



# Copy Number Variation in Fungi and Its Implications for Wine Yeast Genetic Diversity and Adaptation

Jacob L. Steenwyk and Antonis Rokas\*

Department of Biological Sciences, Vanderbilt University, Nashville, TN, United States

## OPEN ACCESS

### Edited by:

Aline Lonvaud,  
Université de Bordeaux, France

### Reviewed by:

Jan Steensels,  
Flanders Institute for Biotechnology,  
Belgium

Estefani Garcia Rios,  
Consejo Superior de Investigaciones  
Científicas (CSIC), Spain

### \*Correspondence:

Antonis Rokas  
antonis.rokas@vanderbilt.edu

### Specialty section:

This article was submitted to  
Food Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 12 December 2017

**Accepted:** 07 February 2018

**Published:** 22 February 2018

### Citation:

Steenwyk JL and Rokas A (2018)  
Copy Number Variation in Fungi  
and Its Implications for Wine Yeast  
Genetic Diversity and Adaptation.  
*Front. Microbiol.* 9:288.  
doi: 10.3389/fmicb.2018.00288

In recent years, copy number (CN) variation has emerged as a new and significant source of genetic polymorphisms contributing to the phenotypic diversity of populations. CN variants are defined as genetic loci that, due to duplication and deletion, vary in their number of copies across individuals in a population. CN variants range in size from 50 base pairs to whole chromosomes, can influence gene activity, and are associated with a wide range of phenotypes in diverse organisms, including the budding yeast *Saccharomyces cerevisiae*. In this review, we introduce CN variation, discuss the genetic and molecular mechanisms implicated in its generation, how they can contribute to genetic and phenotypic diversity in fungal populations, and consider how CN variants may influence wine yeast adaptation in fermentation-related processes. In particular, we focus on reviewing recent work investigating the contribution of changes in CN of fermentation-related genes in yeast wine strains and offer notable illustrations of such changes, including the high levels of CN variation among the *CUP* genes, which confer resistance to copper, a metal with fungicidal properties, and the preferential deletion and duplication of the *MAL1* and *MAL3* loci, respectively, which are responsible for metabolizing maltose and sucrose. Based on the available data, we propose that CN variation is a substantial dimension of yeast genetic diversity that occurs largely independent of single nucleotide polymorphisms. As such, CN variation harbors considerable potential for understanding and manipulating yeast strains in the wine fermentation environment and beyond.

**Keywords:** structural variation, alcohol fermentation, sugar metabolism, gene duplication, gene loss, population genomics

## INTRODUCTION

Genetic variation in natural populations is shaped by diverse biological processes, such as genetic drift and natural selection (Chakravarti, 1999), and is, in part, responsible for phenotypic variation. For example, arginine auxotrophy in the baker's yeast *Saccharomyces cerevisiae* is a Mendelian inherited trait due to polymorphisms in the *ARG4* locus (Brauer et al., 2006), whereas variation in *S. cerevisiae* colony morphology is a complex trait driven by variants in several different genes (Taylor et al., 2016). The aforementioned yeast phenotypes are all caused by SNPs or small insertions and deletions, which are by far the most well characterized types of genetic variation

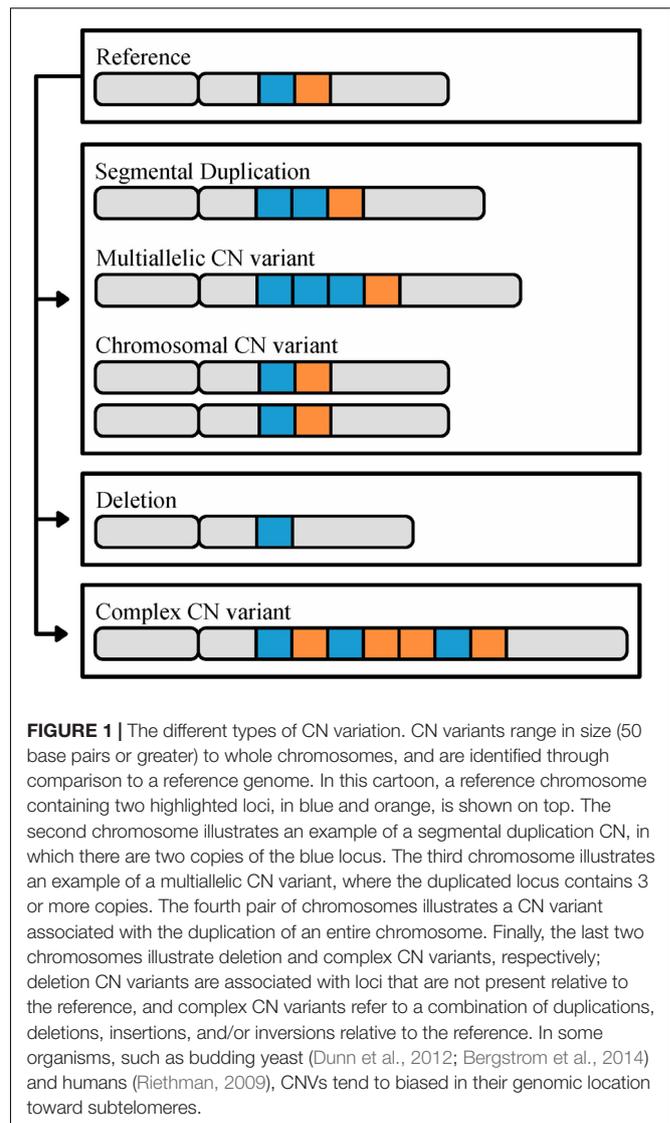
**Abbreviations:** BIR, break-induced recombination; CN, copy number; HR, homologous recombination; NHR, non-homologous repair; SNPs, single nucleotide polymorphisms.

not only in yeast, but in any kind of organism (Sachidanandam et al., 2001; McNally et al., 2009; Schacherer et al., 2009). In recent years, however, several studies in diverse organisms have revealed that genomes also harbor an abundance of structural variation, which too contributes to populations' genetic and phenotypic diversity (Stranger et al., 2007; Zhang et al., 2009).

