

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

Yuan O. Zhu,^{*,†,‡} Gavin Sherlock,^{*,2} and Dmitri A. Petrov^{†,1,2}

^{*}Department of Genetics and [†]Department of Biology, Stanford University, California 94305, and [‡]Genome Institute of Singapore, Singapore 138672

ORCID IDs: 0000-0001-9419-6291 (Y.O.Z.); 0000-0002-1692-4983 (G.S.)

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 eight 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 $>2\times$ increase in aneuploidy rates as compared to other lineages. This dataset provides a rich source of information on 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

Yeast species classified under the *Saccharomyces sensu stricto* species complex, in particular *S. cerevisiae*, have played vital cultural and economic roles in human history (see Sicard and Legras 2011 for review). 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 (*e.g.*, Farrell *et al.* 2006).

With extensive human contact over nearly 7000 yrs, 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.* 1999; 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; Munoz *et al.* 2005; Enache-Angoulvant and Hennequin 2005; de Llanos *et al.* 2006a,b). 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 cases of fungemia, the most severe clinical manifestation of fungal infections

Copyright © 2016 Zhu *et al.*

doi: 10.1534/g3.116.029397

Manuscript received March 21, 2016; accepted for publication May 30, 2016; published Early Online June 13, 2016.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.029397/-/DC1

¹Corresponding author: Department of Biology, 371 Serra St., Stanford University, Stanford, CA 94305-5020. E-mail: dpetrov@stanford.edu

²These authors contributed equally to this work.

(Piarroux *et al.* 1999; Smith *et al.* 2002; Munoz *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; Munoz *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 immunocompromised, had underlying medical conditions, underwent recent antibiotic treatments, used intravenous catheters, or a combination thereof (Enache-Angoulvant and Hennequin 2005; Munoz *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 (Perez-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 nonclinical strains showed that some clinical strains appeared more virulent than nonclinical strains, but virulence phenotypes generally overlap between the two groups, suggesting that preexisting 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 nonclinical 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 2009a). 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 (see Dequin and Casaregola 2011 for review). 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, 2015; Pavelka *et al.* 2010; Rancati and Pavelka 2013; Thorburn *et al.* 2013; Zörgö *et al.* 2013). 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 $80\times$ per strain. We found that 32% of strains assayed contained >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 as 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. One hundred base pair paired-end Illumina sequencing libraries were prepared at The Genome Institute, Washington University School of Medicine, and run on Illumina GAII with 12 samples pooled per lane (TruSeq adapters AD001-AD012), to an average depth of $80\times$ 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 600 kb of the genome – these regions were annotated in the SGD database as simple repeats, centromeric regions, telomeric regions, or LTRs (SGD project – http://downloads.yeastgenome.org/curation/chromosomal_feature/SGD_features.tab, downloaded 4th August 2012) were excluded from analysis due to their susceptibility to mismapping and associated miscalls. For a full list of commands, refer to Supplemental Material and Zhu *et al.* (2014).

Aneuploidy calling, gene copy number, and GO term analysis

The aneuploidy calling process was similar to the method used in Zhu *et al.* (2014) and carried out by computing coverage depth of nonoverlapping 1 kb windows. Whole chromosome aneuploidies were called if average coverage of a chromosome differs $>35\%$ from other

Tree scale: 0.1

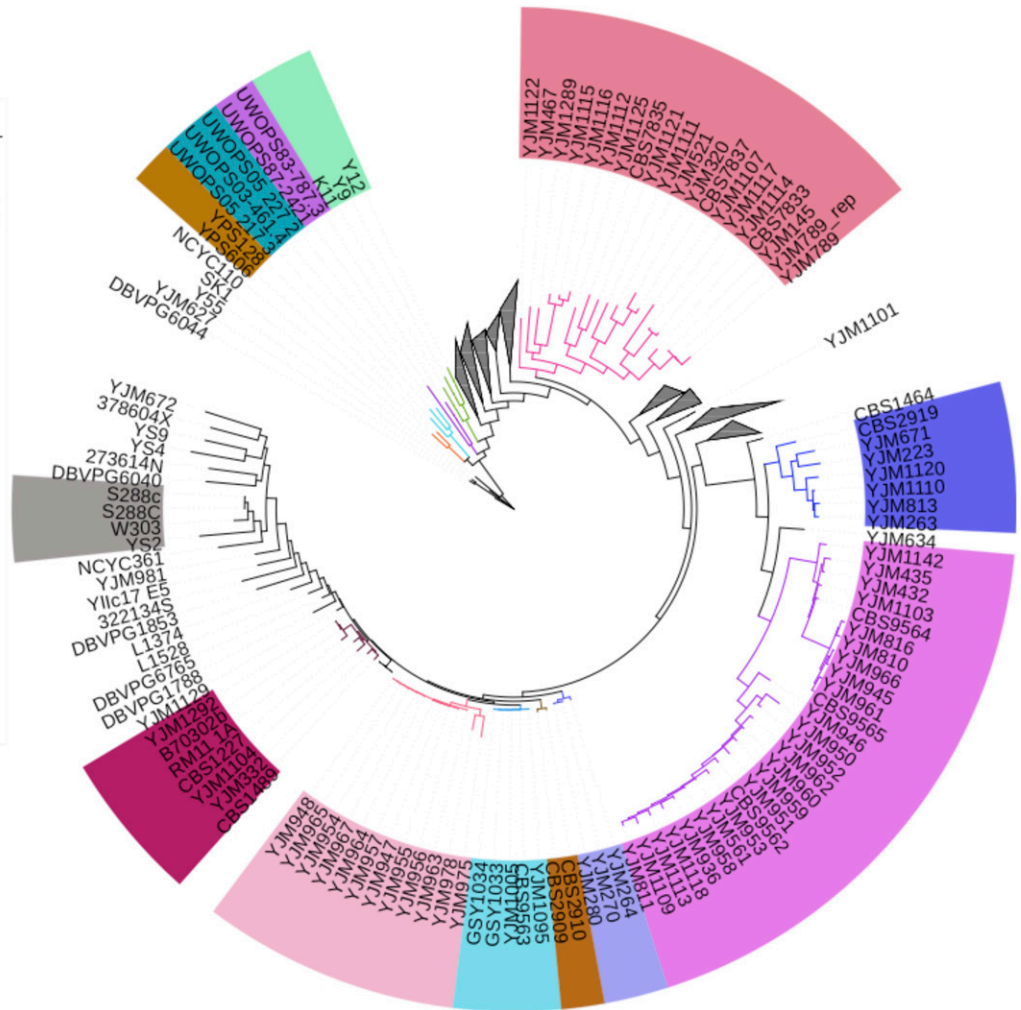


Figure 1 Neighbor-joining tree of 144 sequenced strains and 35 Liti *et al.* (2009) strains, based on SNP differences within the *S. cerevisiae* genome. Groups of strains that fall within a cluster containing noncommercial strains from a single known source are highlighted in color blocks. Clades containing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

chromosomes in the same line (likelihood $P < 0.001$, χ^2 -test), and visually confirmed through coverage traces (see Supplemental Material 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 single nucleotide polymorphisms (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 neighbor-joining (NJ) tree construction. The NJ 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 2006). Structure (version 2.3.2) analysis was run with independent allele frequencies under an admixture model, for a 15,000 burn in period and 5000 iterations of sampling.

