

# Whole Genome Analysis of 132 Clinical *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation

Yuan O. Zhu<sup>\*†‡</sup>, Gavin Sherlock<sup>\*§</sup> and Dmitri A. Petrov<sup>†§,1</sup>

<sup>\*</sup>Department of Genetics, Stanford University, CA 94305, <sup>†</sup>Department of Biology, Stanford University, CA 94305, <sup>‡</sup>Genome Institute of Singapore, Singapore 138672, <sup>§</sup>These authors contributed equally to this work

**ABSTRACT** Budding yeast has undergone several independent transitions from commercial to clinical lifestyles. The frequency of such transitions suggests that clinical yeast strains are derived from environmentally available yeast populations, including commercial sources. However, despite their important role in adaptive evolution, the prevalence of polyploidy and aneuploidy has not been extensively analyzed in clinical strains. In this study, we have looked for patterns governing the transition to clinical invasion in the largest screen of clinical yeast isolates to date. In particular, we have focused on the hypothesis that ploidy changes have influenced adaptive processes. We sequenced 144 yeast strains, 132 of which are clinical isolates. We found pervasive large-scale genomic variation in both overall ploidy (34% of strains identified as 3n/4n) and individual chromosomal copy numbers (36% of strains identified as aneuploid). We also found evidence for the highly dynamic nature of yeast genomes, with 35 strains showing partial chromosomal copy number changes and 8 strains showing multiple independent chromosomal events. Intriguingly, a lineage identified to be Baker's/commercial derived with a unique damaging mutation in NDC80 was particularly prone to polyploidy, with 83% of its members being triploid or tetraploid. Polyploidy was in turn associated with a >2x increase in aneuploidy rates as compared to other lineages. This dataset provides a rich source of information of the genomics of clinical yeast strains and highlights the potential importance of large-scale genomic copy variation in yeast adaptation.

## KEYWORDS

population structure  
ploidy  
aneuploidy  
adaptation  
population sequencing

## INTRODUCTION

Yeast species classified under the *Saccharomyces sensu stricto* species complex, in particular *S. cerevisiae*, have played vital cultural and economic roles in human history (Sicard and Legras 2011). Indeed, yeast is probably one of the first species domesticated by man. Ancient evidence for the role of *Saccharomyces* in fermentation dates as far back as the Neolithic period, including wine jars dating to 7000 B.C. China (McGovern *et al.* 2004), 5400-5000 B.C. Iran (McGovern *et al.* 1997), and 3150 B.C. Egypt (Cavaliere *et al.* 2003). Another Egyptian study revealed some of the oldest evidence for the active use of yeasts in bread and beer making around 1300-1500 B.C.

(Samuel 1996). Yeasts are widely used in industrial applications, with yeast fermentation being harnessed in the development of biofuel technologies that produce ethanol as a renewable source of energy (Farrell *et al.* 2006).

With extensive human contact over nearly 7000 years, it is not surprising that *S. cerevisiae* is often found as an asymptomatic human gut commensal (Zerva *et al.* 1996; McCullough *et al.* 1998; Posteraro *et al.* 2000; Salonen *et al.* 2000; Erdem *et al.* 2003). However, it has become clear that not all human-*S. cerevisiae* interactions are harmless. Indeed, *S. cerevisiae* has been identified with increasing frequency as the causal invasive species in fungal infections, often but not necessarily associated with the use of *S. cerevisiae* var. *boulardii* (Perapoch *et al.* 2000; Lherm *et al.* 2002; Wheeler *et al.* 2003; Muñoz *et al.* 2005; Enache-Angoulvant and Hennequin 2005; De Llanos *et al.* 2006; de Llanos *et al.* 2006). While most fungal infections are caused by *C. albicans* and *C. glabrata*, *S. cerevisiae* has been estimated to be responsible for up to 3.6% of all known

Copyright © 2016 Dmitri A. Petrov *et al.*

Manuscript compiled: Tuesday 7<sup>th</sup> June, 2016%

<sup>1</sup> Department of Biology 371 Serra St. Stanford University Stanford, CA94305-5020

Phone (office): (650) 736 1169 Phone (lab): (650) 736 2249 Fax: (650) 723 6132 Email:

dpetrov@stanford.edu

cases of fungemia, the most severe clinical manifestation of fungal infections (Piarroux *et al.* 1999; Smith *et al.* 2002; Muñoz *et al.* 2005). It is also known to cause or be associated with a multitude of other symptoms that contribute to poor medical outcomes, including pneumonia, peritonitis, and liver abscesses (Dougherty and Simmons 1982; Aucott *et al.* 1990; Doyle *et al.* 1990; Mydlik *et al.* 1996; Muñoz *et al.* 2005). In extreme cases, yeast appears to be the direct cause of mortality, usually by inducing sepsis (Hennequin *et al.* 2000; Piarroux *et al.* 1999). Because of such reports, *S. cerevisiae* is now considered a human pathogen, though one of relatively low virulence (Malgoire *et al.* 2005).

Unlike species that are fully adapted to a pathogenic lifestyle, it is unclear what factors play a role in the opportunistic virulence of a usually benign commensal. There is clear epidemiological evidence supporting impaired host immunity as an important factor. For example, a large proportion of identified cases were of low virulence and occurred in patients who were immuno-compromised, had underlying medical conditions, underwent recent antibiotic treatments, used intravenous catheters, or a combination thereof (Enache-Angoulvant and Hennequin 2005; Muñoz *et al.* 2005). Infections appear to occur frequently through inadvertent exposure to environmental sources such as contaminated catheters, probiotic treatments, or close contact with other infected individuals (Perapoch *et al.* 2000; Lherm *et al.* 2002; Cassone *et al.* 2003; de Llanos *et al.* 2004). This is supported by evidence that *S. cerevisiae* cannot cross the epithelial barrier *in vitro* in the absence of epithelial compromise (Pérez-Torrado *et al.* 2012).

However, even given fortuitous exposure and host condition, there is some evidence that at least two inherent phenotypic traits of yeast strains, the ability to survive at high temperatures and the propensity for pseudohyphal formation, are highly associated with virulence (McCusker *et al.* 1994). Virulent isolates are more likely to produce multiple distinct colony phenotypes as compared to avirulent isolates, which may contribute to their ability to survive in complex host environments (Clemons *et al.* 1996). A study using mice infected with clinical and non-clinical strains showed that some clinical strains appeared more virulent than non-clinical strains, but virulence phenotypes generally overlap between the two groups, suggesting that pre-existing virulence phenotypes play a complementary role to opportunistic exposure (Clemons *et al.* 1994). It should be noted, however, that another study found no appreciable differences between clinical and non-clinical strains (Klingberg *et al.* 2008), although this result does not contradict the possibility that a subset of 'safe' commercial yeast strains are phenotypically more virulent (De Llanos *et al.* 2011), and may be more likely to cause infections given the opportunity.

Unlike phenotypic traits, the genetic factors that may influence or drive opportunistic yeast infections, if any, have been more difficult to pin down. Some gene candidates have been reported to be associated with virulence phenotypes (Muller *et al.* 2011; Strobe *et al.* 2015), but most such studies were conducted on haploid derivatives from clinical isolates, whereas a microsatellite study showed that clinical yeasts are often highly aberrant in their ploidy (Muller and McCusker 2009b). This variability in ploidy and chromosomal copy number is unsurprising yet important to our understanding of yeast adaptation. Historically, yeast has had to adapt to new, stressful and drastically different environments repeatedly in the process of domestication and commercialization. Some of the best examples of how yeast strains accomplished this come from our understanding of how *S. cerevisiae* adapted to commercial fermentation (Dequin and Casaregola 2011). Commercial yeasts are extremely tolerant of large-scale copy number

changes and both bread and brewing strains are often found to have >2n genomes (Albertin *et al.* 2009). Certain strains of sherry yeasts have stable aneuploid chromosomes and a few ale strains have 4n alleles at several loci (Legras *et al.* 2007; Albertin *et al.* 2009). Large structural rearrangements and copy number variants are widespread and may be associated with selective forces in stressful environments (Bidenne *et al.* 1992; Carro *et al.* 2003; Wei *et al.* 2007; James *et al.* 2008; Schacherer *et al.* 2009; Dunn *et al.* 2012). Aneuploidy, whole chromosomal copy number changes that typically have large fitness effects (Sunshine *et al.* 2015), are also common in wild as well as clinical strains (Hose *et al.* 2015; Gasch *et al.* 2016). Ploidy changes and aneuploidy are thus two large-scale genomic changes known to result in substantial fitness effects in yeast (Cromie and Dudley 2015), and both have been associated with adaptation (Selmecki *et al.* 2009; Pavelka *et al.* 2010; Rancati and Pavelka 2013; Thorburn *et al.* 2013; Zörgö *et al.* 2013; Selmecki *et al.* 2015). However, a large scale, whole genome survey of ploidy in clinical yeast strains has not been conducted. In order to investigate the true extent of ploidy variation and aneuploidy abundance in clinical yeast strains, we sequenced 144 strains of yeast in their natural ploidy states, the majority of which are clinical yeasts isolates, to an average genomic coverage of 80x per strain. We found that 32% of strains assayed contained more than 2 copies of each chromosome (3n-5n), and 35% contained aneuploidies at one or more chromosomes. This highlights the importance of whole genome and chromosomal copy number variation in yeast population genomics and adaptation.

## MATERIALS AND METHODS

### Library Prep and Sequencing

DNA extraction and sequencing was carried out in conjunction with strains prepared in (Lang *et al.* 2013). Briefly, DNA was extracted using a modified glass bead lysis protocol and libraries were constructed with TruSeq DNA library kits. 100bp paired-end Illumina sequencing libraries were prepared at The Genome Institute, Washington University School of Medicine, and run on Illumina GAI with 12 samples pooled per lane (TruSeq adapters AD001-AD012), to an average depth of 80x coverage per sample.

### Mapping and SNP Calling

The mapping pipeline was similar to that described in (Zhu *et al.* 2014). Briefly, fastq files were mapped to a reference genome panel (consisting of *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. uvarum*) with bwa v0.5.9 (Li and Durbin 2009), sorted and indexed with samtools v0.1.18 (Li *et al.* 2009), and assigned strain ID with Picard tools v1.55. Duplicated read pairs were removed and remaining reads locally realigned with GATK v2.1-8 (McKenna *et al.* 2010). The UnifiedGenotyper was used to call candidate variants across all samples simultaneously. The resulting VCF file was filtered for variants with coverage depth >8X, >2 reads and >15% of reads supporting alternative variant. Around 600kb of the genome – these regions were annotated in the SGD database as simple repeats, centromeric regions, telomeric regions, or LTRs (SGD project, downloaded 4th Aug 2012) were excluded from analysis due to their susceptibility to mismapping and associated miscalls. For full list of commands, refer to Supplementary materials (Zhu *et al.* 2014).

