCAGGAATAC	To amplify <i>KanMx4</i> cassette; Fwd.
JAO466	TACGAAGTGCTAGGAGGTATATTATTAATATGATAGTTTTAGTTTATAATA GTGGATCTGATATCATCG	To amplify <i>KanMX4</i> cassette; Rev. Homology tail to <i>CEN5</i> region.
JAO464	CATCGTGTAGTCAAGCAGCA	To confirm integration of <i>GFP-KanMx4</i> cassette; Fwd. <i>CEN5</i> region.
JAO467	CATTTACAGATTCATAGTTC	To confirm integration of <i>GFP-KanMx4</i> cassette; Rev. <i>CEN5</i> region.
JAO1563	ACAGTACTAGCTTTTAACTTGTATCCTAGGTTATCTATGCTGTCTCACCATAGGGAATATTACCTATTTCA G CTTGCCTCGTCCCGCCG	To amplify <i>AmdS</i> cassette; Fwd. Homology tail to <i>CEN4</i> region.
JAO1566	GTATTTTAAAGTTTTTAAAAAGTTGATTAATAGCATGTGAC CTCCAGTATAGCGACCAG	To amplify <i>AmdS</i> cassette; Rev. Homology tail to region immediately upstream to <i>CEN4</i> .
JAO1567	AAAAAGTATTTTAAAGTTTTTAAAAAGTTGATTAATAGCAT CTCCAGTATAGCGACCAG	To amplify <i>AmdS</i> cassette; Rev. Integration results in 5bp deletion of <i>CEN4</i> element 1.
JAO1568	GTATTTTAAAGTTTTTAAAAAGTTGATTAATAGCATGTGAC AGTTATGGAGTAACAACG	To amplify <i>AmdS</i> cassette excluding its terminator region; Rev. Homology tail to region immediately upstream to <i>CEN4</i> .
JAO1569	CTGCAAAACAGTACTAGCTTTTAACTTG	To confirm integration of <i>AmdS</i> at <i>CEN4</i> region; Fwd.
JAO1570	GGAATATATAGCAGTAGTCAATTTAGCAC	To confirm integration of <i>AmdS</i> at <i>CEN4</i> region; Rev. Used for Sanger sequencing.
JAO1574	TCCAAGGTGGTTGCATCATA	Chr4 710,025 SNP <i>Bgl</i> II RFLP marker Fwd
JAO1575	TGGCTGGAGTTTCGTCTTCT	Chr4 710,025 SNP <i>Bgl</i> II RFLP marker Rev
JAO1576	GTATCTTAATGAACTATGCAATGG	Chr4 378,184 SNP <i>Bam</i> HI RFLP marker Fwd
JAO1577	CCTCATCGGCACATTAAGCTG	Chr4 378,184 SNP <i>Bam</i> HI RFLP marker Rev
JAO1659	AAAGAAAAAATTACTGCAAAACAGTACTAGCTTTTAACTTGTATCCTAGGTTATCTATGCTGTCTACCA T CTTGCCTCGTCCCGCCG	To amplify <i>AmdS</i> cassette; Fwd. Homology tail to <i>CEN4</i> region
JAO1681	TGGAGGAAAAGCATTGGTATTAAGTACTTTGCATTCTCTGGAGAAGAACTTGATCAATTGACGGGTATAG C CTTGCCTCGTCCCGCCG	To amplify <i>AmdS</i> cassette; Fwd. Homology tail to <i>CEN14</i> region
JAO1682	AAAATGTTTTAAAATATTTTTAAAAAGCTGCACGTGACTAAC AGTTATGGAGTAACAACG	To amplify <i>AmdS</i> cassette excluding its terminator region; Rev. Homology tail to region immediately upstream to <i>CEN14</i> .
JAO1683	GGACTACTGATGTACTGAAGTTTG	To confirm integration of <i>AmdS</i> at <i>CEN14</i> region; Fwd
JAO1684	CCTGCTGTCTAATCCGTATTCATTC	To confirm integration of <i>AmdS</i> at <i>CEN14</i> region; Rev
JAO1685	CTGAATTGGACGTTTGGTTC	Inside <i>AmdS</i> gene. To confirm insertion of <i>AmdS</i> cassettes; Fwd.
JAO1686	TCATCACTGTTCTTTTCAGATACTAGTTTCAAAAATTCCTTGACAGAACCATTTTCATGTTCAATAATGAAA A CTTGCCTCGTCCCGCCG	To amplify <i>AmdS</i> cassette; Fwd. Homology tail to <i>CEN15</i> region
JAO1687	AATGTTTAAATATTTAATGTATATGACTTCCGAAAAATATAT AGTTATGGAGTAACAACG	To amplify <i>AmdS</i> cassette excluding its terminator region; Rev. Homology tail to region immediately upstream to <i>CEN15</i> .
JAO1688	GGTAGTATAAAGAGAATAACTTCCC	To confirm integration of <i>AmdS</i> at <i>CEN15</i> region; Fwd
JAO1689	GCTCTCACCACATTTATCACC	To confirm integration of <i>AmdS</i> at <i>CEN15</i> region; Rev
JAO1692	CACAAAATATGTTTCGTACCCATCC	Chr14 259,198 SNP <i>Mfe</i> II RFLP marker Fwd
JAO1693	GTGACGGAGTTTGATACTTAAC	Chr14 259,198 SNP <i>Mfe</i> II RFLP marker Rev