Variation in the structure of chromosomes, or structural variation, encompasses a wide array of mutations including insertions, inversions, translocations, and CN variants (i.e., duplications and deletions) (Feuk et al., 2006) and, in humans, accounts for an estimated average of 74% of the nucleotide differences between two genomes (Rahim et al., 2008). The major influence of several types of structural variation, such as large-scale inversions, translocations, and insertions, on phenotype is better understood because many such variants can be microscopically examined and lead to classic human genetic disorders, such as Down's syndrome (Youings et al., 2004; Rausch et al., 2012; Gu et al., 2016). In contrast, many CN variants are submicroscopic and eschewed attention until the advent of whole genome sequencing technologies (Feuk et al., 2006).

Copy number variants are defined as duplications or deletions that range from 50 base pairs to whole chromosomes (Figure 1) and can significantly influence phenotypic diversity (Lieber, 2008; Riethman, 2009; Zhang et al., 2009; Arlt et al., 2014). For example, in humans, the CN of the salivary amylase gene, *AMY1*, is higher in populations with high-starch diets and correlated with salivary protein abundance thereby improving digestion of starchy foods (Perry et al., 2007). Levels of CN variation have been examined in diverse organisms across the tree of life, including animals (e.g., Humans; *Homo sapiens*: Sudmant et al., 2015, House mouse; *Mus musculus*: Pezer et al., 2015), plants (e.g., soybean; *Glycine max*: Cook et al., 2012, maize; *Zea mays*: Swanson-Wagner et al., 2010) and fungi (e.g., *Cryptococcus neoformans*: Hu et al., 2011, *Brettanomyces bruxellensis*: Curtin et al., 2012, *Batrachochytrium dendrobatidis*: Farrer et al., 2013, *Zyloseptoria tritici*: Hartmann and Croll, 2017). Additionally, CN variants spanning genes can be a major platform for functional divergence of gene duplicates (e.g., through subfunctionalization or the partitioning of a set of ancestral functions across duplicates), including the evolution of new functions (neofunctionalization) (Lynch and Conery, 2000; Soria et al., 2014; Reams and Roth, 2015). For example, duplicated phospholipase genes that have undergone neofunctionalization are responsible for the evolution and diversification of snake venom and snake species (Lynch, 2007), whereas clusters of tandemly duplicated genes are associated with phenotypic diversity in many traits and organisms (Ortiz and Rokas, 2017).

*Saccharomyces cerevisiae* has been an important model for genetics, genomics, and evolution (Goffeau et al., 1996; Botstein et al., 1997; Winzeler et al., 1999). Much of what we know about the evolutionary history of *S. cerevisiae* stems from investigating genome-wide patterns of SNPs among globally distributed strains. Examination of genome-wide patterns of SNP variation has yielded valuable insights into yeast function in the wine fermentation environment. For example, 13 SNPs in *ABZ1*, a gene associated with nitrogen biosynthetic pathways, have been shown to modify the rate of fermentation and nitrogen utilization during fermentation (Ambroset et al., 2011).



Interrogations of genome-wide patterns of SNPs have also shown that industrial lineages – including those of beer, bread, cacao, sake, and wine – often mirror human history (Schacherer et al., 2009; Sicard and Legras, 2011; Cromie et al., 2013; Gallone et al., 2016; Gonçalves et al., 2016), suggesting that human activity has greatly influenced *S. cerevisiae* genome evolution (Yue et al., 2017). Furthermore, SNP-based studies have repeatedly found that wine strains of *S. cerevisiae* exhibit low levels of genetic diversity (Liti et al., 2009; Schacherer et al., 2009; Sicard and Legras, 2011; Cromie et al., 2013; Borneman et al., 2016), consistent with a historical population bottleneck event that reduced wine yeast genetic variation. The low SNP diversity among wine yeast strains has led some to suggest that wine strain development may benefit from the introduction of genetic variation from yeasts outside the wine lineage (Borneman et al., 2016). However, recent studies examining CN variation among wine associated strains of *S. cerevisiae* have identified considerable genetic diversity (Gallone et al., 2016; Gonçalves

et al., 2016; Steenwyk and Rokas, 2017), suggesting that standing CN variation in wine strains may be industrially relevant.

In the present review, we begin by surveying the molecular mechanisms that lead to CN variant formation, we next discuss the contribution of CN variation to the genetic and phenotypic diversity in fungal populations, and close by examining the CN variation in wine yeasts and the likely phenotypic impact of CN variants in the wine fermentation environment.