### Aneuploidy Calling, Gene Copy Number, and GO Term Analysis

Aneuploidy calling process was similar to method used in (Zhu *et al.* 2014) and carried out by computing coverage depth of non-overlapping 1kb windows. Whole chromosome aneuploidies were

called if average coverage of a chromosome differs more than 35% from other chromosomes in the same line (likelihood  $P < 0.001$ , Chi-Squared Test), and visually confirmed through coverage traces (see Supplemental Figures for examples). Gene copy number was estimated using normalized gene coverage against genome wide distribution, assuming binomial distribution of read depth. GO term analysis was conducted with g:Profiler (Reimand *et al.* 2011) and AmiGO 2 (Carbon *et al.* 2009).

### Neighbor-Joining Tree and Structure

Variants for the Liti strains (Liti *et al.* 2009) were identified according to an older version 2 of the yeast genome. Only sequenced SNPs from the Liti strains (not those inferred by imputation) were converted to version 3 coordinates using LiftOver scripts from the UCSC genome browser website. SNPs called in both datasets with coverage in at least half of the strains within each dataset, were retained for NJ tree construction. The neighbor-joining tree was constructed using PhyML-3.1 (<http://www.atgc-montpellier.fr/phyml>), with missing data included as 'N'. All tree figures were constructed with the online program Interactive Tree of Life (Letunic and Bork 2007). Structure (version 2.3.2) analysis was run with independent allele frequencies under an admixture model, for a 15,000 burn in period and 5,000 iterations of sampling.

## RESULTS

### High Coverage Whole Genome Sequencing

145 *S. cerevisiae* strains were cultured and sequenced in their natural ploidy states. 13 of these strains were from non-clinical sources; the remaining 132 came from clinical sources. 8 strains were recorded as recently derived from strains already present in the collection, and were expected to be genetically very similar. The clinical strains were collected from at least 12 different geographical locations and isolated from at least 14 human tissue types or procedures. Some of the strains have been previously studied in one or more earlier publications: YJM798 was one of the best-studied clinical yeast strains with a fully sequenced genome (Wei *et al.* 2007), 14 strains were part of a clinical yeast pathogenesis study (Clemons *et al.* 1994), 4 strains were part of a clinical yeast phenotypic study (de Llanos *et al.* 2006), 85 strains were part of a genetic diversity by microsatellite study (Muller and McCusker 2009a), 13 strains were part of a GWAS study (Muller *et al.* 2011), and 45 strains were sporulated and a single segregant fully sequenced in (Strope *et al.* 2015) [Table S1]. Unlike earlier studies, such as (Liti *et al.* 2009) and (Strope *et al.* 2015), we sequenced the clinical strains as is, without sporulation. While extensive heterozygosity, coupled with polyploidy and aneuploidy makes *de novo* genome assembly a challenging task, it does provide easily accessible information on chromosomal copy number and heterozygous sites.

Library sequencing, mapping, and analysis protocols are detailed in full under Materials and Methods. Briefly, libraries were constructed for each strain, and 101 bp, paired-end Illumina reads were generated, with an intended average genome-wide read depth of 80X. As it is also known that both hybridization and introgression are not uncommon in clinical budding yeasts (Muller and McCusker 2009a; Strope *et al.* 2015), the fastq reads were mapped to a combined *Saccharomyces sensu stricto* genome that included sequences from *Saccharomyces cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. uvarum*. The S288C reference sequence was downloaded from the *Saccharomyces* Genome Database (Sheppard *et al.* 2015); all other species were downloaded from the *Saccharomyces sensu stricto* Resources site ([www.saccharomycessensustricto.org](http://www.saccharomycessensustricto.org)). One strain was subsequently removed from analysis due to a likely

contamination during library construction, with the vast majority of reads failing to map entirely [Methods]. The coverage depth for the other 144 individual strains ranged from 57X to 215X within the *S. cerevisiae* sequences. Reads were then deduplicated, locally realigned, and recalibrated following best recommended practices on the Genome Analysis Toolkit (GATK) website. Single nucleotide polymorphic sites (SNPs) were filtered by coverage depth, mapping quality, and minimum number of reads supporting variant allele. SNPs falling within known repeat sequences and telomeric regions were subsequently excluded from analysis.

### No Clinical Specific Point Mutations Found

Within *S. cerevisiae* genomic sequences, a total of 477,512 single nucleotide polymorphic sites (SNPs) were called across all non-S288C strains. Individual strains each carry between 36,786 – 97,758 variants (0.3%–0.8%) as compared to S288C. Although the number of non-clinical strains was small, allele frequencies at diallelic positions could be compared between clinical and non-clinical strains. Correlation between the clinical and nonclinical populations was high, with an  $r^2$  value of 0.9346 [Figure S1]. Fst values were calculated for every position, but values did not exceed 0.15, signifying that no SNPs were differentiated between clinical and non-clinical strains [Figure S2]. Thus, no further analysis to identify clinical specific alleles was performed.

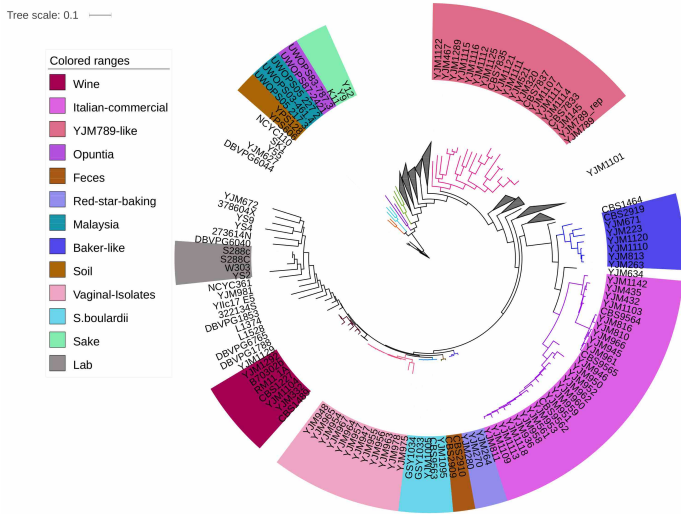
### Population Phylogeny Recapitulates Earlier Studies

A neighbor-joining (NJ) tree was constructed with the newly sequenced strains as well as 35 sequences previously published in (Liti *et al.* 2009) (Figure 1). The two NA soil strains, the three Malaysian strains, the three sake strains, the two baker strains, and the lab strains W303 and S288C cluster into independent clades, as would be expected (Strope *et al.* 2015). The clade consisting of YJM789, a strain isolated from lung, also included other strains isolated from lung, paracentesis fluid, and blood. Six clinical strains appeared to be closely related to Fleischmann's baking yeast. Twenty-three clinical strains fell with four commercial yeast strains from Italy on the NJ tree (although it was unclear what precisely the commercial niche was). Four of these 23 strains were documented as clinical isolates from nearby Greece. Three clinical strains further down the tree appeared to be nearly identical to a pro-biotic strain labeled as *S. boulardii*. Ten strains were closest to two known vaginal isolates, although information regarding their sites of isolation was not available. A handful of strains fell with RM11\_1A, a derivative of a natural vineyard isolate from California. Finally, a couple of clinical strains appeared most similar to Baker's yeast strains.

### Population Structure Identifies Multiple Origins Of Clinical Strains

Overall, the clinical strains were scattered throughout the clades of non-clinical strains on the phylogenetic tree, suggesting multiple independent origins. To obtain a more explicit grouping of ancestral source population, we used Structure (Pritchard *et al.* 2000; Falush *et al.* 2003, 2007; Hubisz *et al.* 2009) on the same set of SNPs used to create the NJ tree [Methods]. The clinical strains were distinguished into 5 main populations in our analysis. The non-clinical strains that corresponded to these populations included baking strains, sake/soil strains, wine strains, and unspecified commercial strains from Italy. A fifth population consisted of only clinical strains, including YJM789, similar to previous studies that found an unknown source population for clinical strains (Liti *et al.* 2009; Strope *et al.* 2015). The remainder of the clinical strains





**Figure 1** Neighbour-joining tree of 144 sequenced strains and 35 (Liti *et al.* 2009) strains, based on SNP differences within the *S. cerevisiae* genome. Groups of strains that fall within a cluster containing non-commercial strains from a single known source are highlighted in color blocks. Clades containing only clinical strains without a known source are collapsed, with the full tree available in Figure 4.

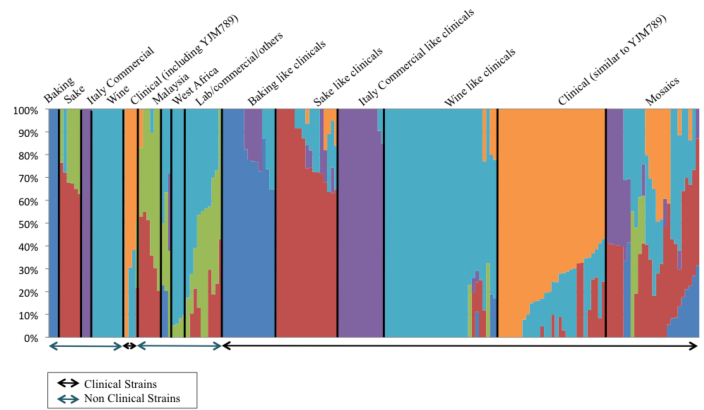
showed varying amounts of admixture from various population sources (Figure 2).

### Introgessions From *S. paradoxus* and *S. kudriavzevii*

In addition to reads mapping to the *S. cerevisiae* genome, three strains also had some sequence reads that mapped to other *Saccharomyces* species' genomes.

YJM264 is a known hybrid (Muller and McCusker 2009b). Reads covered complete chromosomes from both *S. cerevisiae* and *S. kudriavzevii* species. Coverage plots across the genomes showed the presence of at least one copy of all *S. kudriavzevii* chromosomes except V, XI, XIV and XVI, and likely two copies of *S. kudriavzevii* chromosomes IX and XV. *S. kudriavzevii* chromosomes XII and XIII were partially present, with reads mapping to the first 435Kb of XII and the first 364Kb of XIII. Coverage across the *S. cerevisiae* chromosomes showed complementary patterns, including lower copy numbers for the partial segments on chromosomes XII (first 450Kb) and XIII (first 390Kb), such that combined across both genomes, there are three copies of all chromosomes. The only exceptions were chromosome V, which appeared to have 4 copies, and chromosome II, which appeared to have a segmental duplication spanning the first 200Kb [Figure S3].