**Table S3.2. (continued).** Oligonucleotides used in this study.

Name	5' – 3' sequence <sup>1</sup>	Description
JAO1694	GTTGAAGCAATGAAGGAATCTTCG	Chr14 758,547 SNP <i>MseI</i> RFLP marker Fwd
JAO1695	CCTCAGTCATGCCAAGTGCATCAAC	Chr14 758,547 SNP <i>MseI</i> RFLP marker Rev
JAO1696	CGTCTCGTATTATACTTGCCTG	Chr15 16,733 SNP <i>SalI</i> RFLP marker Fwd
JAO1697	CATCGAGGAACAACAGCATCTAG	Chr15 16,733 SNP <i>SalI</i> RFLP marker Rev
JAO1588	GAGTATTGACAGGGAATAC	Chr15 1,058,946 SNP <i>NdeI</i> RFLP marker Fwd
JAO1589	GGAGCTTACTTCACTCATTTTC	Chr15 1,058,946 SNP <i>NdeI</i> RFLP marker Rev

1. Nucleotides in the 5' end (**bold**) have homology to the respective chromosomal insertion sites. 3' end nucleotides (not in bold) correspond to primer sequences to amplify the desired amplicons.

**Table S3.3.** Growth conditions tested through plate spotting assay.

<b>Base Media</b>	<b>Additional treatment/supplement</b>	<b>Temperature of incubation</b>	<b>Hours of incubation</b>
2% YPD	-	30°C	48
2% YPD	-	39°C	96
2% YPGE	-	30°C	48
2% YP Galactose	-	30°C	48
2% YP Raffinose	-	30°C	48
2% YPD	7 % EtOH	30°C	48
2% YPD	11 % EtOH	30°C	48
2% YPD	30 mM furfural	30°C	48
2% YPD	0.75 mM menadione	30°C	48
2% YPD	0.01% MMS	30°C	48
2% YPD	100 mM hydroxyurea	30°C	96
2% YPD	100 J/m <sup>2</sup> UV light	30°C	48
2% YPD	150 J/m <sup>2</sup> UV light	30°C	48

**Table S3.4.** Summary of phenotypic data measure in competition assays (CA) and tolerance to high temperature assays (HT) for each inbred strain.

Strain	Phenotype in CA				Score in HT
	Cycle 0	Cycle 2	Cycle 5	Cycle 8	
JAY1598	50.50	33.72	19.15	13.32	3.3
JAY1600	54.40	36.87	22.27	10.03	1.0
JAY1602	50.65	18.00	5.72	2.39	3.5
JAY1604	54.45	38.85	25.68	12.97	1.5
JAY1606	54.00	59.88	74.95	78.28	1.3
JAY1608	54.10	43.32	32.38	26.13	3.4
JAY1610	54.20	39.92	37.22	25.55	5.0
JAY1612	49.95	36.57	25.97	14.63	3.8
JAY1614	47.15	11.48	1.27	0.81	3.5
JAY1616	50.35	57.08	73.27	70.05	2.7
JAY1618	53.55	57.93	73.47	76.57	0.8
JAY1620	47.45	22.00	6.85	3.11	3.3
JAY1622	49.75	52.28	62.38	59.13	0.3
JAY1624	47.95	38.35	31.53	21.43	3.3
JAY1626	48.25	53.23	69.08	76.03	1.3
JAY1628	47.70	67.25	92.93	98.15	2.0
JAY1630	47.60	50.40	64.62	71.98	3.3
JAY1632	48.75	48.18	63.85	70.22	4.7
JAY1634	45.95	46.48	51.25	51.07	2.6
JAY1636	49.05	33.93	16.63	7.22	2.5