## COPY NUMBER VARIATION AND THE MOLECULAR MECHANISMS THAT GENERATE IT

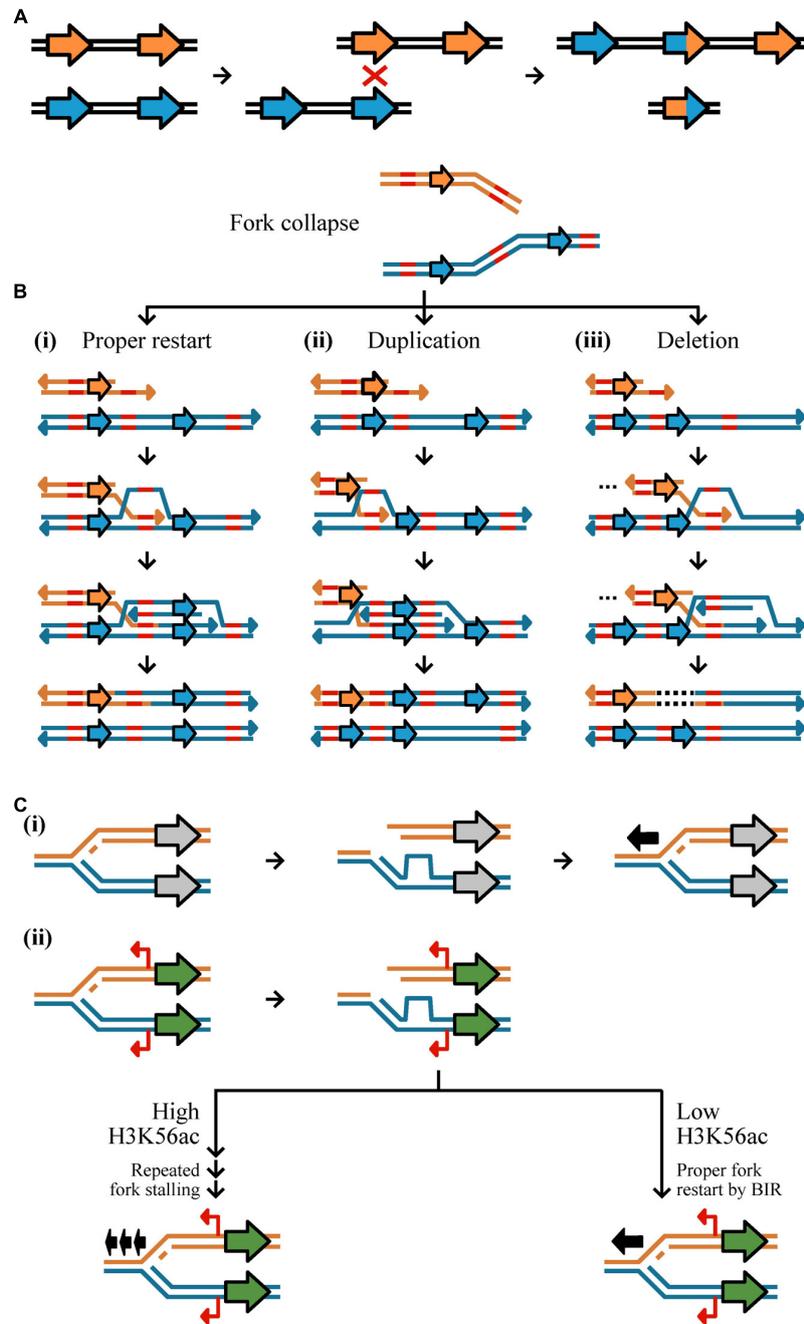
Copy number variants, a class of structural variants, are duplicated or deleted loci that range from 50 base pairs (bp) to whole chromosomes in length (**Figure 1**) and have a mutation rate 100–1,000 times greater than SNPs (Zhang et al., 2009; Arlt et al., 2014; Sener, 2014). CN variable loci can in turn be broken down into three subclasses (**Figure 1**) (Estivill and Armengol, 2007). The first subclass encompasses variants that originate via duplications; in the genome, these can appear as either identical or nearly identical copies, or multi-allelic CN variants (Bailey and Eichler, 2006; Usher and McCarroll, 2015). The extreme version of this subclass are chromosomal CN variants that correspond to duplications of entire chromosomes. The second subclass encompasses CN variants that originate via deletion leading to the loss of the sequence of a locus in the genome. The third subclass includes complex CN variants where a locus exhibits a combination of duplication, deletion, insertion, and inversion events (Usher and McCarroll, 2015).

Copy number variants are commonly generated from aberrant DNA repair via three mechanisms: HR, NHR, and environmental stimulation (**Figure 2**) (Hastings et al., 2009b; Hull et al., 2017). HR is a universal process associated with DNA repair and requires high sequence similarity across 60–300 bps (Hua et al., 1997; Petukhova et al., 1998). HR is initiated by double-strand breaks caused by ionizing radiation, reactive oxygen species, and mechanical stress on chromosomes such as those associated with collapsed or broken replication forks (Khanna and Jackson, 2001; Aylon and Kupiec, 2004; Hastings et al., 2009b). Improper repair by HR can result in duplication, deletion, or inversion of genetic material (Reams and Roth, 2015). Non-allelic HR (also known as ectopic recombination), defined as recombination between two different loci of the same or different chromosomes that share sequence similarity and are  $\geq 300$  base pairs in length, is among the most well-studied examples of improper repair (Kupiec and Petes, 1988; Prado et al., 2003). Most evidence of non-allelic HR resulting in CN variation is directly associated with low copy repeats or transposable elements (Xu and Boeke, 1987; Hurles, 2005). For example, a duplication and deletion may result during unequal crossing over of homologous sequences (**Figure 2A**) (Carvalho and Lupski, 2016). Improper HR may also occur at collapsed or broken replication forks by BIR (**Figure 2B**). BIR requires 3' strand invasion at the allelic site of stalled replication to properly restart DNA synthesis (**Figure 2B-i**) (Llorente et al., 2008), however, template switching, the non-allelic pairing of

homologous sequences, in the backward (**Figure 2B-ii**) or forward (**Figure 2B-iii**) direction can result in a duplication or deletion, respectively (Morrow et al., 1997; Smith et al., 2007). Although HR occurs with high fidelity, errors in the process, which are thought to increase in frequency during mitosis and meiosis, can generate CN variants (Hastings et al., 2009b).

In contrast to HR, NHR utilizes microhomologies (typically defined as  $\sim 65\%$  or more sequence similarity of short sequences up to ten bases long) or does not require homology altogether, and can too lead to CN variant formation (Daley et al., 2005; McVey and Lee, 2008). NHR can occur by two mechanisms: non-replicative and replicative (Hastings et al., 2009b). Non-replicative mechanisms include non-homologous end joining and microhomology-mediated end-joining (Lieber, 2008; McVey and Lee, 2008). Non-homologous end-joining refers to the direct ligation of sequences in a double-strand break (Daley et al., 2005). Prior to ligation, there may be a loss of genetic material or the addition of free DNA (e.g., from transposable elements or mitochondrial DNA) (Yu and Gabriel, 2003). Microhomology-mediated end joining is similar to non-homologous end-joining but occurs more frequently, requires different enzymes, and leverages homologies 1–10 base pairs in length to ensure more efficient annealing (Yu et al., 2004; Lieber, 2008). Non-homologous end-joining and microhomology-mediated non-homologous end-joining are primarily associated with small insertions and deletions and therefore are not likely to be a major driver of CN variation (Yu and Gabriel, 2003; Gu et al., 2008). Replicative mechanisms of CN variant formation include replication slippage, fork stalling, and microhomology BIR. Replication slippage occurs along repetitive stretches of DNA resulting in the duplication or deletion of sequence between repetitive regions (Hastings et al., 2009b). Fork stalling is thought to cause large CNVs of 20 kb average length through template switching between distal replication forks rather than within a replication fork (Slack et al., 2006). However, fork stalling without distal template switching can also be highly mutagenic and induce CN variants (Paul et al., 2013; Hull et al., 2017). Lastly, microhomology-mediated break-induced replication occurs when the 3' end of a collapsed fork anneals with any single-stranded template that it shares microhomology with to reinitiate DNA synthesis (**Figure 2B**) (Hastings et al., 2009b). Annealing can occur in the backward (**Figure 2B-ii**) or forward (**Figure 2B-iii**) direction of the allelic site causing a duplication or deletion, respectively, and is thought to be the primary cause of low copy repeats (Hastings et al., 2009a).