CBS2909 and CBS2910, which were isolated from human feces, are located immediately adjacent on the phylogeny. Both strains showed highly concordant, though not perfectly identical, regions of high coverage along the *S. paradoxus* genome, ranging between 2kb-40kb in size. 818 such 1kb windows were observed in CBS2909 and 778 such 1kb windows were observed in CBS2910. This mapping pattern was not observed in any other strain. Mapping results suggest that they contain mainly *S. cerevisiae* genomes with *S. paradoxus* or *S. paradoxus*-like fragments interspaced across the genome [Figure S4]. Removing regions that may be prone to mismapping artifacts (artificial coverage in >50% of all strains), 47 remaining regions consisting of 292 *S. paradoxus* genes were suspected of introgression into CBS2909 (279 genes) and CBS2910 (274 genes) [Table S2]. 261 genes were shared by both strains. Other yeast



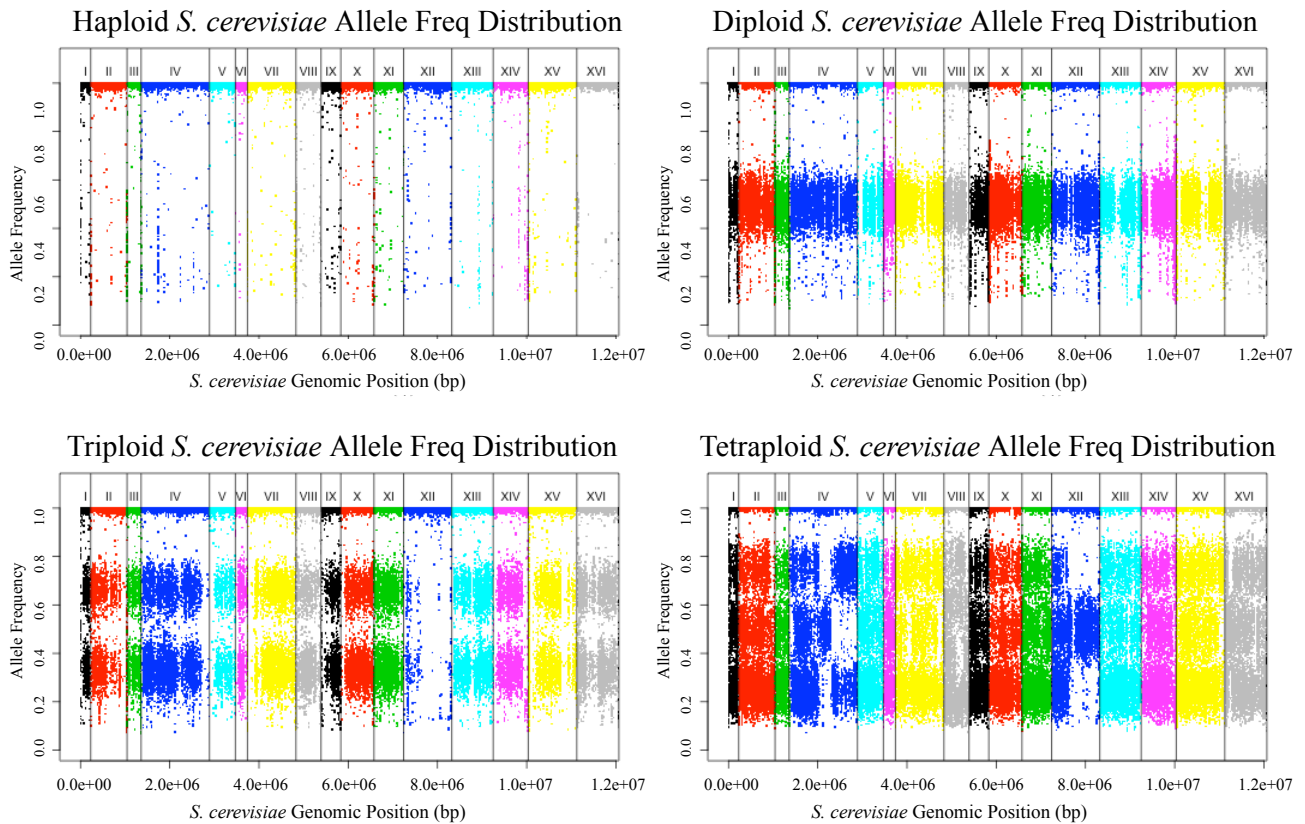
**Figure 2** Structure assignment of source populations for all 180 strains. Strains are aligned from left to right (X-axis) and sorted by their percentage composition of each population (Y-axis) as computed by Structure. Source populations are annotated with text above the figure. Clinical or non-clinical strains are annotated with arrows below the figure. Structure (version 2.3.2) analysis was run with correlated frequencies, admixture model based on base pair distance from preceding SNP, 15,000 burn in period and 5,000 iterations of sampling.

strains have also been previously found to contain introgressions from *S. paradoxus* (Strope *et al.* 2015). *S. paradoxus* is often found in the same habits as *S. cerevisiae*, and such results may indicate ancient hybridization (Sampaio and Gonçalves 2008; Sniegowski *et al.* 2002). No other clinical strains showed significant coverage across non-*S. cerevisiae* sequences [Table S3].

### Ploidy Variation Is Common In Clinical Strains

In *S. cerevisiae* strains, ploidy variation is common. For 96 of the strains that have been included in previously published work, sporulation patterns indicated that 11 strains were haploids, 56 were diploids, 15 were triploids, and 14 were tetraploids (Muller and McCusker 2009a). However, because a number of generations had elapsed since the strains were first made or studied, we assigned ploidy state to each of our sequenced strains *de novo*, using observed SNP frequency distributions within the sequence reads. Haploid strains should contain only SNP frequencies of 1; diploids should contain frequencies of both 0.5 and 1, as some SNPs will be heterozygous, while others will be homozygous, and so on (Figure 3). We note however, that homozygosing events will eliminate heterozygous SNPs, while not changing ploidy. Thus, we were unable to unequivocally rule out euploidy in homozygous strains in the absence of aneuploidy or previous published work.

All 11 haploid strains matched expectations. Three of the diploid strains, YJM1104/YJM1105/YJM436, appear to have been recently homozygosed, possibly during isolation and propagation of the strain – their SNP distributions had no visible clusters of SNPs at  $p=0.5$  as compared to other diploid strains. YJM525 and YJM954 were listed as diploid strains yet had profiles similar to that expected of triploids, while YJM464, YJM1094, YJM1099, and YJM1114 were listed as triploid strains, yet had SNP frequency distributions that indicated diploidy. YJM1146 was also listed as triploid but appeared to be tetraploid based on our SNP read frequency data. Finally, YJM950 and YJM952 were listed as tetraploids, but resembled triploids in our data. Many of these strains possessed aneuploid chromosomes that may have affected sporulation efficiency or viability, leading to the discrepancy be-



**Figure 3** Top left) Allele frequency plot of a typical haploid with SNPs at allele frequencies at 1. Top right) Allele frequency plot of a typical diploid with SNPs at allele frequencies at 0.5 and 1. Bottom left) Allele frequency plot of a typical triploid with SNPs at allele frequencies at 0.33, 0.67, and 1. Bottom right). Allele frequency plot of a typical tetraploid with SNPs at allele frequencies at 0.25, 0.5, 0.75, and 1. Chromosomes are lined in increasing order (X-axis), with raw SNP allele frequencies along the genome plotted (Y-axis).

tween sporulation patterns and our SNP frequency profiles. Across all 144 strains, there were 16 (11%) haploids, 82 (57%) diploids, 23 (16%) triploids, and 23 (16%) tetraploids. Most strains that were 3n or 4n fell within two clades on the phylogeny, the Italian commercial clade and the baker strain clade, suggesting a common recent ancestry (Figure 4).

### Aneuploidy Is Found In More Than One Third Of Clinical Strains Surveyed

Inconsistencies in allele frequency distributions across chromosomes within a single strain were also observed. This could have been due to recent homozygosity of individual chromosomes, or more likely due to individual chromosomes being aneuploid. We assigned the presence of aneuploidies in each strain by examining coverage for 1kb non-overlapping windows across all 16 *S.cerevisiae* chromosomes. 53 of 144 strains (36%) were observed to contain aneuploidies, a number nearly 5 times higher than the 7.5% previously observed in newly isolated segregants (Strope et al. 2015). Some chromosomes were more often observed in an aneuploid state than others, possibly correlated with chromosomal size (Figure 5). The most commonly observed whole chromosomal aneuploidy was chromosome IX, with 22 observed events across all strains, as compared to the 1-13 events observed in all other chromosomes.

When the rate of observed aneuploidies was compared between strains with ploidy states that are thought to be more stable (haploids and diploids) and strains with ploidy states that are thought to be unstable ( $\geq 3n$ ), there was a significantly higher aneuploidy rate in the latter ( $1n/2n = 27.3\%$ ,  $\geq 3n = 56.5\%$ , Fisher's exact test  $p=0.0009$ ). There were also 35 instances of inconsistencies within a chromosome due to a partial chromosomal copy number change (Figure 6). Partial chromosomal copy number changes were most often seen in chromosomes III and VII. There appeared to be no correlation with chromosomal size (Figure 5).

Combined information from SNP frequency distributions and aneuploidy calls suggested that multiple chromosomal scale copy number events might be present in some strains. One diploid strain YJM1098 appeared to have a 3n aneuploid chromosome XII, but SNP frequencies also suggested that chromosome VIII has been recently homozygosed (Figure 7).

Another triploid strain, YJM466, had four chromosomal copies of both chromosome VI and IX, but chromosome VI showed SNP a frequency distribution of a tetraploid, while chromosome IX showed that of a diploid (Figure 8). YJM525, YJM674, YJM810, YJM811, YJM813, and YJM815 also showed such indications of multiple events.

### Damaging Mutation in *NDC80*: Possible Genetic Basis of High Ploidy/Aneuploidy Rates in the Baking Lineage

Strains that were polyploid or carried aneuploid chromosomes were not randomly distributed across the population. One particular clade, containing 36 members from CBS1464 to YJM811 on the phylogeny tree, contained an unusually high number of polyploid and aneuploid strains (Figure 4). 29/36 (81%) strains from the clade made up of Baker's-like and Italian Commercial strains were 3n/4n, whereas only 17/98 (17%) of all other non-segregant strains were (Fisher's exact test  $p=2.05e-11$ ). Similarly, 24/36 (67%) of strains from the baker's/Italian clade contained aneuploid chromosomes, whereas 32/109 (29%) of all other strains were (Fisher's exact test  $p=1.3e-04$ ).

We thus investigated if there were any mutations that might explain the high levels of aneuploidy (and ploidy) observed in

this particular lineage. Hundreds of genes have been associated with genomic stability and faithful replication and segregation during the cell cycle (Ouspenski et al. 1999; Stirling et al. 2011; Cheng et al. 2012). In particular, a number of genes involved in sister chromatid separation and segregation in yeast have been well studied and characterized. We focused on the genes encoding the kinetochore proteins *NDC10* (*CBF2*), *BIR1*, *NDC80* (*YIL144W*), *CEP3*, and the chromosome segregation genes *CSE4*, *IPL1*, *SMT3* (Goh and Kilmartin 1993; Biggins et al. 2001; Cho and Harrison 2012). We used SIFT to look for damaging SNP mutations that are likely to affect the function of these genes (Methods).