Data availability

Sequencing data are available through the National Center for Biotechnology Information Sequence Read Archive, number SRP072079.

RESULTS

High coverage whole genome sequencing

One hundred and forty-five *S. cerevisiae* strains were cultured and sequenced in their natural ploidy states. Thirteen of these strains were from nonclinical sources; the remaining 132 came from clinical sources. Eight 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.* 2006a), 85 strains were part of a genetic diversity by

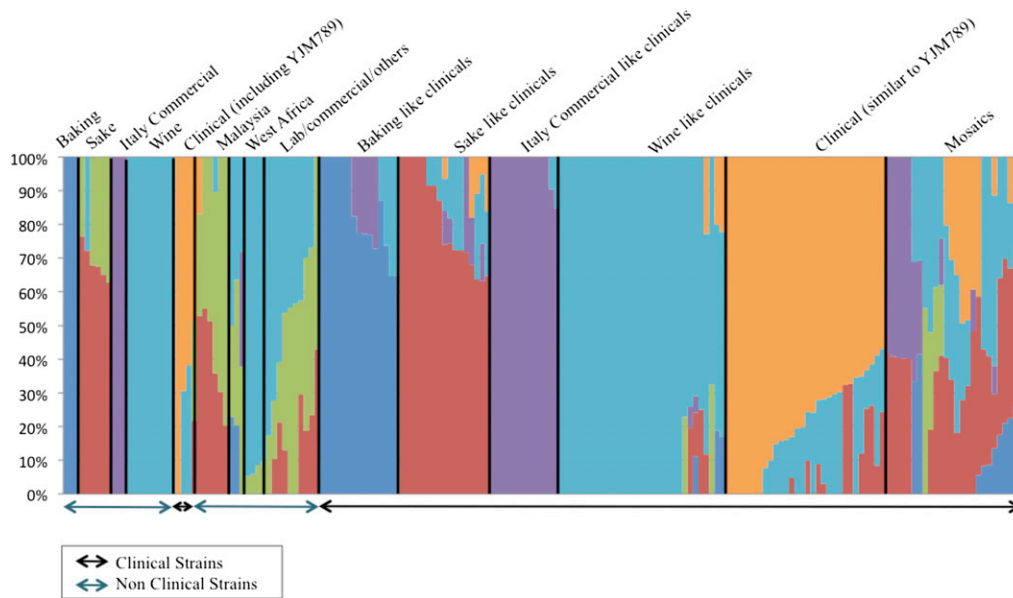


Figure 2 Structure assignment of source populations for all 180 strains. Strains are aligned from left to right (x-axis) and sorted by their percentage composition of each population (y-axis) as computed by Structure. Source populations are annotated with text above the figure. Clinical or non-clinical 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 5000 iterations of sampling.

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) (Supplemental Material, 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 $\sim 80\times$. As it is also known that both hybridization and introgression are not uncommon in clinical budding yeasts (Muller and McCusker 2009b; Strope *et al.* 2015), the fastq reads were mapped to a combined *Saccharomyces sensu stricto* genome that included sequences from *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, and *S. uvarum*. The S288C reference sequence was downloaded from the *Saccharomyces* Genome Database (Sheppard *et al.* 2016); all other species were downloaded from the *Saccharomyces sensu stricto* Resources site (www.saccharomyces-sustricto.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 (*Materials and 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. 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 SNPs were called across all non-S288C strains. Individual strains each carry between 36,786 and 97,758 variants (0.3–0.8%) as compared to S288C. Although the number of nonclinical strains was small, allele frequencies at diallelic positions could be compared between clinical and nonclinical

strains. Correlation between the clinical and nonclinical populations was high, with an r^2 value of 0.9346 (Figure S1). F_{st} values were calculated for every position, but values did not exceed 0.15, signifying that no SNPs were differentiated between clinical and nonclinical strains (Figure S2). Thus, no further analysis to identify clinical specific alleles was performed.

Population phylogeny recapitulates earlier studies

An 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's 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 probiotic 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 nonclinical 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 (*Materials and Methods*). The clinical strains were distinguished into five main populations in our analysis. The nonclinical strains that corresponded to these populations included baking strains, sake/soil strains, wine strains, and unspecified commercial strains from