The third mechanism is associated with an epigenetic mark that can stimulate the formation of CN variants. Histone acetylation, specifically H3K56ac, is, in part, environmentally driven (Turner, 2009), associated with highly transcribed loci, and can promote CN variant formation through repeated fork stalling or template switching (**Figure 2C**) (Hull et al., 2017). For example, it has been shown that exposure to environmental copper stimulates the generation of CN variation in *CUP1*, a gene that is associated with copper resistance when duplicated (Fogel and Welch, 1982), thereby increasing the likelihood of favorable alleles that exhibit increased copper resistance (Hull et al., 2017). Similarly, environmental formaldehyde exposure was shown to



**FIGURE 2 |** Mechanisms of CN variant formation. CN variants typically occur as a result of aberrant replication via homologous recombination, non-homology based mechanisms, and environmentally stimulated processes. **(A)** Unequal crossing over during recombination may result in duplication and deletion. Here, two equal strands of DNA with two genes (represented by the orange or blue arrows) have undergone unequal crossing over due to the misalignment of a homologous sequence. This results in one DNA strand having three genes and the other one gene. **(B,C)** A major driver of CN variant formation is aberrant DNA replication. **(B, top)** Double strand breaks at replication forks or collapsed forks are often repaired via Break-induced replication (BIR). (i) Proper BIR starts with strand invasion of a homologous or microhomologous sequence (shown in red) to allow for proper fork restart. (ii) If template switching occurs in the backward direction, a segment of DNA will have been replicated twice resulting in a duplication; (iii) in contrast, template switching in the forward direction results in a deletion represented by a dashed line in the DNA sequence. Erroneous BIR may be mediated by microhomologies as well. **(C)** CN variants may be stimulated near genes that are highly expressed due to an increased chance of fork stalling. (i) If a replication fork breaks down near a gene that is not expressed (gray) and restarts once (represented by one black arrow), no mutation will occur. (ii) If a replication fork breaks down near a gene that is expressed (green) with cryptic unstable transcripts (red) then there may be two outcomes dependent on the degree of the H3K56ac acetylation mark. If there are low levels of H3K56ac, it is more likely that there will be proper fork restart by BIR (represented by one black arrow). If there are high levels of H3K56ac, it is more likely that there will be repeated fork stalling (represented by three black arrows) (see Figure 8 from Hull et al., 2017).

stimulate CN variation (Hull et al., 2017) of the *SFA1* gene, which confers formaldehyde resistance at higher CNs (Wehner et al., 1993). Altogether, these experiments provide insight to how perturbations of an environmental parameter may stimulate CN variation at a locus associated with adaptation in the new environment (Hull et al., 2017).

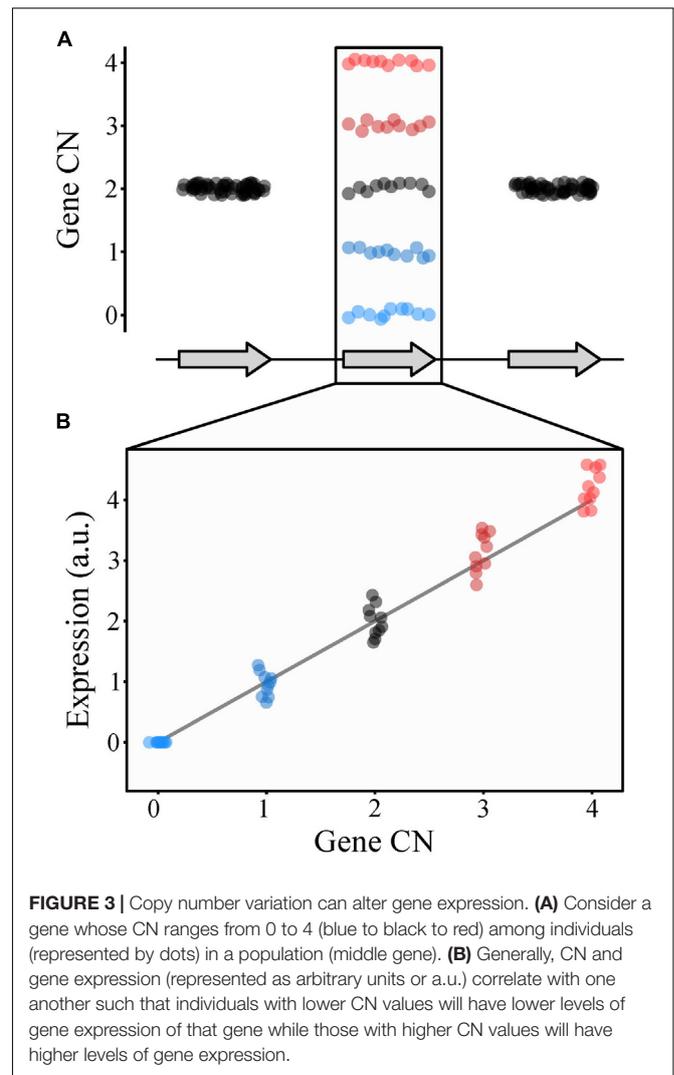
## COPY NUMBER VARIATION AS A SOURCE OF PHENOTYPIC DIVERSITY

Copy number variants can have multiple effects on gene activity, such as changing gene dosage (i.e., gene CN; **Figure 3**) and interrupting coding sequences (Itsara et al., 2009; Sener, 2014). These effects can be substantial; for example, 17.7% of gene expression variation in human populations can be attributed to CN variants (Stranger et al., 2007). Furthermore, changes in human gene expression attributed to CN variants have little overlap with changes in gene expression caused by SNPs, suggesting the two types of variation independently affect gene expression (Stranger et al., 2007). Additionally, gene CN tends to correlate with levels of both gene expression and protein abundance (Perry et al., 2007; Stranger et al., 2007; Henrichsen et al., 2009). For example, changes in gene expression and therefore protein abundance caused by chromosomal CN variation in human chromosome 21 are thought to contribute to Down syndrome (Kahlem et al., 2004; Aivazidis et al., 2017).