**Table S3.4.** Summary of phenotypic data measure in competition assays (CA) and tolerance to high temperature assays (HT) for each inbred strain.

Strain	Phenotype in CA				Score in HT
	Cycle 0	Cycle 2	Cycle 5	Cycle 8	
JAY1638	46.60	58.07	83.22	93.55	2.4
JAY1640	47.40	63.05	90.82	97.47	1.5
JAY1642	40.25	28.00	24.68	15.60	3.7
JAY1644	47.95	45.53	62.88	61.22	3.8
JAY1646	49.20	55.67	79.45	88.7	3.4
JAY1648	45.40	65.70	93.83	97.42	3.8
JAY1650	49.45	46.10	68.23	70.28	4.3
JAY1652	45.75	54.32	81.95	91.70	3.6
JAY1654	39.90	61.55	88.72	96.30	4.2
JAY1656	45.40	44.32	49.82	50.20	4.3
JAY1658	45.35	65.97	90.13	96.32	4.0
JAY1660	49.60	64.93	87.97	96.30	3.7
JAY1662	40.60	30.02	20.12	12.95	4.3
JAY1664	47.05	67.30	92.93	97.93	0.0
JAY1666	51.83	41.30	28.87	18.93	0.0
JAY1668	52.20	28.53	5.56	2.29	1.8
JAY1670	52.02	63.97	71.73	74.48	2.9
JAY1672	50.53	51.68	45.70	25.42	0.7
JAY1674	49.07	37.38	18.28	9.05	2.8
JAY1676	50.70	51.02	45.40	29.93	0.7

**Table S3.4.** Summary of phenotypic data measure in competition assays (CA) and tolerance to high temperature assays (HT) for each inbred strain.

Strain	Phenotype in CA				Score in HT
	Cycle 0	Cycle 2	Cycle 5	Cycle 8	
JAY1678	49.55	41.27	50.87	46.52	0.8
JAY1680	45.35	59.17	28.92	38.80	3.2
JAY1682	50.23	45.62	47.33	34.42	2.7
JAY1684	49.00	36.33	14.77	5.20	4.3
JAY1686	52.93	31.18	28.03	17.82	4.3
JAY1688	52.82	30.47	24.83	14.92	5.0
JAY1690	49.32	28.15	8.07	2.89	4.8
JAY1692	43.13	32.80	24.40	15.27	4.0
JAY1694	48.32	30.17	27.12	14.50	3.0
JAY1696	44.93	44.48	31.73	33.77	4.4
JAY1698	52.42	48.17	31.33	16.68	3.0
JAY1700	48.68	32.05	26.65	2.93	5.0
JAY1702	50.48	30.12	17.13	2.96	2.5
JAY1704	49.87	55.72	71.50	84.67	4.4
JAY1706	50.80	64.72	86.90	96.1	3.7
JAY1708	53.53	81.92	92.57	97.25	3.3
JAY1710	46.03	25.37	25.25	24.93	4.8
JAY1712	51.80	19.22	16.70	16.70	4.0
JAY1714	52.30	24.70	22.67	23.55	2.5
JAY1716	51.90	7.05	3.56	4.00	2.0

**Table S3.4.** Summary of phenotypic data measure in competition assays (CA) and tolerance to high temperature assays (HT) for each inbred strain.