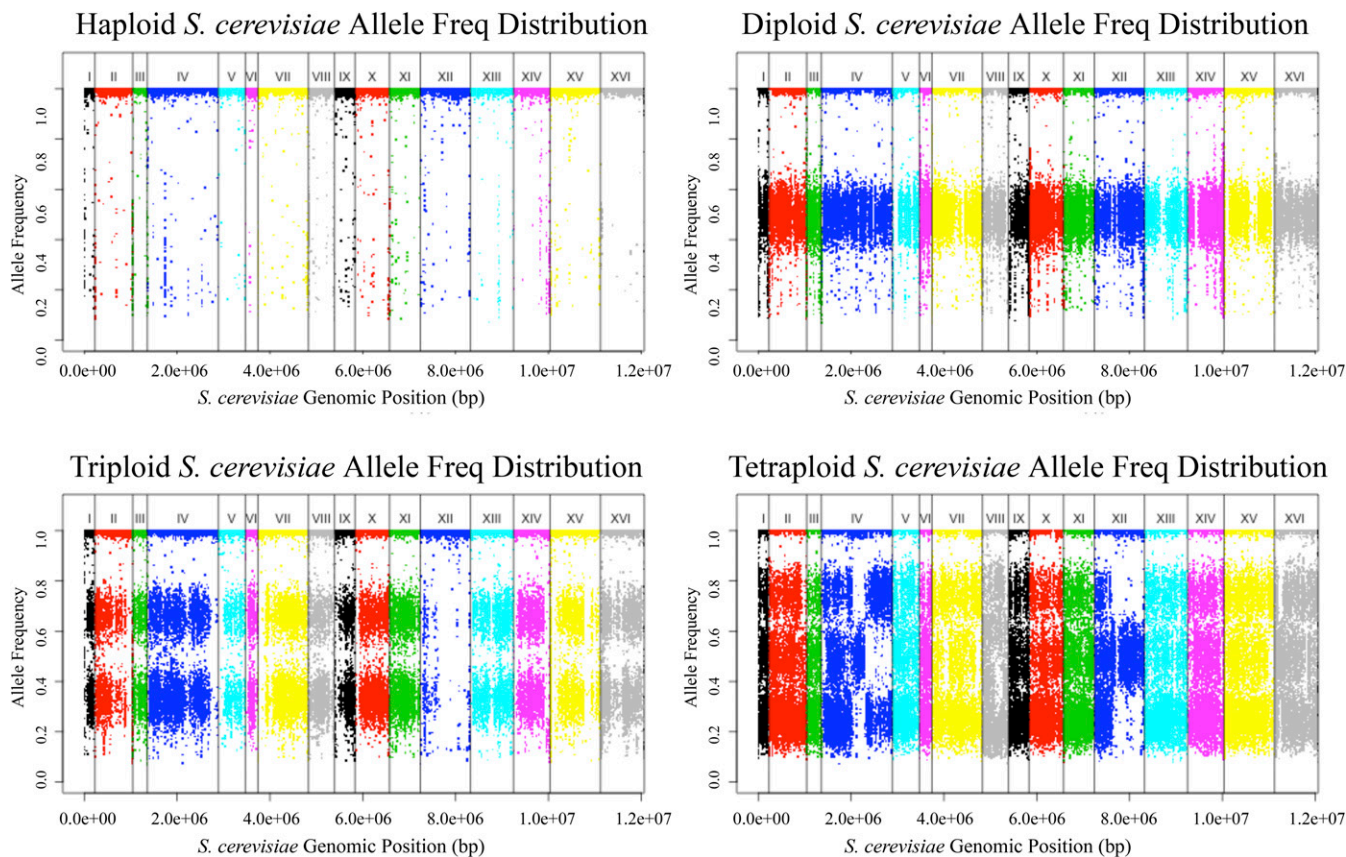


Figure 3 Top left: allele frequency plot of a typical haploid with SNPs at allele frequencies at 1. Top right: allele frequency plot of a typical diploid with SNPs at allele frequencies at 0.5 and 1. Bottom left: allele frequency plot of a typical triploid with SNPs at allele frequencies at 0.33, 0.67, 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 up in increasing order (x-axis), with raw SNP allele frequencies along the genome plotted (y-axis).

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; Strobe *et al.* 2015). The remainder of the clinical strains showed varying amounts of admixture from various population sources (Figure 2).

Introgressions 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 2009). 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 435 kb of XII and the first 364 kb of XIII. Coverage across the *S. cerevisiae* chromosomes showed complementary patterns, including lower copy numbers for the partial segments on chromosomes XII (first 450 kb) and XIII (first 390 kb), such that combined across both genomes, there are three copies of all chromosomes. The only exceptions were chromosome V, which appeared to have four copies, and chromosome II, which appeared to have a segmental duplication spanning the first 200 kb (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 2 kb and 40 kb in size. Eight hundred and eighteen such 1 kb windows were observed in CBS2909 and 778 such 1 kb 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). Two hundred and sixty-one genes were shared by both strains. Other yeast strains have also been previously found to contain introgressions from *S. paradoxus* (Strobe *et al.* 2015). *S. paradoxus* is often found in the same habitats 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

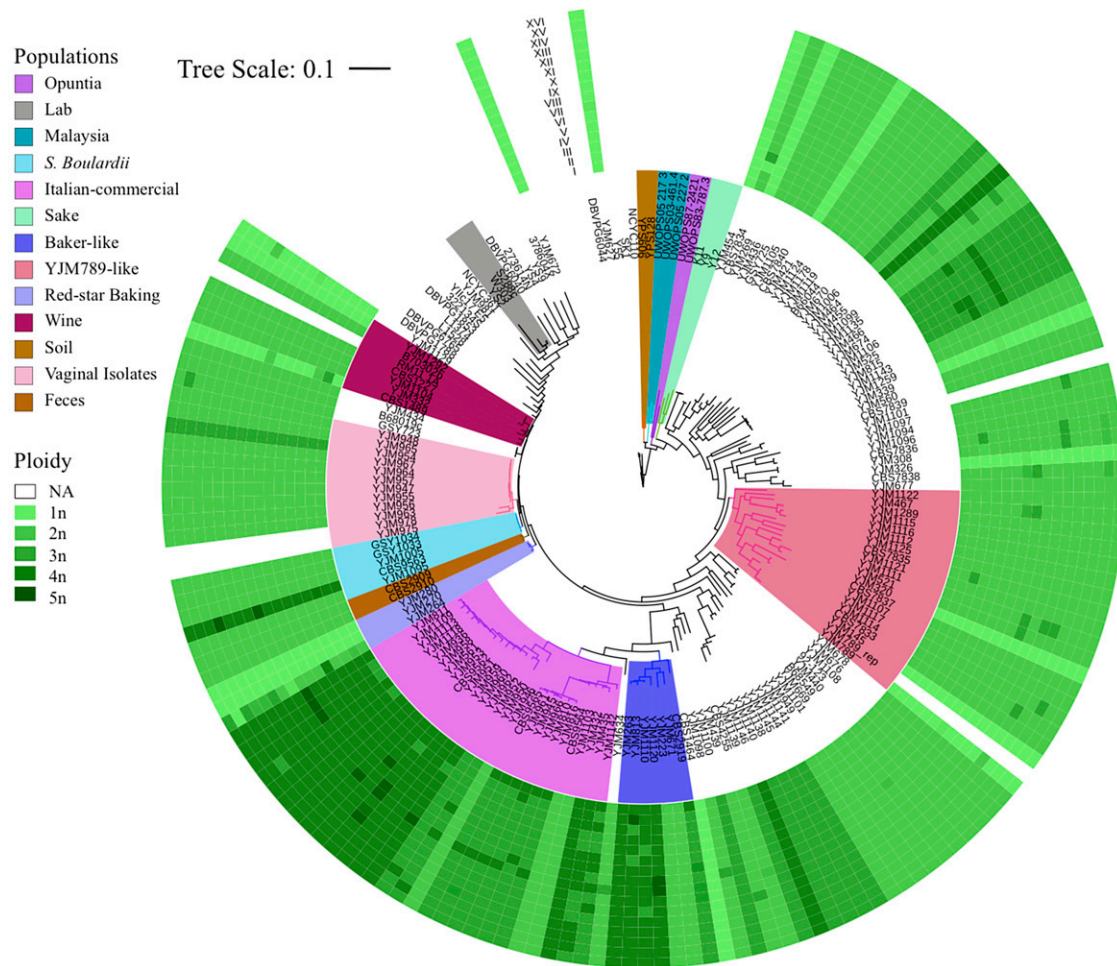


Figure 4 Heatmap of chromosomal copy number (0n–5n) for each of 144 strains aligned next to each strain on the phylogenetic tree. The 35 Liti 2009 strains do not have chromosomal copy number information panels (in white). The 16 chromosomes of *S. cerevisiae* are sorted in increasing order in an outwards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

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 homozygous 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 between 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 homozygosing 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 1 kb nonoverlapping windows across all 16 *S. cerevisiae* chromosomes. Fifty-three of 144 strains (36%) were observed to contain aneuploidies, a number nearly five 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

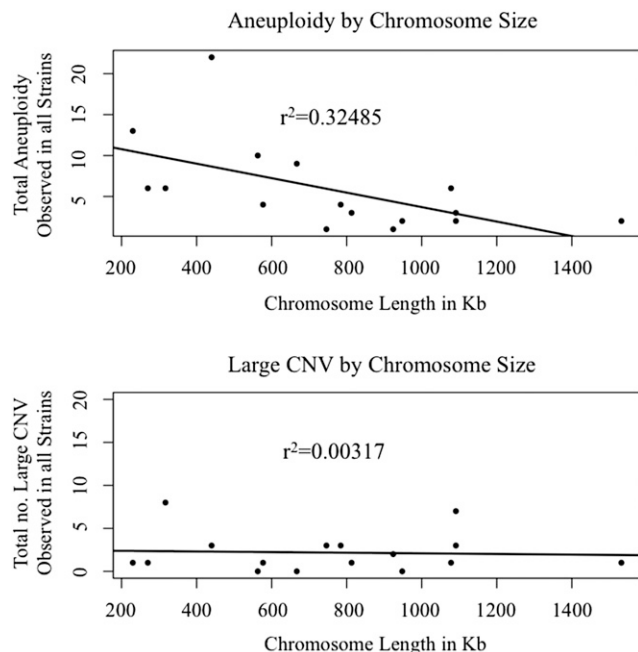


Figure 5 Top: correlation between whole chromosome aneuploid events observed (y-axis) and chromosomal size (x-axis). Bottom: correlation between partial chromosomal copy number changes observed (y-axis) and chromosomal size (x-axis).