Most of these genes are essential and had very few damaging mutations in the strains sampled (*BIR1*: 0 strains, *CEP3*: 3 strains, *SMT3*: 1 strain, *CSE4*: 4 strains). However, a much larger number of strains carried copies of damaging mutations in *NDC80* (41 strains), *IPL1* (57 strains), and *NDC10* (27 strains). Of these, one particular proline to glutamine mutation in *NDC80* at ChrIX:78597, mostly in the heterozygous form, was only found in the baker/commercial lineage, affected 32/36 strains, and may be part of the explanation for the high rate of aberrant chromosomal copies observed in this clade. Additionally, all 36 baker/commercial lineage strains contained damaging mutations in either *NDC80* (89%) or *IPL1* (89%), with half also carrying affected copies of *NDC10* (54%). These numbers are significantly different from strains outside the lineage, where only 29/98 strains carried any damaging mutations in these genes, and the relative frequencies are just 7% in *NDC80* (Fisher's exact test  $p=2.2e-16$ ), 26% in *IPL1* ( $p=4.5e-11$ ), and 10% in *NDC10* ( $p=0.035$ ). Strikingly, almost all other damaging mutations in *NDC80* found outside of the baker/commercial lineage were in polyploid strains, with just two exceptions in two closely related segregant strains that shared a unique mutation not found in any other strain.

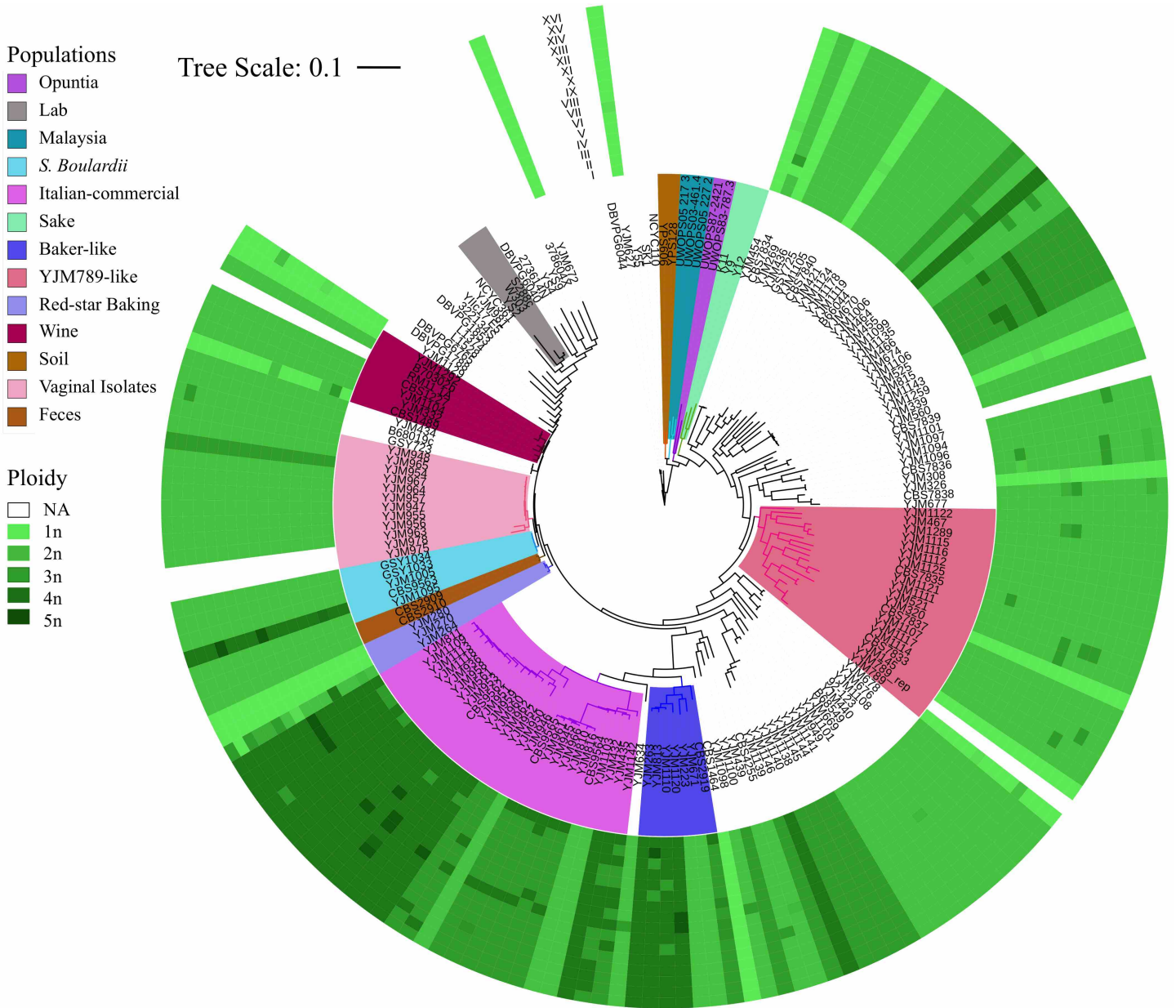
### Gene Loss and Copy Number Changes In Clinical Strains

To look for gene copy number changes that may be adaptive in all clinical strains, we investigated recent gene losses and gains for verified genes in the 144 sequenced clinical strains. Genes lost or gained in a few strains may be recent sporadic events, but a common gene loss or gain found in most clinical strains could indicate an adaptive advantage.

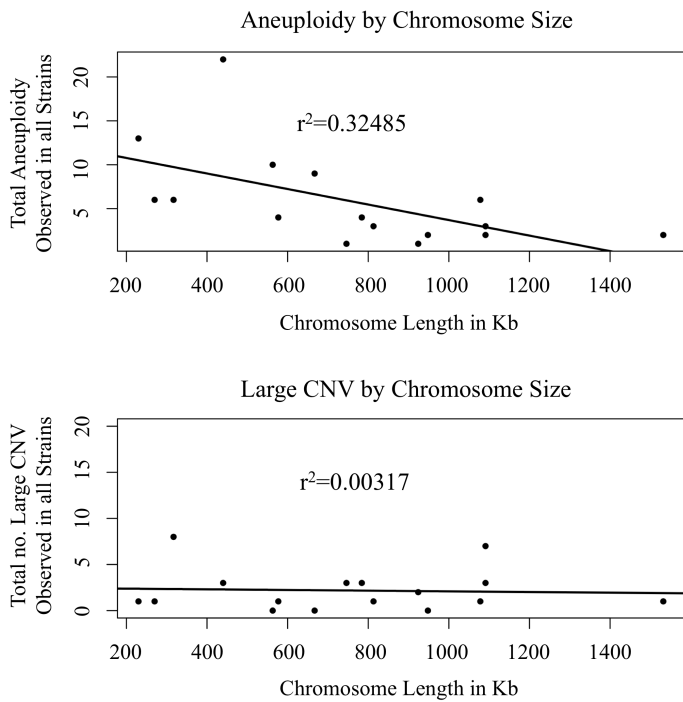
Conservatively, genes were classified as lost only if no reads mapped to any position within the coding sequence of a gene, and the Bonferroni corrected likelihood that a gene of that length is not sequenced given each strain's average genomic coverage was  $<0.05$ . 95 genes passed the filter [Table S4]. We focused on recent gene losses – those found only in  $<5\%$  (1-7 strains) of the sampled population. The 10 most significant GO terms associated with the 52 deleted genes suggested a depletion of deletion in genes involved in cellular component organization and biogenesis [Table 1]. There was enrichment in vitamin B6 related biosynthetic and metabolic pathways, largely due to a four-gene deletion (*SNZ2*, *SN02*, *SNZ3*, *SNO3*) in one strain from chromosomes VI and XIV. These genes are found in the subtelomeric region, and are known to vary in copy number (Stambuk et al. 2009). Conversely, common gene deletions (7-86 strains) showed enrichment in deletion of genes involved in transmembrane glucoside and saccharide transporter activities [Table 1].

To observe patterns in gene copy number changes that are not deletions, we included only the 87 strains with no observed or suspected recent aneuploidies, where effects from those events are likely to dominate. 109 genes were significant for duplication [Table S5]. Rare duplications (63 genes found in 1-2 strains each)

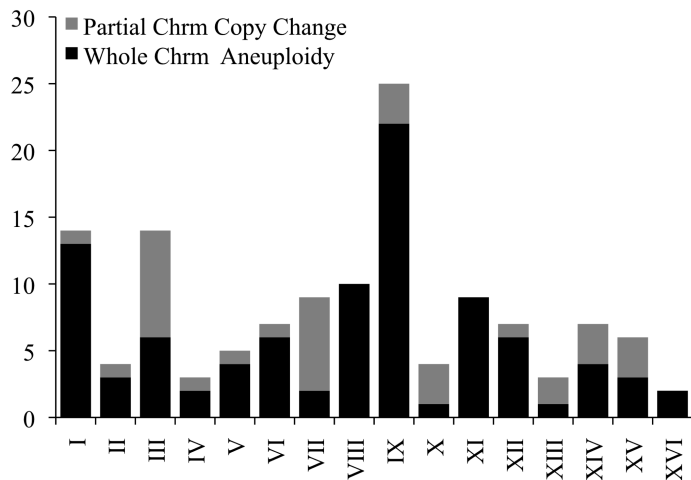




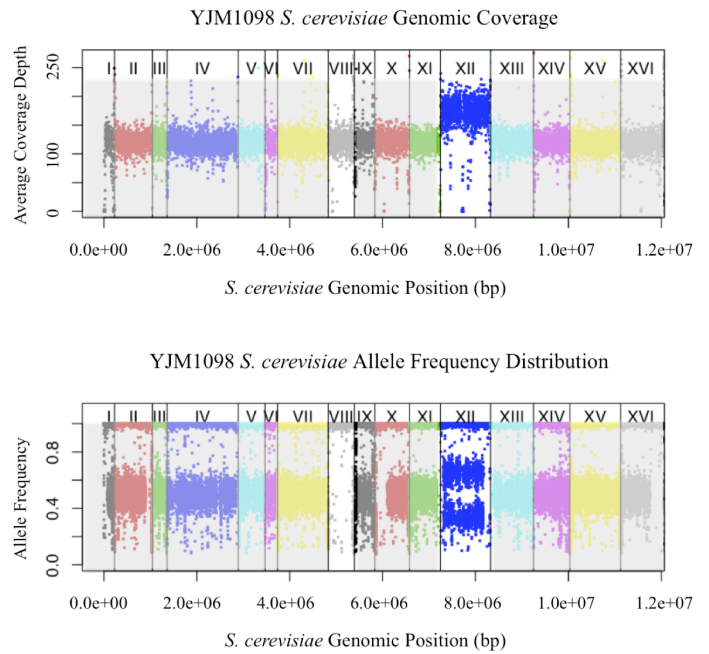
**Figure 4** Heatmap of chromosomal copy number (0n-5n) for each of 144 strains aligned next to each strain on the phylogenetic tree. The 35 Liti 2009 strains do not have chromosomal copy number information panels (in white). The 16 chromosomes of *S. cerevisiae* are sorted in increasing order in an outwards direction. Panel color from lightest to darkest green represent chromosomal copy number in increasing order.