## COPY NUMBER VARIATION AS A SOURCE OF GENETIC AND PHENOTYPIC DIVERSITY IN FUNGAL POPULATIONS

Copy number variant loci contribute to population genetic and phenotypic diversity (**Box 1**), such as virulence (Hu et al., 2011; Farrer et al., 2013), in diverse fungal species, including the baker's yeast *Saccharomyces cerevisiae* (ASCOMYCOTA, Saccharomycetes) (Strope et al., 2015; Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk and Rokas, 2017), *Saccharomyces paradoxus* (ASCOMYCOTA, Saccharomycetes) (Bergstrom et al., 2014), the fission yeast *Schizosaccharomyces pombe* (ASCOMYCOTA, Schizosaccharomycetes) (Jeffares et al., 2017), the wheat pathogen *Zymoseptoria tritici* (ASCOMYCOTA, Dothideomycetes) (Hartmann and Croll, 2017), the human fungal pathogens *Cryptococcus deuterogattii* (BASIDIOMYCOTA, Tremellomycetes) (previously known as *Cryptococcus gattii* VGII; Steenwyk et al., 2016) and *C. neoformans* (Hu et al., 2011), and the amphibian pathogen *Batrachochytrium dendrobatidis* (CHYTRIDIOMYCOTA, Chytridiomycetes) (Farrer et al., 2013).

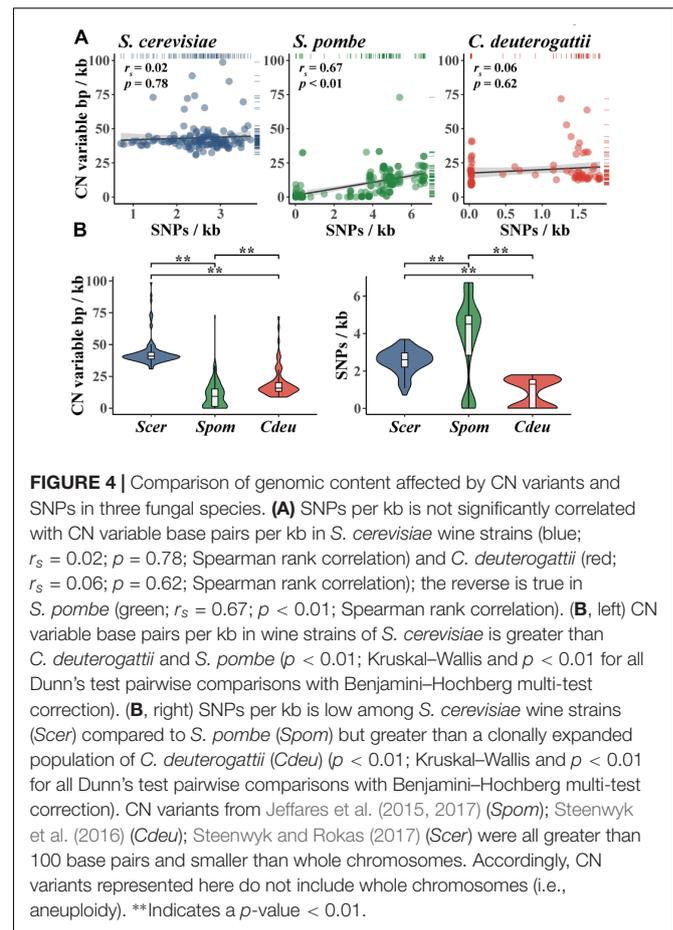
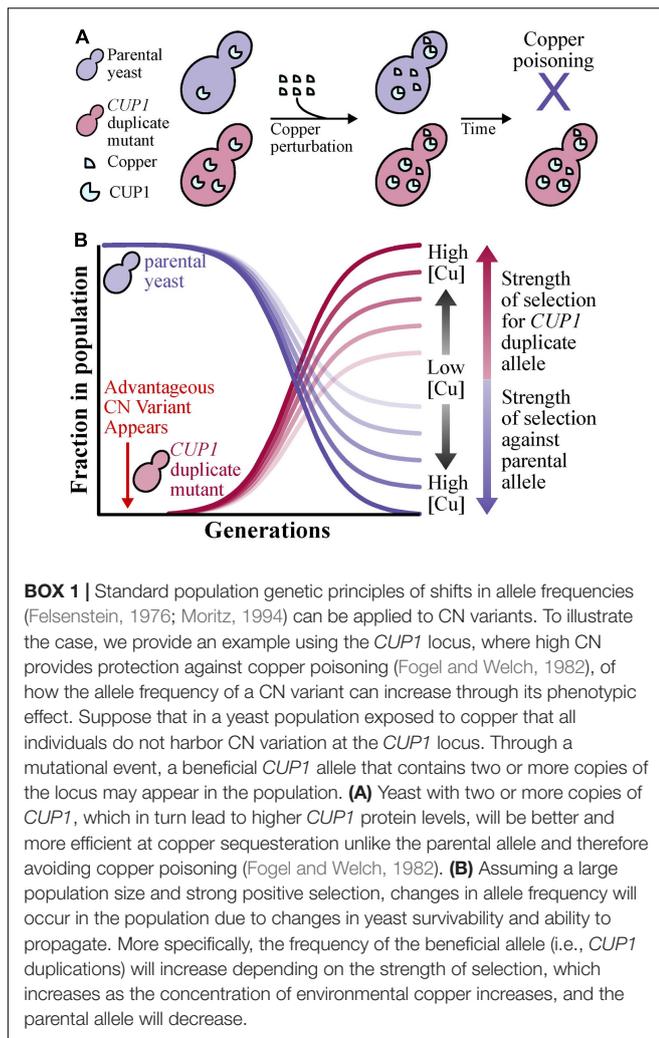
Importantly, the degree of CN variation (which can be represented by CN variable base pairs per kilobase) in fungal populations is not always correlated to the degree of SNP variation (which can be represented by SNPs per kilobase) (**Figure 4A**). For example, there is no correlation between CN variable base pairs per kilobase and SNPs per kilobase among



**FIGURE 3 |** Copy number variation can alter gene expression. **(A)** Consider a gene whose CN ranges from 0 to 4 (blue to black to red) among individuals (represented by dots) in a population (middle gene). **(B)** Generally, CN and gene expression (represented as arbitrary units or a.u.) correlate with one another such that individuals with lower CN values will have lower levels of gene expression of that gene while those with higher CN values will have higher levels of gene expression.