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 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). Twenty-nine out of 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 nonsegregant 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

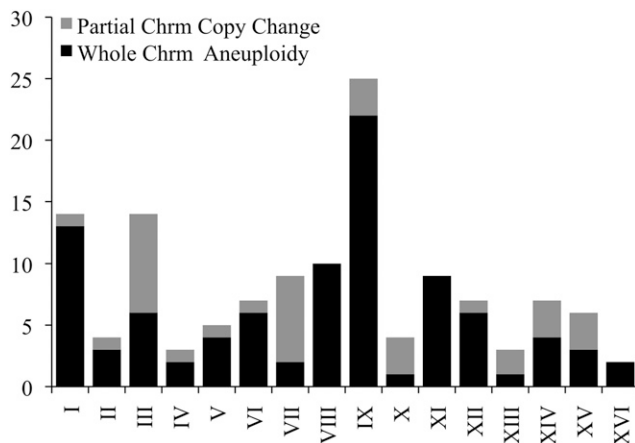


Figure 6 Histogram of the total number of complete and partial aneuploid events observed within all 144 sequenced strains, by chromosome.

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 2011). We used SIFT to look for damaging SNP mutations that are likely to affect the function of these genes (*Materials and 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

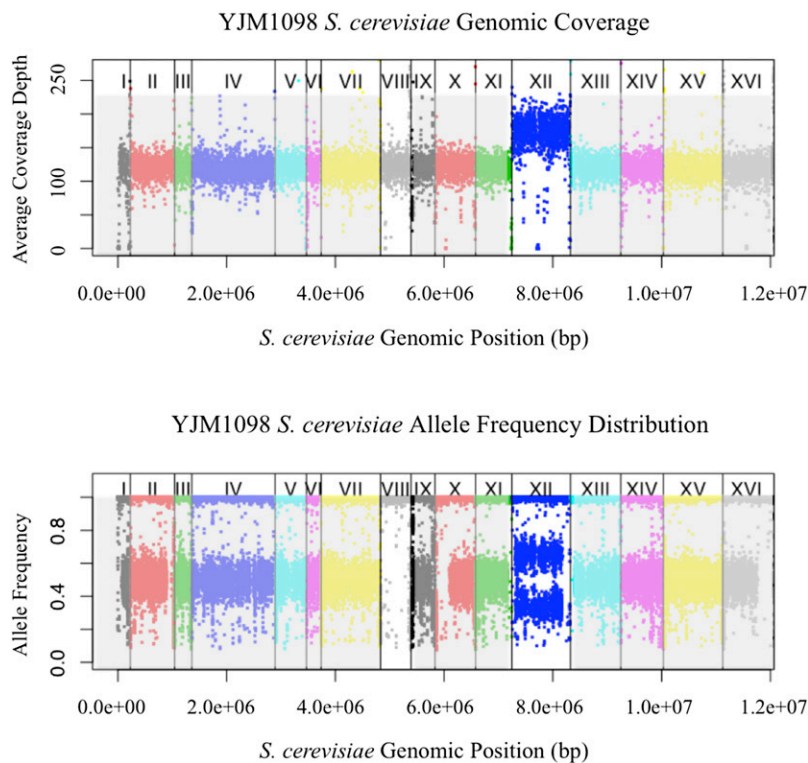


Figure 7 Coverage and allele frequency plots for YJM1098. Chromosomes are lined in increasing order (x-axis), with total coverage in 1 kb windows (top panel) or raw SNP allele frequencies (bottom panel) plotted (y-axis). YJM1098 showed a copy number change in chromosome XII with corresponding shift in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

strains' average genomic coverage was <0.05 . Ninety-five 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. One hundred and nine genes were significant for duplication (Table S5). Rare duplications (63 genes found in 1–2 strains each) 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). One hundred and nineteen 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) were enriched for membrane components (Table 1). Overall, 90/215 genes were members in two 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.

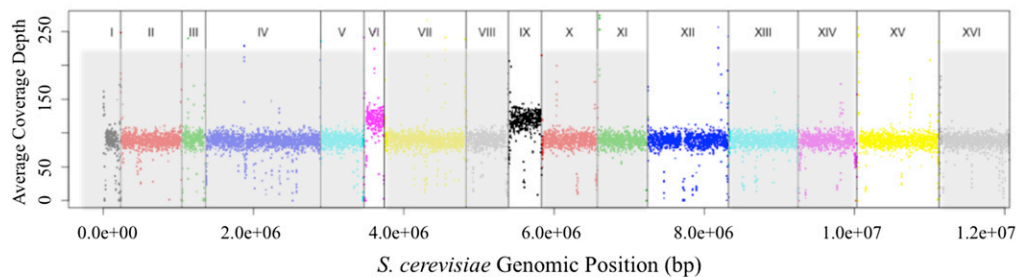
DISCUSSION

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 2009b; Sicard and Legras 2011; Wang *et al.* 2012; Liti 2015; Peter and Schacherer 2015). 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 (Gayevskiy 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 nonclinical strains.

This task is hampered by the fact that a strain's 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 toward 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 nonclinical 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 *et al.* 2011), and many remain genetically most similar to commercial strains (likely their source populations) rather than other clinical strains (de Llanos *et al.* 2006a,b; McKenzie *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

YJM466 *S. cerevisiae* Genomic Coverage



YJM466 *S. cerevisiae* Allele Frequency

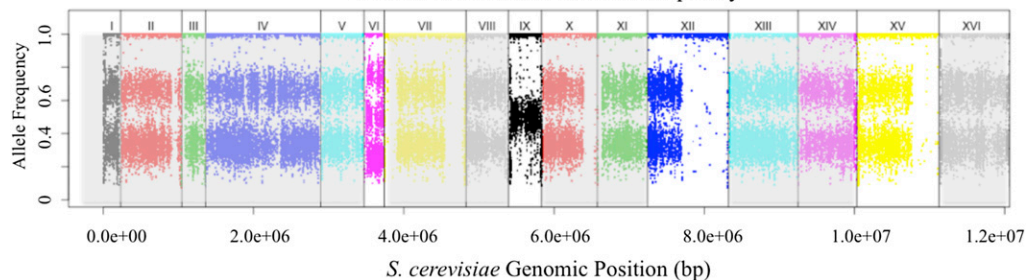


Figure 8 Coverage and allele frequency plots for YJM466. Chromosomes are lined in increasing order (x-axis), with total coverage in 1k b windows (top panel) or raw SNP allele frequencies (bottom panel) plotted (y-axis). YJM466 is a triploid strain that showed a copy number increase at chromosomes VI 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 $2n$.

lifestyle at all. To fully study the transition of nonpathogenic yeast to a pathogenic lifestyle, either models are needed that can select for clinical behavior in nonclinical strains, to determine precise changes that might afford this change in lifestyle, or thousands of clinical and nonclinical yeast strains must be sequenced for a comprehensive analysis. Here, we start the process by resequencing 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. NJ 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 toward 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 (see Dequin and Casaregola 2011 for review). 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 2009b; 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 noncommercial sources, or are mainly able to do so as a side effect of domestication prolonging exposure and opportunity.