Strain	Phenotype in CA				Score in HT
	Cycle 0	Cycle 2	Cycle 5	Cycle 8	
JAY1718	51.47	46.05	57.65	75.33	1.9
JAY1720	49.82	41.17	43.68	35.63	1.5
JAY1722	46.68	7.64	5.25	4.94	3.0
JAY1724	43.30	18.58	14.02	8.49	1.5
JAY1726	43.73	8.52	4.18	5.11	3.8
JAY1728	39.18	34.82	25.07	16.42	2.7
JAY1730	40.37	33.60	23.30	17.52	1.3
JAY1732	38.78	13.20	4.97	1.68	3.3
JAY1734	51.42	36.17	12.00	3.91	4.2
JAY1736	50.88	59.33	69.98	79.22	3.0
JAY1738	51.72	55.88	58.42	61.85	2.7
JAY1740	47.28	35.82	19.05	10.22	4.0
JAY1742	53.95	69.72	90.65	97.78	1.5
JAY1744	55.13	65.48	86.57	95.35	0.5
JAY1746	48.22	42.45	41.25	38.33	2.8
JAY1748	51.38	38.30	22.48	12.52	2.5
JAY1750	48.55	15.60	4.05	1.83	4.2
JAY1752	48.17	15.15	4.11	2.31	3.2

**Table S3.5.** List of phased JAY270 HetSNPs interrogated for confirmation of UPD strains.

Marker SGD coordinates		PCR primers	Polymorphism detection method	SGD Watson bases <sup>1</sup>	JAY270 Watson phased bases	
Chromosome	Nucleotide				Chr-P	Chr-M
Chr4	378,184	JAO1576/JAO1577	<i>Bam</i> HI	<u>C</u>	T	<u>C</u>
Chr4	710,025	JAO1574/JAO1575	<i>Bgl</i> II	<u>I</u>	G	<u>I</u>
Chr14	259,198	JAO1692/JAO1693	<i>Mfe</i> I	<u>I</u>	<u>I</u>	A
Chr14	758,547	JAO1694/JAO1695	<i>Mse</i> I	<u>A</u>	G	<u>A</u>
Chr15	16,733	JAO1696/JAO1697	<i>Sal</i> I	<u>C</u>	T	<u>C</u>
Chr15	1,058,946	JAO1588/JAO1589	<i>Nde</i> I	T	T	<u>C</u>

1. Nucleotides that are underlined indicate the sequences that are cut by the respective restriction endonucleases, whereas lack of underlining corresponds to no cut site.

**Table S3.6.** Details on genomic regions showing significant association to the growth kinetics and heat tolerance phenotypes.

Phenotype Assay	Chromosome Number	Genome Coordinates	Number of Markers	Number of Genes in Region	Max LOD value	Quantitative Inheritance Model
GROWTH	4	Chr04_558889..572184 (15.2kb)	17	8	5.382	Additive Maternal
GROWTH	10	Chr10_467589..569699 (102.1kb)	48	52	6.280	Additive Paternal
GROWTH	15	Chr15_1024820..1069385 (44.6kb)	138	21	4.324	Dominant Paternal Low
GROWTH	16	Chr16_289873..361102 (71.2kb)	134	36	5.351	Dominant Maternal
TEMPERATURE	5	Chr05_549677..565641 (16.0kb)	40	9	5.561	Dominant Paternal
TEMPERATURE	6	Chr06_85688..97759 (12.1kb)	33	4	6.042	Additive Maternal
TEMPERATURE	7	Chr07_82753..95670 (12.9kb)	30	7	5.230	Dominant Maternal
TEMPERATURE	7	Chr07_673392..730469 (57.1kb)	9	27	7.574	Additive Paternal
TEMPERATURE	7	Chr07_849524..887971 (38.4kb)	14	18	5.403	Dominant Paternal
TEMPERATURE	13	Chr13_75040..92531 (17.5kb)	50	9	8.172	Additive Maternal
TEMPERATURE	13	Chr13_715996..749967 (34.0kb)	17	19	9.800	Additive Paternal
TEMPERATURE	13	Chr13_754723..818486 (63.8kb)	62	36	6.387	Additive Paternal
TEMPERATURE	14	Chr14_528514..579106 (50.6kb)	37	29	10.638	Additive Maternal
TEMPERATURE	14	Chr14_596319..605480 (9.2kb)	2	4	10.638	Additive Maternal
TEMPERATURE	14	Chr14_662399..662537 (0.14kb)	6	0	4.311	Dominant Maternal
TEMPERATURE	14	Chr14_691878..695118 (3.2kb)	6	3	5.072	Dominant Maternal
TEMPERATURE	15	Chr15_764174..781222 (17.0kb)	13	9	5.787	Additive Maternal