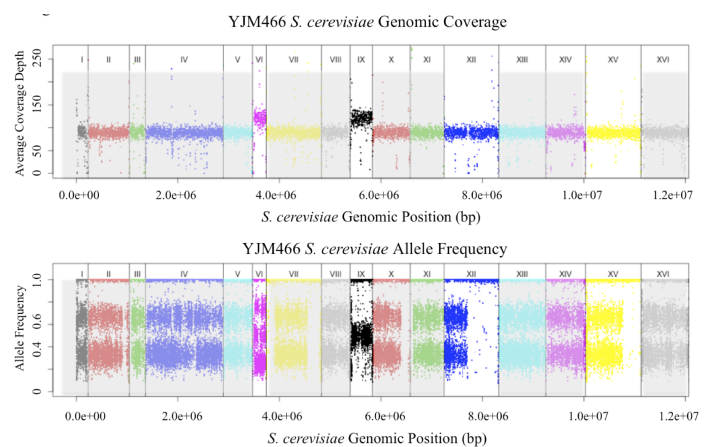
**Figure 5** Top). Correlation between whole chromosome aneuploid events observed (Y-axis) and chromosomal size (X-axis). Bottom). Correlation between partial chromosomal copy number changes observed (Y-axis) and chromosomal size (X-axis).



**Figure 6** Histogram of the total number of complete and partial aneuploid events observed within all 144 sequenced strains, by chromosome.



**Figure 7** Coverage and allele frequency plots for YJM1098. Chromosomes are lined in increasing order (X-axis), with total coverage in 1kb windows [top panel] or raw SNP allele frequencies [bottom panel] plotted (Y-axis). YJM1098 showed a copy number change in chromosome XII with corresponding shift in SNP allele frequency distribution, as well as no copy number change in chromosome VIII, but a allele frequency distribution suggesting recent chromosomal homozygosing event.



**Figure 8** Coverage and allele frequency plots for YJM466. Chromosomes are lined in increasing order (X-axis), with total coverage in 1kb windows [top panel] or raw SNP allele frequencies [bottom panel] plotted (Y-axis). YJM466 is a triploid strain that showed a copy number increase at chromosomes VI and IX. Chromosome IX showed a corresponding shift in SNP allele frequency distribution to that of 4n, but chromosome VI showed a shift in SNP allele frequency distribution that resembles 2n.



were weakly enriched for thiosulfate sulfurtransferase activity, whereas common duplications (45 genes in 3-67 strains each) were enriched for genes in membrane transport [Table 1]. 119 genes were significant for copy number loss (excluding haploid strains) [Table S6]. Rare losses (35 genes; 1-2 strains each) were depleted for genes involved in organelle components, whereas common losses (82 genes; 3-68 strains each) was enriched for membrane components [Table 1]. Overall, 90/215 genes were members in 2 or more of the gene loss, gene copy number gain, and gene copy number loss lists, and likely represent genes that are highly variable in copy number among yeast strains.

## DISCUSSION

There has been a renewal in interest in natural populations of ‘wild’ *S. cerevisiae*, and the adaptive events that lead to commercial yeast strains (Barrio *et al.* 2006; Blondin *et al.* 2009; Muller and McCusker 2009a; Sicard and Legras 2011; Wang *et al.* 2012; Liti 2015; Peter and Schacherer 2016). A survey extending beyond the classic European strains used in lab or commercial environments showed that there was larger diversity in undomesticated *S. cerevisiae* strains from New Zealand (Gayeviskiy and Goddard 2012). Another study on Chinese yeast strains found twice the amount of diversity as compared to previously surveyed strains, suggesting that Chinese yeasts may be a candidate source population (Wang *et al.* 2012). Geography thus plays a large part in wild yeast population structure, in stark contrast with commercial strains that have been spread across far reaching human communities through human activities (Goddard *et al.* 2010). To understand what differentiates virulent, commensal, wild, and commercially ‘safe’ strains in *S. cerevisiae*, it is necessary to understand the genetics of pathogenic yeast within this ever-growing context of non-clinical strains.

This task is hampered by the fact that a strains’ source substrate is a poor indication of whether it is truly a virulent, pathogenic clinical strain. The process of isolating *S. cerevisiae* colonies is often heavily biased towards strains that survive well in sugar rich conditions, and it is not always clear if *S. cerevisiae* is the dominant species occupying those substrates, or is even particularly well adapted to them (Goddard and Greig 2015). Earlier surveys of the *S. cerevisiae* population structure and phenotypic diversity suggested that the shift from non-clinical to clinical yeast occurred multiple times, ruling out a single selective event that would have left clear genetic signatures. Instead, clinical strains isolated from human patients do not cluster into genetically distinct groups (Liti *et al.* 2009; Schacherer *et al.* 2009; Muller and McCusker 2011), and many remain genetically most similar to commercial strains (likely their source populations) rather than other clinical strains (De Llanos *et al.* 2006; de Llanos *et al.* 2006; MacKenzie *et al.* 2008). The fact that multiple origins of clinical yeast strains have been found suggest a scenario where transitions are common, originate from multiple genetic backgrounds, and are unlikely to share a single adaptive strategy, if they are adapted to a pathogenic lifestyle at all. To fully study the transition of non-pathogenic yeast to a pathogenic lifestyle, either models are needed that can select for clinical behavior in non-clinical strains, to determine precise changes that might afford this change in lifestyle, or thousands of clinical and non-clinical yeast strains must be sequenced for a comprehensive analysis. Here, we start the process by re-sequencing 144 yeast strains, including 132 clinical isolates, from a variety of geographical locations and tissue substrates. We find that earlier results hold over this larger set of clinical strains. Neighbor-joining tree and Structure results both confirm that many clinical isolates are more closely related to specific commercial strains instead of

each other. We find evidence that there were likely independent pathogenic invasions from each of lab, sake, wine, and baking yeast strains. Within the rest of the clinical strains, there was no clear segregation of substrate tissue or geographical origin, and closely related strains often include isolates from entirely different tissue types and locations. It is possible that some of these strains may have been isolated from different tissues in the same patient, although we do not have access to the clinical data to confirm or refute that. It is also worth noting that due to considerations of comparability across datasets, we only used SNPs found and sufficiently covered in all datasets, which may bias the analysis towards relatively common polymorphisms.

Some clues about how yeast strains have adapted to drastic changes and stresses in new environments come from our understanding of how *S. cerevisiae* adapted to commercial fermentation (Dequin and Casaregola 2011). Functionally, whole genome or whole chromosome copy number changes can be beneficial under certain extreme conditions in yeast. Chromosome VIII, which carries the *CUP1-1* and *CUP1-2* genes, consistently increases in copy number in a copper-rich environment (Whittaker *et al.* 1988), and in general yeast appears to have the ability to maintain individual chromosomal copy numbers (Waghmare and Bruschi 2005). Chromosome IX, the most often observed aneuploid in the clinical strains, is especially common in aneuploid state, even compared to other chromosomes (II, III and VIII) that are comparably stable as aneuploids (Zhu *et al.* 2012). These structural variants have been suggested to provide one-step reversible short cuts when adapting to a harsh new habitat (Hittinger and Carroll 2007; Coyle and Kroll 2008; Rancati *et al.* 2008; Querol and Bond 2009; Stambuk *et al.* 2009; Muller and McCusker 2009a; Berman 2016), and a similar process may be involved in clinical yeast adaptation going by the high frequency of polyploid and aneuploid strains. Association between the two also agrees well with the thinking that polyploid genomes are less stable than diploids/haploids, making them important drivers of rapid adaptation (Hickman *et al.* 2015; Selmecki *et al.* 2015). More importantly, aneuploidies are highly unstable if not maintained by strong selective forces (Zhu *et al.* 2012; Kumaran *et al.* 2013; Zhu *et al.* 2014), and high aneuploidy rates could be indications that this is a selectable trait that may have been beneficial during the transition to pathogenic lifestyle (Hirakawa *et al.* 2015), and would be worth detailed phenotypic studies, especially in clades such as the Baker’s/commercial strains.

It must be noted that the majority of the clinical strains represented in our dataset were of North American and European origin. For a significant number of the clinical strains collected, we were unable to confidently identify the source population. Many of these strains are hybrids of known populations, but some may represent the genetic diversity of uncharacterized commercial or environmental *S. cerevisiae*. In addition, the lack of clinical strains that cluster with strains collected from Malaysia, Hawaii, and West Africa could indicate sampling bias. The role of increased human contact in opportunistic invasions may favor wine and baking strains, or it could indicate genetic predisposition of domesticated strains to transition to pathogenic lifestyles, perhaps precisely due to their genetic predisposition to easily adapt to drastically different carbon sources. More likely, it could be a combination of the two. Clinical strains thus present a selective view into a subset of medically relevant *S. cerevisiae*. More clinical strains collected from other parts of the world will help understand if clinical strains regularly colonize human hosts from non-commercial sources, or are mainly able to do so as a side effect of domestication prolonging exposure and opportunity.