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

■ **Table 1 GO enrichment terms for genes that showed complete deletion, copy number gain, and copy number loss**

	#	Expected	Fold Enrichment	+/-	P Value
Deleted genes (1–7 strains)					
Biological processes					
GO biological process complete					
Vitamin B6 biosynthetic process (GO:0042819)	4	0.06	>5	+	1.00E-03
Pyridoxine metabolic process (GO:0008614)	4	0.06	>5	+	1.00E-03
Vitamin B6 metabolic process (GO:0042816)	4	0.07	>5	+	1.60E-03
Cellular macromolecule metabolic process (GO:0044260)	4	19.42	0.21	-	1.95E-03
Pyridoxal phosphate biosynthetic process (GO:0042823)	4	0.09	>5	+	3.53E-03
Pyridoxal phosphate metabolic process (GO:0042822)	4	0.09	>5	+	3.53E-03
Aldehyde biosynthetic process (GO:0046184)	4	0.09	>5	+	3.53E-03
Cellular component organization or biogenesis (GO:0071840)	3	17.35	<0.2	-	3.83E-03
Macromolecule metabolic process (GO:0043170)	5	20.23	0.25	-	4.30E-03
Molecular function					
GO molecular function complete					
Amine-lyase activity (GO:0016843)	4	0.05	>5	+	1.71E-04
Pyridoxal 5'-phosphate synthase (GO:0036381)	4	0.05	>5	+	1.71E-04
Lyase activity (GO:0016829)	8	0.88	>5	+	2.55E-03
Carbon-nitrogen lyase activity (GO:0016840)	4	0.11	>5	+	4.84E-03
Cellular component					
GO cellular component complete					
Intracellular (GO:0005622)	24	38.94	0.62	-	5.74E-03
Intracellular part (GO:0044424)	24	38.81	0.62	-	6.98E-03
Macromolecular complex (GO:0032991)	4	16.95	0.24	-	1.40E-02
Intrinsic component of membrane (GO:0031224)	24	11.14	2.15	+	3.96E-02
Deleted genes (>8 strains)					
Biological processes					
GO biological process complete					
Maltose metabolic process (GO:0000023)	3	0.06	>5	+	4.34E-02
Molecular function					
GO molecular function complete					
Oligosaccharide transmembrane transporter activity (GO:0015157)	3	0.01	>5	+	2.92E-04
Disaccharide transmembrane transporter activity (GO:0015154)	3	0.01	>5	+	2.92E-04
Maltose transmembrane transporter activity (GO:0005363)	3	0.01	>5	+	2.92E-04
α -Glucoside transmembrane transporter activity (GO:0015151)	3	0.02	>5	+	6.91E-04
Glucoside transmembrane transporter activity (GO:0042947)	3	0.02	>5	+	6.91E-04
Sugar transmembrane transporter activity (GO:0051119)	4	0.08	>5	+	1.56E-03
Carbohydrate transmembrane transporter activity (GO:0015144)	4	0.1	>5	+	2.71E-03
Carbohydrate transporter activity (GO:1901476)	4	0.1	>5	+	2.71E-03
Cadmium ion binding (GO:0046870)	2	0.01	>5	+	3.31E-02
Cellular component					
GO cellular component complete					
Plasma membrane (GO:0005886)	11	2	>5	+	1.08E-03
Ccell periphery (GO:0071944)	11	2.86	3.85	+	3.32E-02
Duplicated genes (1–2 strains)					
Molecular function					
GO molecular function complete					
Thiosulfate sulfurtransferase activity (GO:0004792)	3	0.05	>5	+	1.45E-02
Duplicated genes (>2 strains)					
Biological processes					
GO biological process complete					
Transmembrane transport (GO:0055085)	13	2.8	4.65	+	4.08E-03
Cellular component organization or biogenesis (GO:0071840)	3	15.01	<0.2	-	4.30E-02
Molecular function					
GO molecular function complete					
Transporter activity (GO:0005215)	14	3.04	4.61	+	8.68E-04
Sucrose α -glucosidase activity (GO:0004575)	3	0.04	>5	+	9.93E-03
β -Fructofuranosidase activity (GO:0004564)	3	0.04	>5	+	9.93E-03
Transmembrane transporter activity (GO:0022857)	11	2.39	4.61	+	1.69E-02
Glucosidase activity (GO:0015926)	4	0.19	>5	+	4.03E-02
Cellular component					
GO cellular component complete					
Macromolecular complex (GO:0032991)	3	14.67	0.2	-	2.34E-02

(continued)