*S. cerevisiae* wine strains (Steenwyk and Rokas, 2017) and a population of *Cryptococcus deuterogattii* (Steenwyk et al., 2016). Interestingly, both populations harbor low levels of SNP diversity; for *S. cerevisiae* wine strains this is due to a single domestication-associated bottleneck event (Liti et al., 2009; Schacherer et al., 2009; Sicard and Legras, 2011; Cromie et al., 2013), whereas for *C. deuterogattii* this is because the samples stem from three clonally evolved subpopulations from the Pacific Northwest, United States (Engelthaler et al., 2014). In contrast, a significant correlation is observed between CN variable base pairs per kilobase and SNPs per kilobase among individuals in a globally distributed population of *S. pombe* (Jeffares et al., 2015).

The proportion of the genome exhibiting CN and SNP variation also varies across *S. cerevisiae*, *S. pombe*, and *C. deuterogattii* populations. For example, CN variable base pairs per kilobase are significantly different between the three populations (**Figure 4B**), with the fraction of CN variable base pairs per kilobase being greatest in *S. cerevisiae* wine strains, followed by *C. deuterogattii*, and then *S. pombe*. Notably, wine strains of *S. cerevisiae* exhibit higher levels of CN variation



than sake strains but lower than beer strains (Gallone et al., 2016). In contrast, there are fewer SNPs per kilobase in the *S. cerevisiae* population compared to *S. pombe* but more compared to *C. deuterogattii* (Figure 4B). Additionally, several different *S. cerevisiae* lineages (e.g., wine, sake, etc.) have more CN variation but less SNP variation than the sister species, *S. paradoxus*, further highlighting the importance of CN variation to *S. cerevisiae* genome evolution (Bergstrom et al., 2014). Interestingly, *S. cerevisiae* CN variants are not evenly distributed across the genome, but tend to occur most frequently within subtelomeric regions (Dunn et al., 2012; Bergstrom et al., 2014). For example, across 132 wine yeast strains, 46 and 67% of the most CN diverse loci and genes, respectively, are observed in the subtelomeric regions (Steenwyk and Rokas, 2017).

How CN variants influence gene expression and phenotype in fungi is not well known. Examination of the contribution of CN variants to gene expression and phenotypic variation in *S. pombe* shows that partial aneuploidies (i.e., large CN variants) influence both local and global gene expression (Chikashige et al., 2007); in addition, CN variants are positively correlated with gene expression changes ( $r_s = 0.71$ ;  $p = 0.01$ ; Spearman

rank correlation; reported in Jeffares et al., 2017). Genome-wide association analyses of numerous phenotypes in *S. pombe* showed that structural variants accounted for 11% of phenotypic variation (CN variants accounted for 7% of that variation and rearrangements for 4%; Jeffares et al., 2017). The phenotypes significantly influenced by CN variants included growth rate, growth in various free amino acids (e.g., tryptophan, isoleucine), growth in the presence of various stressors (e.g., hydrogen peroxide, ultraviolet radiation, minimal media), and sugar utilization in winemaking (Jeffares et al., 2017). However, how much of the phenotypic impact of CN variants is due to genetic drift or adaptation remains largely unknown. Functional analyses of single genes have provided some insight for adaptive CN variants. For example, in *S. cerevisiae*, CN variants have been shown to influence ecologically-relevant phenotypes; *CUP1* duplications have been repeatedly associated with resistance to copper (Fogel and Welch, 1982; Strope et al., 2015) and duplications in the *MAL* loci, which facilitate the utilization of maltose, the main carbon source during beer fermentation and present in sake fermentations, are frequently observed among beer and sake yeast strains, (Vidgren et al., 2005; Gallone et al., 2016; Gonçalves et al., 2016).

Although more studies are needed, these findings argue that CN variation may be a substantial contributor to the total genetic

and phenotypic variation of fungal populations. Additionally, the variation in the correlation between CN and SNP variation across fungal populations (Figure 4) suggests that levels of SNP variation are not always a good proxy for levels of CN variation.

## COPY NUMBER VARIATION AND ITS IMPACT ON WINE YEAST ADAPTATION IN FERMENTATION-RELATED PROCESSES

During the wine making process, *S. cerevisiae* yeasts are barraged with numerous stressors such as high acidity, ethanol, osmolarity, sulfites, and low levels of oxygen and nutrient availability (Marsit and Dequin, 2015). Not surprisingly, *S. cerevisiae* strains isolated from wine making environments tend to be more robust to acid, copper, and sulfite stressors than yeasts isolated from beer and sake environments (Gallone et al., 2016). These biological differences are, at least partially, explained by variants, including CN variants, found at different frequencies or uniquely in wine yeasts. Although it is not known whether most of these CN variant differences are driven by natural selection or genetic drift, CN variation in several cases is associated with ecologically-relevant genes and traits. Below, we discuss what is known about the CN profile of genes from *S. cerevisiae* wine yeast strains associated with these stressors that may reflect diversity in stress tolerance or metabolic capacity and efficiency (Figure 5).