**Table 1: Deleted genes (1-7 strains)**

<b>Biological Processes</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO biological process complete					
vitamin B6 biosynthetic process (GO:0042819)	<u>4</u>	0.06	> 5	+	1.00E-03
pyridoxine metabolic process (GO:0008614)	<u>4</u>	0.06	> 5	+	1.00E-03
vitamin B6 metabolic process (GO:0042816)	<u>4</u>	0.07	> 5	+	1.60E-03
cellular macromolecule metabolic process (GO:0044260)	<u>4</u>	19.42	0.21	-	1.95E-03
pyridoxal phosphate biosynthetic process (GO:0042823)	<u>4</u>	0.09	> 5	+	3.53E-03
pyridoxal phosphate metabolic process (GO:0042822)	<u>4</u>	0.09	> 5	+	3.53E-03
aldehyde biosynthetic process (GO:0046184)	<u>4</u>	0.09	> 5	+	3.53E-03
cellular component organization or biogenesis (GO:0071840)	<u>3</u>	17.35	< 0.2	-	3.83E-03
macromolecule metabolic process (GO:0043170)	<u>5</u>	20.23	0.25	-	4.30E-03
<b>Molecular Function</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO molecular function complete					
amine-lyase activity (GO:0016843)	<u>4</u>	0.05	> 5	+	1.71E-04
pyridoxal 5'-phosphate synthase (GO:0036381)	<u>4</u>	0.05	> 5	+	1.71E-04
lyase activity (GO:0016829)	<u>8</u>	0.88	> 5	+	2.55E-03
carbon-nitrogen lyase activity (GO:0016840)	<u>4</u>	0.11	> 5	+	4.84E-03
<b>Cellular Component</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO cellular component complete					
intracellular (GO:0005622)	<u>24</u>	38.94	0.62	-	5.74E-03
intracellular part (GO:0044424)	<u>24</u>	38.81	0.62	-	6.98E-03
macromolecular complex (GO:0032991)	<u>4</u>	16.95	0.24	-	1.40E-02
intrinsic component of membrane (GO:0031224)	<u>24</u>	11.14	2.15	+	3.96E-02

**Deleted genes (>8 strains)**

<b>Biological Processes</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO biological process complete					
maltose metabolic process (GO:0000023)	<u>3</u>	0.06	> 5	+	4.34E-02
<b>Molecular Function</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO molecular function complete					
oligosaccharide transmembrane transporter activity (GO:0015157)	<u>3</u>	0.01	> 5	+	2.92E-04
disaccharide transmembrane transporter activity (GO:0015154)	<u>3</u>	0.01	> 5	+	2.92E-04
maltose transmembrane transporter activity (GO:0005363)	<u>3</u>	0.01	> 5	+	2.92E-04
alpha-glucoside transmembrane transporter activity (GO:0015151)	<u>3</u>	0.02	> 5	+	6.91E-04
glucoside transmembrane transporter activity (GO:0042947)	<u>3</u>	0.02	> 5	+	6.91E-04
sugar transmembrane transporter activity (GO:0051119)	<u>4</u>	0.08	> 5	+	1.56E-03
carbohydrate transmembrane transporter activity (GO:0015144)	<u>4</u>	0.1	> 5	+	2.71E-03
carbohydrate transporter activity (GO:1901476)	<u>4</u>	0.1	> 5	+	2.71E-03
cadmium ion binding (GO:0046870)	<u>2</u>	0.01	> 5	+	3.31E-02
<b>Cellular Component</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO cellular component complete					
plasma membrane (GO:0005886)	<u>11</u>	2	> 5	+	1.08E-03
cell periphery (GO:0071944)	<u>11</u>	2.86	3.85	+	3.32E-02

**Duplicated genes (1-2 strains)**

<b>Molecular Function</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO molecular function complete					
thiosulfate sulfurtransferase activity (GO:0004792)	<u>3</u>	0.05	> 5	+	1.45E-02

**Duplicated genes (>2 strains)**

<b>Biological Processes</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO biological process complete					
transmembrane transport (GO:0055085)	<u>13</u>	2.8	4.65	+	4.08E-03
cellular component organization or biogenesis (GO:0071840)	<u>3</u>	15.01	< 0.2	-	4.30E-02
<b>Molecular Function</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO molecular function complete					
transporter activity (GO:0005215)	<u>14</u>	3.04	4.61	+	8.68E-04
sucrose alpha-glucosidase activity (GO:0004575)	<u>3</u>	0.04	> 5	+	9.93E-03
beta-fructofuranosidase activity (GO:0004564)	<u>3</u>	0.04	> 5	+	9.93E-03
transmembrane transporter activity (GO:0022857)	<u>11</u>	2.39	4.61	+	1.69E-02
glucosidase activity (GO:0015926)	<u>4</u>	0.19	> 5	+	4.03E-02
<b>Cellular Component</b>					
	#	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO cellular component complete					
macromolecular complex (GO:0032991)	<u>3</u>	14.67	0.2	-	2.34E-02
organelle (GO:0043226)	<u>16</u>	29.44	0.54	-	2.71E-02
intracellular organelle (GO:0043229)	<u>16</u>	29.37	0.54	-	2.97E-02
organelle part (GO:0044422)	<u>7</u>	19.76	0.35	-	3.36E-02
plasma membrane (GO:0005886)	<u>12</u>	3.21	3.74	+	3.47E-02
intracellular organelle part (GO:0044446)	<u>7</u>	19.68	0.36	-	3.70E-02



## Loss of copy genes (1-2 strains)

<b>Biological Processes</b>					
	<u>#</u>	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO biological process complete					
cellular component organization or biogenesis (GO:0071840)	<u>1</u>	11.68	< 0.2	-	2.25E-02
cellular macromolecule metabolic process (GO:0044260)	<u>2</u>	13.07	< 0.2	-	3.27E-02
<b>Molecular Function</b>					
	<u>#</u>	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO molecular function complete					
oligo-1,6-glucosidase activity (GO:0004574)	<u>3</u>	0.03	> 5	+	2.68E-03
acid phosphatase activity (GO:0003993)	<u>3</u>	0.05	> 5	+	1.54E-02
alpha-glucosidase activity (GO:0090599)	<u>3</u>	0.06	> 5	+	2.79E-02
<b>Cellular Component</b>					
	<u>#</u>	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO cellular component complete					
organelle (GO:0043226)	<u>6</u>	22.9	0.26	-	3.84E-06
intracellular organelle (GO:0043229)	<u>6</u>	22.85	0.26	-	4.30E-06
membrane-bounded organelle (GO:0043227)	<u>5</u>	21.18	0.24	-	1.46E-05
intracellular membrane-bounded organelle (GO:0043231)	<u>5</u>	21.13	0.24	-	1.62E-05
intracellular (GO:0005622)	<u>11</u>	26.21	0.42	-	5.17E-05
intracellular part (GO:0044424)	<u>11</u>	26.12	0.42	-	6.29E-05
organelle part (GO:0044422)	<u>4</u>	15.37	0.26	-	2.42E-02
intracellular organelle part (GO:0044446)	<u>4</u>	15.31	0.26	-	2.63E-02

## Loss of copy genes (>2 strains)

<b>Biological Processes</b>					
	<u>#</u>	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO biological process complete					
transmembrane transport (GO:0055085)	<u>20</u>	5.1	3.92	+	1.91E-04
cellular component organization or biogenesis (GO:0071840)	<u>8</u>	27.36	0.29	-	1.11E-03
maltose metabolic process (GO:0000023)	<u>5</u>	0.16	> 5	+	1.28E-03
macromolecule metabolic process (GO:0043170)	<u>12</u>	31.91	0.38	-	2.51E-03
regulation of biological process (GO:0050789)	<u>4</u>	19.5	0.21	-	7.50E-03
cellular macromolecule metabolic process (GO:0044260)	<u>12</u>	30.63	0.39	-	9.19E-03
cellular component organization (GO:0016043)	<u>7</u>	23.83	0.29	-	1.00E-02
ion transport (GO:0006811)	<u>15</u>	3.92	3.83	+	1.30E-02
regulation of cellular process (GO:0050794)	<u>4</u>	18.43	0.22	-	2.28E-02
thiamine-containing compound biosynthetic process (GO:0042724)	<u>4</u>	0.15	> 5	+	3.07E-02
thiamine biosynthetic process (GO:0009228)	<u>4</u>	0.15	> 5	+	3.07E-02
thiamine metabolic process (GO:0006772)	<u>4</u>	0.16	> 5	+	4.19E-02
<b>Molecular Function</b>					
	<u>#</u>	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO molecular function complete					
transmembrane transporter activity (GO:0022857)	<u>20</u>	4.35	4.6	+	7.55E-06
transporter activity (GO:0005215)	<u>21</u>	5.53	3.8	+	7.92E-05
substrate-specific transmembrane transporter activity (GO:0022891)	<u>17</u>	3.92	4.34	+	2.74E-04
substrate-specific transporter activity (GO:0022892)	<u>17</u>	4.75	3.58	+	3.80E-03
ion transmembrane transporter activity (GO:0015075)	<u>13</u>	2.86	4.55	+	5.03E-03
sugar transmembrane transporter activity (GO:0051119)	<u>5</u>	0.25	> 5	+	5.51E-03
oligosaccharide transmembrane transporter activity (GO:0015157)	<u>3</u>	0.04	> 5	+	7.76E-03
disaccharide transmembrane transporter activity (GO:0015154)	<u>3</u>	0.04	> 5	+	7.76E-03
maltose transmembrane transporter activity (GO:0005363)	<u>3</u>	0.04	> 5	+	7.76E-03
sodium-exporting ATPase activity, phosphorylative mechanism (GO:0008554)	<u>3</u>	0.04	> 5	+	7.76E-03
<b>Cellular Component</b>					
	<u>#</u>	<u>expected</u>	<u>Fold Enrichment</u>	<u>+/-</u>	<u>P value</u>
GO cellular component complete					
plasma membrane (GO:0005886)	<u>25</u>	5.85	4.27	+	1.88E-07
cell periphery (GO:0071944)	<u>27</u>	8.38	3.22	+	1.35E-05
intrinsic component of membrane (GO:0031224)	<u>40</u>	17.56	2.28	+	2.51E-05
protein complex (GO:0043234)	<u>2</u>	19.2	< 0.2	-	6.80E-05
integral component of membrane (GO:0016021)	<u>38</u>	16.82	2.26	+	9.03E-05
membrane part (GO:0044425)	<u>40</u>	20.68	1.93	+	2.37E-03



It is undeniable that there is a large environmental shift when yeasts move into a pathogenic lifestyle. However, the corresponding genetic changes that resulted from these selective pressures are complex. While our knowledge of yeast diversity, complexity, and ecological traits are expanding, there remains a need for the whole genome deep sequencing of more yeast strains (Peter and Schacherer 2016), as well as the better documentation and classification of what exactly defines a strain as pathogenic and virulent. Better means of analyzing genomic data with variable ploidy across and within individuals will also be important in the study of *S. cerevisiae* isolates. While our dataset is limited in the types of analysis that can be conducted, in contrast to earlier studies on haploid sequencing data such as the 100-genomes from (Strope *et al.* 2015), it provides information on a different scale and is complimentary to our understanding of the many levels of genetic variation available to *S. cerevisiae*. Such information, combined with human host traits and immune information, may provide the comprehensive picture required to tease apart yeast pathogenicity.