■ Table 1, continued

	#	Expected	Fold Enrichment	+/-	P Value
Organelle (GO:0043226)	16	29.44	0.54	-	2.71E-02
Intracellular organelle (GO:0043229)	16	29.37	0.54	-	2.97E-02
Organelle part (GO:0044422)	7	19.76	0.35	-	3.36E-02
Plasma membrane (GO:0005886)	12	3.21	3.74	+	3.47E-02
Intracellular organelle part (GO:0044446)	7	19.68	0.36	-	3.70E-02
Loss of copy genes (1-2 strains)					
Biological processes					
GO biological process complete					
Cellular component organization or biogenesis (GO:0071840)	1	11.68	<0.2	-	2.25E-02
Cellular macromolecule metabolic process (GO:0044260)	2	13.07	<0.2	-	3.27E-02
Molecular function					
GO molecular function complete					
Oligo-1,6-glucosidase activity (GO:0004574)	3	0.03	>5	+	2.68E-03
Acid phosphatase activity (GO:0003993)	3	0.05	>5	+	1.54E-02
α-Glucosidase activity (GO:0090599)	3	0.06	>5	+	2.79E-02
Cellular component					
GO cellular component complete					
Organelle (GO:0043226)	6	22.9	0.26	-	3.84E-06
Intracellular organelle (GO:0043229)	6	22.85	0.26	-	4.30E-06
Membrane-bounded organelle (GO:0043227)	5	21.18	0.24	-	1.46E-05
Intracellular membrane-bounded organelle (GO:0043231)	5	21.13	0.24	-	1.62E-05
Intracellular (GO:0005622)	11	26.21	0.42	-	5.17E-05
Intracellular part (GO:0044424)	11	26.12	0.42	-	6.29E-05
Organelle part (GO:0044422)	4	15.37	0.26	-	2.42E-02
Intracellular organelle part (GO:0044446)	4	15.31	0.26	-	2.63E-02
Loss of copy genes (>2 strains)					
Biological processes					
GO biological process complete					
Transmembrane transport (GO:0055085)	20	5.1	3.92	+	1.91E-04
Cellular component organization or biogenesis (GO:0071840)	8	27.36	0.29	-	1.11E-03
Maltose metabolic process (GO:0000023)	5	0.16	>5	+	1.28E-03
Macromolecule metabolic process (GO:0043170)	12	31.91	0.38	-	2.51E-03
Regulation of biological process (GO:0050789)	4	19.5	0.21	-	7.50E-03
Cellular macromolecule metabolic process (GO:0044260)	12	30.63	0.39	-	9.19E-03
Cellular component organization (GO:0016043)	7	23.83	0.29	-	1.00E-02
Ion transport (GO:0006811)	15	3.92	3.83	+	1.30E-02
Regulation of cellular process (GO:0050794)	4	18.43	0.22	-	2.28E-02
Thiamine-containing compound biosynthetic process (GO:0042724)	4	0.15	>5	+	3.07E-02
Thiamine biosynthetic process (GO:0009228)	4	0.15	>5	+	3.07E-02
Thiamine metabolic process (GO:0006772)	4	0.16	>5	+	4.19E-02
Molecular function					
GO molecular function complete					
Transmembrane transporter activity (GO:0022857)	20	4.35	4.6	+	7.55E-06
Transporter activity (GO:0005215)	21	5.53	3.8	+	7.92E-05
Substrate-specific transmembrane transporter activity (GO:0022891)	17	3.92	4.34	+	2.74E-04
Substrate-specific transporter activity (GO:0022892)	17	4.75	3.58	+	3.80E-03
Ion transmembrane transporter activity (GO:0015075)	13	2.86	4.55	+	5.03E-03
Sugar transmembrane transporter activity (GO:0051119)	5	0.25	>5	+	5.51E-03
Oligosaccharide transmembrane transporter activity (GO:0015157)	3	0.04	>5	+	7.76E-03
Disaccharide transmembrane transporter activity (GO:0015154)	3	0.04	>5	+	7.76E-03
Maltose transmembrane transporter activity (GO:0005363)	3	0.04	>5	+	7.76E-03
Sodium-exporting ATPase activity, phosphorylative mechanism (GO:0008554)	3	0.04	>5	+	7.76E-03
Cellular component					
GO cellular component complete					
Plasma membrane (GO:0005886)	25	5.85	4.27	+	1.88E-07
Cell periphery (GO:0071944)	27	8.38	3.22	+	1.35E-05
Intrinsic component of membrane (GO:0031224)	40	17.56	2.28	+	2.51E-05
Protein complex (GO:0043234)	2	19.2	<0.2	-	6.80E-05
Integral component of membrane (GO:0016021)	38	16.82	2.26	+	9.03E-05
Membrane part (GO:0044425)	40	20.68	1.93	+	2.37E-03

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 2015), 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 Strobe *et al.* (2015), it provides information on a different scale and is complementary 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.

ACKNOWLEDGMENTS

We thank Greg Lang, Patrick Gibney, and the WashU Genome Sequencing Center for generation of the DNA sequencing data. We thank John McCusker for sharing background knowledge on the clinical strains sequenced in this study. Y.O.Z. was supported by the A*STAR National Science Scholarship PhD. G.S. was supported by R01 HG003328. D.A.P. was supported by the National Institutes of Health grants RO1GM100366 and RO1GM097415.

LITERATURE CITED

Albertin, W., P. Marullo, M. Aigle, A. Bourgeois, M. Bely *et al.*, 2009 Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. *J. Evol. Biol.* 22: 2157–2170.

Aucott, J. N., J. Fayen, H. Grossnicklas, A. Morrissey, M. M. Lederman *et al.*, 1990 Invasive infection with *Saccharomyces cerevisiae*: report of three cases and review. *Rev. Infect. Dis.* 12: 406–411.

Barrio, E., S. González, A. Arias, C. Belloch, and A. Querol, 2006 Molecular mechanisms involved in the adaptive evolution of industrial yeasts, pp. 153–174 in *The Yeast Handbook*, edited by Querol, A., and G. H. Fleet. Springer Verlag, Berlin, Germany.

Berman, J., 2016 Ploidy plasticity: a rapid and reversible strategy for adaptation to stress. *FEMS Yeast Res.* 16(3): abstract Available at <http://femsyr.oxfordjournals.org/content/16/3/fov020>.

Bidenne, C., B. Blondin, A. Querol, and F. Vezinhet, 1992 Analysis of the chromosomal DNA polymorphism of wine strains of *Saccharomyces cerevisiae*. *Curr. Genet.* 22: 1–7.

Biggins, S., N. Bhalla, A. Chang, D. L. Smith, and A. W. Murray, 2001 Chromatid separation and segregation in the budding yeast *Saccharomyces cerevisiae*. *Genetics* 159(2): 453–470.

Blondin, B., S. Dequin, A. Querol, and J. L. Legras, 2009 Genome of *Saccharomyces cerevisiae* and related yeasts, pp. 361–378 in *Biology of Microorganisms on Grapes, in Must and in Wine*. 19

Carbon, S., A. Ireland, C. J. Mungall, S. Q. Shu, B. Marshall *et al.*, 2009 AmiGO: online access to ontology and annotation data. *Bioinformatics* 25(2): 288–289.

Carro, D., E. Bartra, and B. Pina, 2003 Karyotype rearrangements in a wine yeast strain by rad52-dependent and rad52-independent mechanisms. *Appl. Environ. Microbiol.* 69: 2161–2165.

Cassone, M., P. Serra, F. Mondello, A. Girolamo, S. Scafetti *et al.*, 2003 Outbreak of *Saccharomyces cerevisiae* subtype *boulardii* fungemia in patients neighboring those treated with a probiotic preparation of the organism. *J. Clin. Microbiol.* 41(11): 5340–5343.

Cavaliieri, D., P. E. McGovern, D. L. Hartl, R. Mortimer, and M. Polsinelli, 2003 Evidence for *S. cerevisiae* fermentation in ancient wine. *J. Mol. Evol.* 57: S226–S232.

Cheng, E., J. A. Vaisica, J. Ou, A. Baryshnikova, Y. Lu *et al.*, 2012 Genome rearrangements caused by depletion of essential DNA replication proteins in *Saccharomyces cerevisiae*. *Genetics* 192(1): 147–160.

Cho, U. S., and S. C. Harrison, 2011 Ndc10 is a platform for inner kinetochore assembly in budding yeast. *Nat. Struct. Mol. Biol.* 19: 48–55.

Clemons, K. V., J. H. McCusker, R. W. Davis, and D. A. Stevens, 1994 Comparative pathogenesis of clinical and nonclinical isolates of *Saccharomyces cerevisiae*. *J. Infect. Dis.* 169: 859–867.