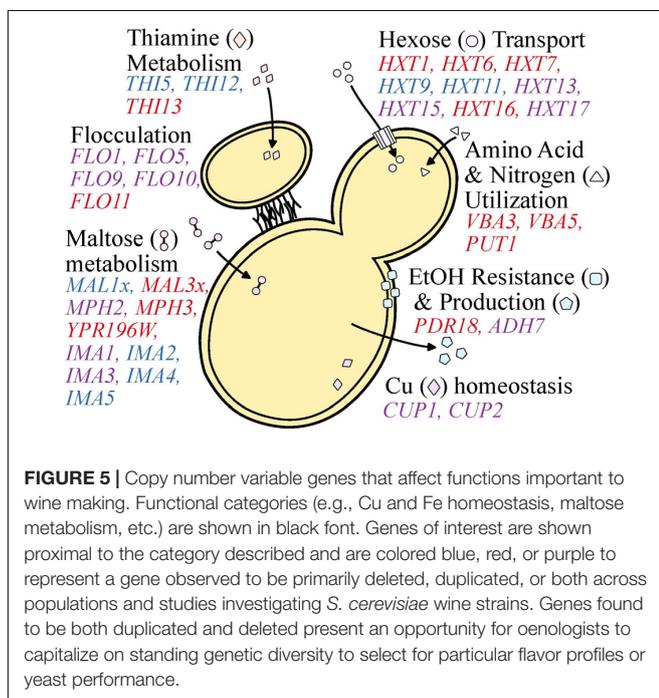
### CN Variable Genes Related to Stress

Many of the CN variable genes that have been identified among wine strains of *S. cerevisiae* (Ibáñez et al., 2014; Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk and Rokas, 2017)

are associated with fermentation processes (Table 1), which supports the hypothesis that CN variation plays a significant role in microbial domestication (Gibbons and Rinker, 2015). For example, *CUP1* is commonly duplicated among wine yeast strains, but not among yeasts in the closely related natural oak lineage (Almeida et al., 2015; Strobe et al., 2015). Duplications in *CUP1* have been shown to confer copper resistance (Warringer et al., 2011) and their occurrence in wine yeast strains may have been driven by the human use of copper as a fungicide to combat powdery mildews in vineyards since the 1800's (Fay et al., 2004; Almeida et al., 2015).

Wine yeasts have also evolved strategies that favor survival in the wine fermentation environment, such as flocculation. This aggregation of yeast cells is associated with escape from hypoxic conditions, as it promotes floating and reaching the air-liquid interface where oxidative metabolism is possible (Martínez et al., 1997; Fidalgo et al., 2006). Flocculation is also favorable for oenologists as it facilitates yeast removal in post-processing (Soares, 2011) and is associated with the production of flavor enhancing ester-containing compounds (Pretorius, 2000). Flocculation is controlled by the *FLO* family of genes (Fidalgo et al., 2006; Govender et al., 2008). Examination of patterns of CN variation in *FLO* gene family members shows frequent duplications in *FLO11* as well as numerous duplications and deletions in *FLO1*, *FLO5*, *FLO9*, and *FLO10* (Gallone et al., 2016; Steenwyk and Rokas, 2017). Additionally, multiple independent studies have reported the GO terms CELL AGGREGATION (GO:0098743) and AGGREGATION OF UNICELLULAR ORGANISMS (GO:0098630) to be significantly enriched among CN variable genes in wine yeasts (Gallone et al., 2016; Steenwyk and Rokas, 2017). Interestingly, the same GO terms are only enriched among deleted genes in the beer and Asia/sake lineages (Gallone et al., 2016) suggesting these genes may be particularly important for wine yeasts. In fact, this has been demonstrated for “flor” or “sherry” yeasts, where partial duplications in the Serine/Threonine-rich hydrophobic region of *FLO11* are associated with the adaptive phenotype of floating to the air-liquid interface to access oxygen (Fidalgo et al., 2006). Furthermore, the same partial duplications have also been observed in the more general wine lineage (Steenwyk and Rokas, 2017), suggesting that the benefits associated with this phenotype may not be unique to “flor” yeasts.

Copy number variation is also observed in genes related to stuck (incomplete) or sluggish (delayed) fermentations. Stuck fermentations are caused by a multitude of factors including nitrogen availability, nutrient transport, and decreased resistance to starvation (Salmon, 1989; Thomsson et al., 2005). Two genes associated with decrease resistance to starvation, *ADH7* and *AAD3*, are sometimes duplicated or deleted among wine yeast strains (Steenwyk and Rokas, 2017). Diverse CN profiles of *ADH7*, an alcohol dehydrogenase that reduces acetaldehyde to ethanol during glucose fermentation, and *AAD3*, an aryl-alcohol dehydrogenase whose null mutant displays greater starvation sensitivity (Walker et al., 2014), suggest variable degrees of starvation sensitivity and therefore fermentation performance. Additionally, wine yeasts are enriched for duplication in *PDR18* (Gallone et al., 2016), a transporter that aids in resistance to



**TABLE 1** | Genes associated with fermentation-related processes that exhibit CN variation among *Saccharomyces cerevisiae* wine strains.

Process (organized alphabetically)	Gene	Primarily duplicated, deleted, or both	References (organized by publication date)
Amino acid and nitrogen utilization	<i>VBA3, VBA5, PUT1</i>	Duplicated	Ibáñez et al., 2014; Gallone et al., 2016
Cu and Fe homeostasis	<i>CUP1, CUP2</i>	Both	Fay et al., 2004; Warringer et al., 2011; Almeida et al., 2015; Steenwyk and Rokas, 2017
Ethanol resistance and production	<i>FIT2, FIT3, FRE3, PDR18</i>	Duplicated	Gallone et al., 2016 Gallone et al., 2016
	<i>ADH7</i>	Both	Steenwyk and Rokas, 2017
Flocculation	<i>FLO11</i>	Duplicated	Steenwyk and Rokas, 2017
	<i>FLO1, FLO5, FLO9, FLO10</i>	Both	Gallone et al., 2016; Steenwyk and Rokas, 2017
Hexose transport	<i>HXT1, HXT4, HXT6, HXT7, HXT16</i>	Duplicated	Gallone et al., 2016; Steenwyk and Rokas, 2017
	<i>HXT9, HXT11</i>	Deleted	Gallone et al., 2016; Steenwyk and Rokas, 2017
	<i>HXT13, HXT15, HXT17</i>	Both	Gallone et al., 2016; Steenwyk and Rokas, 2017
Maltose metabolism	<i>MAL3x, MPH3, YPR196W</i>	Duplicated	Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk and Rokas, 2017
	<i>MAL1x, IMA2, IMA4, IMA5</i>	Deleted	Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk and Rokas, 2017
	<i>MPH2, IMA1, IMA3</i>	Both	Gallone et al., 2016; Steenwyk and Rokas, 2017
Thiamine metabolism	<i>THI3</i>	Duplicated	Steenwyk and Rokas, 2017
	<i>THI5, THI2</i>	Deleted	Steenwyk and Rokas, 2017

ethanol stress, one of the traits that differentiates wine from other industrial strains. Another gene associated with decreased resistance to starvation that also exhibits CN variation is *IMA1* (Steenwyk and Rokas, 2017), a major isomaltase with glucosidase activity (Teste et al., 2010).