## LITERATURE CITED

- Albertin, W., P. Marullo, M. Aigle, A. Bourgeois, M. Bely, C. Dillmann, D. De Vienne, and D. Sicard, 2009 Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. *Journal of evolutionary biology* **22**: 2157–2170.
- Aucott, J. N., J. Fayen, H. Grossnicklas, A. Morrissey, M. M. Lederman, and R. A. Salata, 1990 Invasive infection with *Saccharomyces cerevisiae*: report of three cases and review. *Review of Infectious Diseases* **12**: 406–411.
- Barrio, E., S. S. González, A. Arias, C. Belloch, and A. Querol, 2006 Molecular mechanisms involved in the adaptive evolution of industrial yeasts. In *Yeasts in Food and Beverages*, pp. 153–174, Springer.
- Berman, J., 2016 Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS yeast research* **16**: fow020.
- Bidene, C., B. Blondin, S. Dequin, and F. Vezinhet, 1992 Analysis of the chromosomal dna polymorphism of wine strains of *Saccharomyces cerevisiae*. *Current genetics* **22**: 1–7.
- Biggins, S., N. Bhalla, A. Chang, D. L. Smith, and A. W. Murray, 2001 Genes involved in sister chromatid separation and segregation in the budding yeast *Saccharomyces cerevisiae*. *Genetics* **159**: 453–470.
- Blondin, B., S. Dequin, A. Querol, and J.-L. Legras, 2009 Genome of *Saccharomyces cerevisiae* and related yeasts. In *Biology of Microorganisms on Grapes, in Must and in Wine*, pp. 361–378, Springer.
- Carbon, S., A. Ireland, C. J. Mungall, S. Shu, B. Marshall, S. Lewis, W. P. W. Group, *et al.*, 2009 Amigo: online access to ontology and annotation data. *Bioinformatics* **25**: 288–289.
- Carro, D., E. Bartra, and B. Piña, 2003 Karyotype rearrangements in a wine yeast strain by rad52-dependent and rad52-independent mechanisms. *Applied and environmental microbiology* **69**: 2161–2165.
- Cassone, M., P. Serra, F. Mondello, A. Girolamo, S. Scafetti, E. Pistella, and M. Venditti, 2003 Outbreak of *Saccharomyces cerevisiae* subtype bouldarii fungemia in patients neighboring those treated with a probiotic preparation of the organism. *Journal of clinical microbiology* **41**: 5340–5343.
- Cavaliere, D., P. E. McGovern, D. L. Hartl, R. Mortimer, and M. Polsinelli, 2003 Evidence for *S. cerevisiae* fermentation in ancient wine. *Journal of molecular evolution* **57**: S226–S232.
- Cheng, E., J. A. Vaisica, J. Ou, A. Baryshnikova, Y. Lu, F. P. Roth, and G. W. Brown, 2012 Genome rearrangements caused by depletion of essential dna replication proteins in *Saccharomyces cerevisiae*. *Genetics* **192**: 147–160.
- Cho, U.-S. and S. C. Harrison, 2012 Ndc10 is a platform for inner kinetochore assembly in budding yeast. *Nature structural & molecular biology* **19**: 48–55.
- Clemons, K., L. Hanson, and D. Stevens, 1996 Colony phenotype switching in clinical and non-clinical isolates of *Saccharomyces cerevisiae*. *Journal of Medical and Veterinary Mycology* **34**: 259–264.
- Clemons, K. V., J. H. McCusker, R. W. Davis, and D. A. Stevens, 1994 Comparative pathogenesis of clinical and nonclinical isolates of *Saccharomyces cerevisiae*. *Journal of Infectious Diseases* **169**: 859–867.
- Coyle, S. and E. Kroll, 2008 Starvation induces genomic rearrangements and starvation-resilient phenotypes in yeast. *Molecular biology and evolution* **25**: 310–318.
- Cromie, G. A. and A. M. Dudley, 2015 Aneuploidy: Tolerating tolerance. *Current Biology* **25**: R771–R773.
- de Llanos, R., M. T. Fernández-Espinar, and A. Querol, 2006 A comparison of clinical and food *Saccharomyces cerevisiae* isolates on the basis of potential virulence factors. *Antonie Van Leeuwenhoek* **90**: 221–231.
- De Llanos, R., S. Llopis, G. Molero, A. Querol, C. Gil, and M. Fernandez-Espinar, 2011 In vivo virulence of commercial *Saccharomyces cerevisiae* strains with pathogenicity-associated phenotypical traits. *International journal of food microbiology* **144**: 393–399.
- De Llanos, R., A. Querol, J. Pemán, M. Gobernado, and M. T. Fernández-Espinar, 2006 Food and probiotic strains from the *Saccharomyces cerevisiae* species as a possible origin of human systemic infections. *International journal of food microbiology* **110**: 286–290.
- de Llanos, R., A. Querol, A. M. Planes, and M. T. Fernández-Espinar, 2004 Molecular characterization of clinical *Saccharomyces cerevisiae* isolates and their association with non-clinical strains. *Systematic and applied microbiology* **27**: 427–435.
- Dequin, S. and S. Casaregola, 2011 The genomes of fermentative *Saccharomyces*. *Comptes rendus biologiques* **334**: 687–693.
- Dougherty, S. H. and R. L. Simmons, 1982 Postoperative peritonitis caused by *Saccharomyces cerevisiae*. *Archives of Surgery* **117**: 248–249.
- Doyle, M., L. Pickering, N. O'Brien, K. Hoots, *et al.*, 1990 *Saccharomyces cerevisiae* infection in a patient with acquired immunodeficiency syndrome. *Pediatric Infectious Disease Journal* **9**: 850–851.
- Dunn, B., C. Richter, D. J. Kvitek, T. Pugh, and G. Sherlock, 2012 Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome research* **22**: 908–924.
- Enache-Angoulvant, A. and C. Hennequin, 2005 Invasive *Saccharomyces* infection: a comprehensive review. *Clinical Infectious Diseases* **41**: 1559–1568.
- Erdem, H., M. Cetin, T. Timuroglu, A. Cetin, O. Yanar, and A. Pahsa, 2003 Identification of yeasts in public hospital primary care patients with or without clinical vaginitis. *Australian and New Zealand journal of obstetrics and gynaecology* **43**: 312–316.
- Falush, D., M. Stephens, and J. K. Pritchard, 2003 Inference of population structure using multilocus genotype data: linked loci

- and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- Falush, D., M. Stephens, and J. K. Pritchard, 2007 Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes* **7**: 574–578.
- Farrell, A. E., R. J. Plevin, B. T. Turner, A. D. Jones, M. O'hare, and D. M. Kammen, 2006 Ethanol can contribute to energy and environmental goals. *Science* **311**: 506–508.
- Gasch, A. P., J. Hose, M. A. Newton, M. Sardi, M. Yong, and Z. Wang, 2016 Further support for aneuploidy tolerance in wild yeast and effects of dosage compensation on gene copy-number evolution. *eLife* **5**: e14409.
- Gayevskiy, V. and M. R. Goddard, 2012 Geographic delineations of yeast communities and populations associated with vines and wines in new zealand. *The ISME journal* **6**: 1281–1290.
- Goddard, M. R., N. Anfang, R. Tang, R. C. Gardner, and C. Jun, 2010 A distinct population of *saccharomyces cerevisiae* in new zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Environmental Microbiology* **12**: 63–73.
- Goddard, M. R. and D. Greig, 2015 *Saccharomyces cerevisiae*: a nomadic yeast with no niche? *FEMS yeast research* **15**: fov009.
- Goh, P.-Y. and J. V. Kilmartin, 1993 *Ndc10*: a gene involved in chromosome segregation in *saccharomyces cerevisiae*. *The Journal of Cell Biology* **121**: 503–512.
- Hennequin, C., C. Kauffmann-Lacroix, A. Jobert, J. Viard, C. Ricour, J. Jacquemin, and P. Berche, 2000 Possible role of catheters in *saccharomyces boulardii* fungemia. *European Journal of Clinical Microbiology and Infectious Diseases* **19**: 16–20.
- Hickman, M. A., C. Paulson, A. Dudley, and J. Berman, 2015 Paraxenial ploidy reduction drives population heterogeneity through random and transient aneuploidy in *candida albicans*. *Genetics* **200**: 781–794.
- Hirakawa, M. P., D. A. Martinez, S. Sakthikumar, M. Z. Anderson, A. Berlin, S. Gujja, Q. Zeng, E. Zisson, J. M. Wang, J. M. Greenberg, *et al.*, 2015 Genetic and phenotypic intra-species variation in *candida albicans*. *Genome research* **25**: 413–425.
- Hittinger, C. T. and S. B. Carroll, 2007 Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**: 677–681.
- Hose, J., C. M. Yong, M. Sardi, Z. Wang, M. A. Newton, and A. P. Gasch, 2015 Dosage compensation can buffer copy-number variation in wild yeast. *Elife* **4**: e05462.
- Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard, 2009 Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources* **9**: 1322–1332.
- James, T. C., J. Usher, S. Campbell, and U. Bond, 2008 Lager yeasts possess dynamic genomes that undergo rearrangements and gene amplification in response to stress. *Current genetics* **53**: 139–152.
- Klingberg, T. D., U. Lesnik, N. Arneborg, P. Raspor, and L. Jespersen, 2008 Comparison of *saccharomyces cerevisiae* strains of clinical and nonclinical origin by molecular typing and determination of putative virulence traits. *FEMS yeast research* **8**: 631–640.
- Kumaran, R., S.-Y. Yang, and J.-Y. Leu, 2013 Characterization of chromosome stability in diploid, polyploid and hybrid yeast cells. *PLoS One* **8**: e68094.
- Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai, 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**: 571–574.
- Legras, J.-L., D. Merdinoglu, J. CORNUET, and F. Karst, 2007 Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular ecology* **16**: 2091–2102.
- Letunic, I. and P. Bork, 2007 Interactive tree of life (itol): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.
- Lherm, T., C. Monet, B. Nougère, M. Soulier, D. Larbi, C. Le Gall, D. Caen, and C. Malbrunot, 2002 Seven cases of fungemia with *saccharomyces boulardii* in critically ill patients. *Intensive care medicine* **28**: 797–801.
- Li, H. and R. Durbin, 2009 Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**: 1754–1760.
- Li, R., Y. Li, X. Fang, H. Yang, J. Wang, K. Kristiansen, and J. Wang, 2009 Snp detection for massively parallel whole-genome resequencing. *Genome research* **19**: 1124–1132.
- Liti, G., 2015 The fascinating and secret wild life of the budding yeast *s. cerevisiae*. *Elife* **4**: e05835.
- Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts, S. A. James, R. P. Davey, I. N. Roberts, A. Burt, V. Koufopanou, *et al.*, 2009 Population genomics of domestic and wild yeasts. *Nature* **458**: 337–341.
- MacKenzie, D. A., M. Defernez, W. B. Dunn, M. Brown, L. J. Fuller, S. R. de Herrera, A. Günther, S. A. James, J. Eagles, M. Philo, *et al.*, 2008 Relatedness of medically important strains of *saccharomyces cerevisiae* as revealed by phylogenetics and metabolomics. *Yeast* **25**: 501–512.
- Malgoire, J., S. Bertout, F. Renaud, J. Bastide, and M. Mallie, 2005 Typing of *saccharomyces cerevisiae* clinical strains by using microsatellite sequence polymorphism. *Journal of clinical microbiology* **43**: 1133–1137.
- McCullough, M. J., K. V. Clemons, C. Farina, J. H. McCusker, and D. A. Stevens, 1998 Epidemiological investigation of vaginal *saccharomyces cerevisiae* isolates by a genotypic method. *Journal of clinical microbiology* **36**: 557–562.
- McCusker, J. H., K. V. Clemons, D. A. Stevens, and R. W. Davis, 1994 *Saccharomyces cerevisiae* virulence phenotype as determined with *cd-1* mice is associated with the ability to grow at 42 degrees c and form pseudohyphae. *Infection and Immunity* **62**: 5447–5455.
- McGovern, P. E., U. Hartung, V. R. Badler, D. L. Glusker, and L. J. Exner, 1997 The beginnings of winemaking and viniculture in the ancient near east and egypt. *Expedition* **39**: 3–21.
- McGovern, P. E., J. Zhang, J. Tang, Z. Zhang, G. R. Hall, R. A. Moreau, A. Nuñez, E. D. Butrym, M. P. Richards, C.-s. Wang, *et al.*, 2004 Fermented beverages of pre-and proto-historic china. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 17593–17598.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, *et al.*, 2010 The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome research* **20**: 1297–1303.
- Muller, L. A., J. E. Lucas, D. R. Georgianna, and J. H. McCusker, 2011 Genome-wide association analysis of clinical vs. nonclinical origin provides insights into *saccharomyces cerevisiae* pathogenesis. *Molecular ecology* **20**: 4085–4097.
- Muller, L. A. and J. McCusker, 2009a Microsatellite analysis of genetic diversity among clinical and nonclinical *saccharomyces cerevisiae* isolates suggests heterozygote advantage in clinical environments. *Molecular ecology* **18**: 2779–2786.
- Muller, L. A. and J. H. McCusker, 2009b A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *saccharomyces cerevisiae*. *FEMS yeast research*