Clemons, K. V., L. C. Hanson, and D. A. Stevens, 1996 Colony phenotype switching in clinical and non-clinical isolates of *Saccharomyces cerevisiae*. *J. Med. Vet. Mycol.* 34: 259–264.

Coyle, S., and E. Kroll, 2008 Starvation induces genomic rearrangements and starvation-resilient phenotypes in yeast. *Mol. Biol. Evol.* 25: 310–318.

Cromie, G. A., and A. M. Dudley, 2015 Aneuploidy: tolerating tolerance. *Curr. Biol.* 25(17): R771–R773.

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. *Syst. Appl. Microbiol.* 27: 427–435.

de Llanos, R., M. T. Fernandez-Espinar, and A. Querol, 2006a 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., A. Querol, J. Peñan, M. Gobernado, and M. T. Fernández-Espinar, 2006b Food and probiotic strains from the *Saccharomyces cerevisiae* species as a possible origin of human systemic infections. *Int. J. Food Microbiol.* 110: 286–290.

de Llanos, R., S. Llopis, G. Molero, A. Querol, C. Gil *et al.*, 2011 *In vivo* virulence of commercial *Saccharomyces cerevisiae* strains with pathogenicity-associated phenotypic traits. *Int. J. Food Microbiol.* 144(3): 393–399.

Dequin, S., and S. Casaregola, 2011 The genomes of fermentative *Saccharomyces*. *C. R. Biol.* 334: 687–693.

Dougherty, S. H., and R. L. Simmons, 1982 Postoperative peritonitis caused by *Saccharomyces cerevisiae*. *Arch. Surg.* 117: 248.

Doyle, M. G., L. K. Pickering, N. O'Brien, K. Hoots, and J. E. Benson, 1990 *Saccharomyces cerevisiae* infection in a patient with acquired immunodeficiency syndrome. *Pediatr. Infect. Dis. J.* 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 Res.* 22(5): 908–924.

Enache-Angoulvant, A., and C. Hennequin, 2005 Invasive *Saccharomyces* infection: a comprehensive review. *Nephrol. Dial. Transplant.* 41(11): 1559–1568.

Erdem, H., M. Cetin, T. Timuroglu, A. Cetin, O. Yanar *et al.*, 2003 Identification of yeasts in public hospital primary care patients with or without clinical vaginitis. *Aust. N. Z. J. Obstet. Gynaecol.* 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(4): 1567–1587.

Falush, D., M. Stephens, and J. K. Pritchard, 2007 Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7(4): 574–578.

Farrell, A. E., R. J. Plevin, B. T. Turner, A. D. Jones, M. O'Hare *et al.*, 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 *et al.*, 2016 Further support for aneuploid 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. *ISME J.* 6(7): 1281–1290. [10.1038/ismej.2011.195](https://doi.org/10.1038/ismej.2011.195)

Goddard, M. R., and D. Greig, 2015 *Saccharomyces cerevisiae*: a nomadic yeast with no niche? *FEMS Yeast Res.* 15(3): abstract Available at <http://femsyr.oxfordjournals.org/content/15/3/fov009>.

Goddard, M. R., N. Anfang, R. Y. 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. *Environ. Microbiol.* 12(1): 63–73.

Goh, P. Y., and J. V. Kilmartin, 1993 NDC10: A gene involved in chromosome segregation in *Saccharomyces cerevisiae*. *J. Cell Biol.* 121(3): 503–512.

- Hennequin, C., C. Kauffmann-Lacroix, A. Jobert, J. P. Viard, C. Ricour *et al.*, 2000 Possible role of catheters in *Saccharomyces boulardii* fungemia. *Eur. J. Clin. Microbiol. Infect. Dis.* 19: 16–20.
- Hickman, M. A., C. Paulson, A. Dudley, and J. Berman, 2015 Parasexual ploidy reduction drives population heterogeneity through random and transient aneuploidy in *Candida albicans*. *Genetics* 200(3): 781–794.
- Hirakawa, M. P., D. A. Martinez, S. Sakthikumar, M. Z. Anderson, A. Berlin *et al.*, 2015 Genetic and phenotypic intra-species variation in *Candida albicans*. *Genome Res.* 25(3): 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, A. Wang, M. A. Newton *et al.*, 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. *Mol. Ecol. Resour.* 9(5): 1322–1332.
- James, T. C., J. Usher, S. Campbell, and U. Bond, 2008 Lager yeast possess dynamic genomes that undergo rearrangements and amplification in response to stress. *Curr. Genet.* 3: 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 Res.* 8(4): 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* 10.1371/journal.pone.0068094.
- Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock *et al.*, 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500: 571–574.
- Legras, J.-L., D. Merdinoglu, J.-M. Cornuet, and F. Karst, 2007 Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Mol. Ecol.* 16(10): 2091–2102.
- Letunic, I. and P. Bork 2006 Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* 23(1): 127–128.20
- Letunic, I. and P. Bork 2011 Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39(suppl 2): W475–W478.21
- Lherm, T., C. Monet, B. Nougere, M. Soulier, D. Larbi *et al.*, 2002 Seven cases of fungemia with *Saccharomyces boulardii* in critically ill patients. *Intensive Care Med.* 28: 797–801.
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14): 1754–1760.
- Li, R., Y. Li, X. Fang, H. Yang, J. Wang *et al.*, 2009 SNP detection for massively parallel whole genome resequencing. *Genome Res.* 19(6): 1124–1132.
- Liti, G., 2015 The natural history of model organisms: 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 *et al.*, 2009 Population genomics of domestic and wild yeasts. *Nature* 458(7236): 337–341.
- Malgoire, J. Y., S. Bertout, F. Renaud, J. M. Bastide, and M. Mallié, 2005 Typing of *Saccharomyces cerevisiae* clinical strains by using microsatellite sequence polymorphism. *J. Clin. Microbiol.* 43(3): 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. *J. Clin. Microbiol.* 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. *Infect. Immun.* 62(12): 5447–5455.
- McGovern, P. E., U. Hartung, V. R. Badler, D. L. Glusker, and L. J. Exner, 1997 The beginnings of winemaking and viticulture in the ancient Near East and Egypt. *Expedition* 39: 3–21.
- McGovern, P. E., J. Zhang, J. Tang, Z. Zhang, G. R. Hall *et al.*, 2004 Fermented beverages of pre- and proto-historic China. *Proc. Natl. Acad. Sci. USA* 101: 17593–17598.
- McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20(1): 1297–1303.
- McKenzie, D. A., M. Defernez, W. B. Dunn, M. Brown, L. J. Fuller *et al.*, 2008 Relatedness of medically important strains of *Saccharomyces cerevisiae* as revealed by phylogenetics and metabolomics. *Yeast* 25(7): 501–512.
- Muller, L. A., and J. H. McCusker, 2009a A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 9(1): 143–152.
- Muller, L. A., and J. H. McCusker, 2009b Microsatellite analysis of genetic diversity among clinical and nonclinical *Saccharomyces cerevisiae* isolates suggests heterozygote advantage in clinical environments. *Mol. Ecol.* 18(13): 2779–2786.
- Muller, L. A., and J. H. McCusker, 2011 Nature and distribution of large sequence polymorphisms in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 11(7): 587–594.22
- Muller, L. A., J. E. Lucas, D. R. Geogianna, and J. H. McCusker, 2011 Genome-wide association analysis of clinical vs. nonclinical origin provides insights into *Saccharomyces cerevisiae* pathogenesis. *Mol. Ecol.* 20(19): 4085–4097.
- Munoz, P., E. Bouza, M. Cuenca-Estrella, J. M. Eiros, M. J. Perez *et al.*, 2005 *Saccharomyces cerevisiae* fungemia: an emerging infectious disease. *Clin. Infect. Dis.* 40: 1625–1634.
- Mydlik, M., E. Tkacova, K. Szovenyiova, P. Mizla, and K. Derzsiova, 1996 *Saccharomyces cerevisiae* peritonitis complicating CAPD. *Perit. Dial. Int.* 16: 188.
- Ouspenski, I. I., S. J. Elledge, and B. R. Brinkley, 1999 New yeast genes important for chromosome integrity and segregation identified by dosage effects on genome stability. *Nucleic Acids Res.* 27(15): 3001–3008.
- Pavelka, N., G. Rancati, J. Zhu, W. D. Bradford, A. Saraf *et al.*, 2010 Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468(7321): 321–325.
- Perapoch, J., A. M. Planes, A. Querol, V. López, I. Martínez-Bendayán *et al.*, 2000 Fungemia with *Saccharomyces cerevisiae* in two newborns, only one of whom had been treated with Ultra-Levura. *Eur. J. Clin. Microbiol. Infect. Dis.* 19: 468–470.
- Perez-Torrado, R., S. Llopis, L. Jespersen, T. Fernández-Espinar, A. Querol *et al.*, 2012 Clinical *Saccharomyces cerevisiae* isolates cannot cross the epithelial barrier *in vitro*. *Int. J. Food Microbiol.* 157(1): 59–64.
- Peter, J., and J. Schacherer, 2015 Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* 10.1002/yea.3142.
- Piarroux, R., L. Millon, K. Bardouet, O. Vagner, and H. Koenig, 1999 Are live saccharomyces yeasts harmful to patients? *Lancet* 353: 1851–1852.
- Posteraro, B., M. Sanguinetti, G. D'Amore, L. Masucci, G. Morace *et al.*, 1999 Molecular and epidemiological characterization of vaginal *Saccharomyces cerevisiae* isolates. *J. Clin. Microbiol.* 37: 2230–2235.
- 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 Microbiol. Lett.* 293: 1–10.
- Rancati, G., and N. Pavelka, 2013 Karyotypic changes as drivers and catalyzers of cellular evolvability: a perspective from non-pathogenic yeasts. *Semin. Cell Dev. Biol.* 24(4): 332–338.
- Rancati, G., N. Pavelka, B. Fleharty, A. Noll, R. Trimble *et al.*, 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 Res.* 10.1093/nar/gkr378.
- Salonen, J. H., M. D. Richardson, K. Gallacher, J. Issakainen, H. Helenius *et al.*, 2000 Fungal colonization of haematological patients receiving cytotoxic chemotherapy: emergence of azole-resistant *Saccharomyces cerevisiae*. *J. Hosp. Infect.* 45: 293–301.
- Samuel, D., 1996 Investigation of ancient Egyptian baking and brewing methods by correlative microscopy. *Science* 273: 488–490.