- 9: 143–152.
- Muller, L. A. and J. H. McCusker, 2011 Nature and distribution of large sequence polymorphisms in *saccharomyces cerevisiae*. *FEMS yeast research* **11**: 587–594.
- Muñoz, P., E. Bouza, M. Cuenca-Estrella, J. M. Eiros, M. J. Pérez, M. Sánchez-Somolinos, C. Rincón, J. Hortal, and T. Peláez, 2005 *Saccharomyces cerevisiae* fungemia: an emerging infectious disease. *Clinical Infectious Diseases* **40**: 1625–1634.
- Mydlik, M., E. Tkacova, K. Szovenyiova, P. Mizla, and K. Derzsiova, 1996 *Saccharomyces cerevisiae* peritonitis complicating capd. *Peritoneal Dialysis International* **16**: 188–188.
- Ouspenski, I. I., S. J. Elledge, and B. Brinkley, 1999 New yeast genes important for chromosome integrity and segregation identified by dosage effects on genome stability. *Nucleic acids research* **27**: 3001–3008.
- Pavelka, N., G. Rancati, J. Zhu, W. D. Bradford, A. Saraf, L. Florens, B. W. Sanderson, G. L. Hattem, and R. Li, 2010 Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**: 321–325.
- Perapoch, J., A. Planes, A. Querol, V. Lopez, I. Martinez-Bendayan, R. Tormo, F. Fernandez, G. Peguero, and S. Salcedo, 2000 Fungemia with *saccharomyces cerevisiae* in two newborns, only one of whom had been treated with ultra-levura. *European Journal of Clinical Microbiology and Infectious Diseases* **19**: 468–470.
- Pérez-Torrado, R., S. Llopis, L. Jespersen, T. Fernández-Espinar, and A. Querol, 2012 Clinical *saccharomyces cerevisiae* isolates cannot cross the epithelial barrier in vitro. *International journal of food microbiology* **157**: 59–64.
- Peter, J. and J. Schacherer, 2016 Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast*.
- Piarroux, R., L. Millon, K. Bardonnnet, O. Vagner, and H. Koenig, 1999 Are live *saccharomyces cerevisiae* yeasts harmful to patients? *The Lancet* **353**: 1851–1852.
- Posteraro, B., M. Sanguinetti, L. Masucci, L. Romano, G. Morace, and G. Fadda, 2000 Reverse cross blot hybridization assay for rapid detection of pcr-amplified dna from *candida* species, *cryptococcus neoformans*, and *saccharomyces cerevisiae* in clinical samples. *Journal of clinical microbiology* **38**: 1609–1614.
- Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Querol, A. and U. Bond, 2009 The complex and dynamic genomes of industrial yeasts. *FEMS microbiology letters* **293**: 1–10.
- Rancati, G. and N. Pavelka, 2013 Karyotypic changes as drivers and catalyzers of cellular evolvability: a perspective from non-pathogenic yeasts. In *Seminars in cell & developmental biology*, volume 24, pp. 332–338, Elsevier.
- Rancati, G., N. Pavelka, B. Fleharty, A. Noll, R. Trimble, K. Walton, A. Perera, K. Staehling-Hampton, C. W. Seidel, and R. Li, 2008 Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* **135**: 879–893.
- Reimand, J., T. Arak, and J. Vilo, 2011 g: Profiler—a web server for functional interpretation of gene lists (2011 update). *Nucleic acids research* p. gkr378.
- Salonen, J., M. Richardson, K. Gallacher, J. Issakainen, H. Helenius, O.-P. Lehtonen, and J. Nikoskelainen, 2000 Fungal colonization of haematological patients receiving cytotoxic chemotherapy: emergence of azole-resistant *saccharomyces cerevisiae*. *Journal of Hospital Infection* **45**: 293–301.
- Sampaio, J. P. and P. Gonçalves, 2008 Natural populations of *saccharomyces kudriavzevii* in portugal are associated with oak bark and are sympatric with *s. cerevisiae* and *s. paradoxus*. *Applied and environmental microbiology* **74**: 2144–2152.
- Samuel, D., 1996 Investigation of ancient egyptian baking and brewing methods by correlative microscopy. *Science* **273**: 488.
- Schacherer, J., J. A. Shapiro, D. M. Ruderfer, and L. Kruglyak, 2009 Comprehensive polymorphism survey elucidates population structure of *saccharomyces cerevisiae*. *Nature* **458**: 342–345.
- Selmecki, A. M., K. Dulmage, L. E. Cowen, J. B. Anderson, and J. Berman, 2009 Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet* **5**: e1000705.
- Selmecki, A. M., Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shores, A. L. Sorenson, S. De, R. Kishony, F. Michor, R. Dwell, *et al.*, 2015 Polyploidy can drive rapid adaptation in yeast. *Nature* **519**: 349–352.
- Sheppard, T. K., B. C. Hitz, S. R. Engel, G. Song, R. Balakrishnan, G. Binkley, M. C. Costanzo, K. S. Dalusag, J. Demeter, S. T. Hellerstedt, *et al.*, 2015 The *saccharomyces* genome database variant viewer. *Nucleic acids research* p. gkv1250.
- Sicard, D. and J.-L. Legras, 2011 Bread, beer and wine: yeast domestication in the *saccharomyces sensu stricto* complex. *Comptes rendus biologiques* **334**: 229–236.
- Smith, D., D. Metzgar, C. Wills, and J. Fierer, 2002 Fatal *saccharomyces cerevisiae* aortic graft infection. *Journal of clinical microbiology* **40**: 2691–2692.
- Sniegowski, P. D., P. G. Dombrowski, and E. Fingerhahn, 2002 *Saccharomyces cerevisiae* and *saccharomyces paradoxus* coexist in a natural woodland site in north america and display different levels of reproductive isolation from european conspecifics. *FEMS Yeast Research* **1**: 299–306.
- Stambuk, B. U., B. Dunn, S. L. Alves, E. H. Duval, and G. Sherlock, 2009 Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin b1 and b6 biosynthesis. *Genome research* **19**: 2271–2278.
- Stirling, P. C., M. S. Bloom, T. Solanki-Patil, S. Smith, P. Sipahimalani, Z. Li, M. Kofoed, S. Ben-Aroya, K. Myung, and P. Hieter, 2011 The complete spectrum of yeast chromosome instability genes identifies candidate cin cancer genes and functional roles for astra complex components. *PLoS Genet* **7**: e1002057.
- Strope, P. K., D. A. Skelly, S. G. Kozmin, G. Mahadevan, E. A. Stone, P. M. Magwene, F. S. Dietrich, and J. H. McCusker, 2015 The 100-genomes strains, an *s. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome research* **25**: 762–774.
- Sunshine, A. B., C. Payen, G. T. Ong, I. Liachko, K. M. Tan, and M. J. Dunham, 2015 The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biol* **13**: e1002155.
- Thorburn, R. R., C. Gonzalez, G. A. Brar, S. Christen, T. M. Carlile, N. T. Ingolia, U. Sauer, J. S. Weissman, and A. Amon, 2013 Aneuploid yeast strains exhibit defects in cell growth and passage through start. *Molecular biology of the cell* **24**: 1274–1289.
- Waghmare, S. K. and C. V. Bruschi, 2005 Differential chromosome control of ploidy in the yeast *saccharomyces cerevisiae*. *Yeast* **22**: 625–639.
- Wang, Q.-M., W.-Q. Liu, G. Liti, S.-A. Wang, and F.-Y. Bai, 2012 Surprisingly diverged populations of *saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular ecology* **21**: 5404–5417.
- Wei, W., J. H. McCusker, R. W. Hyman, T. Jones, Y. Ning, Z. Cao, Z. Gu, D. Bruno, M. Miranda, M. Nguyen, *et al.*, 2007 Genome sequencing and comparative analysis of *saccharomyces cerevisiae*



- strain yjm789. Proceedings of the National Academy of Sciences **104**: 12825–12830.
- Wheeler, R. T., M. Kupiec, P. Magnelli, C. Abeijon, and G. R. Fink, 2003 A *saccharomyces cerevisiae* mutant with increased virulence. Proceedings of the National Academy of Sciences **100**: 2766–2770.
- Whittaker, S., B. Rockmill, A. Blechl, D. Maloney, M. Resnick, and S. Fogel, 1988 The detection of mitotic and meiotic aneuploidy in yeast using a gene dosage selection system. Molecular and General Genetics MGG **215**: 10–18.
- Zerva, L., R. Hollis, and M. Pfaller, 1996 In vitro susceptibility testing and dna typing of *saccharomyces cerevisiae* clinical isolates. Journal of clinical microbiology **34**: 3031–3034.
- Zhu, J., N. Pavelka, W. D. Bradford, G. Rancati, and R. Li, 2012 Karyotypic determinants of chromosome instability in aneuploid budding yeast. PLoS Genet **8**: e1002719.
- 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**: E2310–E2318.
- Zörgö, E., K. Chwialkowska, A. B. Gjuvsland, E. Garrè, P. Sunnerhagen, G. Liti, A. Blomberg, S. W. Omholt, and J. Warringer, 2013 Ancient evolutionary trade-offs between yeast ploidy states **9**: 1–11.