- Schacherer, J., J. A. Shapiro, D. M. Ruderfer, and L. Kruglyak, 2009 Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458(7236): 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(10): e1000705.
- Selmecki, A. M., Y. E. Maruvka, P. A. Richmond, M. Guillet, N. Shores *et al.*, 2015 Polyploidy can drive rapid adaptation in yeast. *Nature* 519(7543): 349–352.
- Sheppard, T. K., B. C. Hitz, S. R. Engel, G. Song, R. Balakrishnan *et al.*, 2016 The *Saccharomyces* Genome Database Variant Viewer. *Nucleic Acids Res.* 44(D1): D698–D702.
- Sicard, D., and J. L. Legras, 2011 Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C. R. Biol.* 334(3): 229–236.
- Smith, D., D. Metzgar, C. Wills, and J. Fierer, 2002 Fatal *Saccharomyces cerevisiae* aortic graft infection. *J. Clin. Microbiol.* 40(7): 2691–2692.
- Stambuk, B. U., B. Dunn, S. L. Alves, Jr, 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 Res.* 19(12): 2271–2278.
- Stirling, P. C., M. S. Bloom, T. Solanki-Patil, S. Smith, P. Sopahimalani *et al.*, 2011 The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. *PLoS Genet.* 10.1371/journal.pgen.1002057.
- Strope, P. K., D. A. Skelly, S. G. Kozmin, G. Mahadevan, E. A. Stone *et al.*, 2015 The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res.* 25(5): 762–774.
- Sunshine, A. B., C. Payen, G. T. Ong, I. Liachko, K. M. Tan *et al.*, 2015 The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biol.* 13(5): e1002155.
- Thorburn, R. R., C. Gonzalez, G. A. Brar, S. Christen, T. M. Carlile *et al.*, 2013 Aneuploid yeast strains exhibit defects in cell growth and passage through START. *Mol. Biol. Cell* 24(9): 1274–1289.
- Waghmare, S. K., and C. V. Bruschi, 2005 Differential chromosome control of ploidy in the yeast *Saccharomyces cerevisiae*. *Yeast* 22(8): 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. *Mol. Ecol.* 21(22): 5404–5417.
- Wei, W., J. H. McCusker, R. W. Hyman, T. Jones, Y. Ning *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc. Natl. Acad. Sci. USA* 104: 12825–12830.
- Wheeler, R. T., M. Kupiec, P. Magnelli, C. Abeijon, and G. R. Fink, 2003 A *Saccharomyces cerevisiae* mutant with increased virulence. *Proc. Natl. Acad. Sci. USA* 100: 2766–2770.
- Whittaker, S. G., B. M. Rockmill, E. Blechl, D. H. Maloney, M. Resnick *et al.*, 1988 The detection of mitotic and meiotic aneuploidy in yeast using a gene dosage selection system. *Molecular & General Genetics: MGG* 215(1): 10–18.
- Zerva, L., R. J. Hollis, and M. A. Pfaller, 1996 In vitro susceptibility testing and DNA typing of *Saccharomyces cerevisiae* clinical isolates. *J. Clin. Microbiol.* 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.*
- Zhu, Y. O., M. L. Siegal, D. W. Hall, and D. A. Petrov, 2014 Precise estimates of mutation rate and spectrum in yeast. *Proc. Natl. Acad. Sci. USA* 111(22): E2310–E2318.
- Zörgö, E., K. Chwialkowska, A. B. Gjuvsland, E. Garré, P. Sunnerhagen *et al.*, 2013 Evolutionary trade-offs between yeast ploidy states. *PLoS Genet.* 9(3): e1003388.

Communicating editor: B. J. Andrews