## CN Variable Genes Related to Metabolism

Nutrient availability and acquisition is a major driving factor of wine fermentation outcome. Among the most important nutrients dictating the pace and success of wine fermentation is sugar availability (Marsit and Dequin, 2015). The most abundant fermentable hexose sugars in the wine environment include glucose and fructose (Marques et al., 2015), whose transport is largely carried out by genes from the hexose transporter (*HXT*) family (Boles and Hollenberg, 1997). A reproducible evolutionary outcome of yeasts exposed to glucose-limited environments, which are reflective of late wine fermentation, is duplication in the high-affinity hexose transporters, such as *HXT6* and *HXT7* (Brown et al., 1998; Dunham et al., 2002; Gresham et al., 2008, 2010), suggesting that changes in transporter CN are adaptive. Interestingly, GO terms such as HEXOSE TRANSMEMBRANE TRANSPORT (GO:0035428), GLUCOSE IMPORT (GO:0046323), and MONOSACCHARAIDE TRANSPORT (GO:0015749) are significantly enriched among duplicated CN variable genes in the wine lineage primarily due to duplications repeatedly observed in the *HXT* gene

family among wine yeast strains (Dunn et al., 2012; Gallone et al., 2016; Steenwyk and Rokas, 2017). More specifically, *HXT13*, *HXT15*, and *HXT17* exhibit CN variation among wine strains, *HXT1*, *HXT6*, *HXT7*, and *HXT16* are more commonly duplicated, and *HXT9* and *HXT11* are more commonly deleted (Gallone et al., 2016; Steenwyk and Rokas, 2017).

Similarly striking patterns of CN variation are observed for genes associated with maltose metabolism (Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk and Rokas, 2017). The two *MAL* loci in the reference genome of *S. cerevisiae* S288C, *MAL1*, and *MAL3*, that contain three genes which encode for a permease (*MALx1*), a maltase (*MALx2*), and a *trans*-activator (*MALx3*) (Michels et al., 1992; Naumov et al., 1994). The *MAL* loci are primarily associated with the metabolism of maltose (Michels et al., 1992), an abundant sugar during beer fermentation, and are commonly duplicated among beer yeast strains (Gallone et al., 2016; Gonçalves et al., 2016), however, this locus would be expected to be primarily deleted among wine yeasts as maltose is in relatively low abundance compared to other sugars during wine fermentation. As expected, MALTOSE METABOLIC PROCESS (GO:0000023) is among the significantly enriched GO terms across deleted genes in the wine yeast strains (Gallone et al., 2016) due to the deletion of the *MAL1* locus (Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk and Rokas, 2017). In contrast, the *MAL3* locus is primarily duplicated among wine yeast strains (Gonçalves et al., 2016; Steenwyk and Rokas, 2017). Interestingly,

part of the *MAL3* locus, *MAL32*, has been demonstrated to be important for growth on turanose, maltotriose, and sucrose (Brown et al., 2010), which are present in the wine environment, albeit in small quantities (Victoria Moreno-Arribas and Carmen, 2013), suggesting potential function on secondary substrates or perhaps another function.

Equally important as sugar availability in determining fermentation outcome is nitrogen acquisition (Marsit and Dequin, 2015). Genes associated with amino acid and nitrogen utilization are commonly duplicated among wine yeast strains. Notable examples of such duplications are the amino acid permeases, *VBA3* and *VBA5* (Gallone et al., 2016), and *PUT1*, a gene that aids in the recycling or utilization of proline (Ibáñez et al., 2014).

Copy number variation is also observed in genes of the *THI* family, which are all involved in biosynthesis of hydroxymethylpyrimidine, a thiamine, or vitamin B<sub>1</sub>, precursor (Rodríguez-Navarro et al., 2002; Wightman and Meacock, 2003; Li et al., 2010), another important determinant of wine fermentation outcome. Several *THI* gene family members are CN variable; *THI5* and *THI12* are typically deleted, while *THI13* is commonly duplicated (Steenwyk and Rokas, 2017). Expression of *THI5* is commonly repressed or absent in wine strains, as it is associated with an undesirable rotten-egg smell and taste in wine (Bartra et al., 2010; Brion et al., 2014). Interestingly, *THI5* is deleted in greater than 90% of examined wine strains (Steenwyk and Rokas, 2017) but is duplicated in several other strains of *S. cerevisiae*, as well as in its sister species *S. paradoxus* and the hybrid species *S. pastorianus* (Wightman and Meacock, 2003).

## CONCLUSION AND PERSPECTIVES

An emerging body of work suggests that CN variation is an important, largely underappreciated, dimension of fungal genome biology and evolution (Hu et al., 2011; Farrer et al., 2013; Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk et al., 2016; Hartmann and Croll, 2017; Steenwyk and Rokas, 2017). Not surprisingly, numerous questions remain unresolved. For example, we have detailed numerous mechanisms that lead to the generation of CN variation but the relative contribution of each remains unclear. Additionally, both the genomic organization and genetic architecture of CN variants remain largely unknown. For example, are duplicated copies typically found in the same genomic neighborhood or are they dispersed? Similarly, what percentage of phenotypic differences among fungal strains is explained by CN variation?

The same can be said about the role of CN variation in yeast adaptation to the wine fermentation environment. We still lack computational methods for distinguishing the footprint of natural selection and genetic drift on CN variation. Comparison of genome-wide patterns of CN variation among yeast populations responsible for the fermentation of different wines (e.g., white and red), coupled with functional studies, would provide insight to how human activity has shaped the genome of yeasts associated with particular types of

wine. Additionally, most sequenced wine strains originate from Italy, Australia, or France. Genome sequencing of yeasts from underrepresented regions (e.g., Africa and the Americas) may provide further insight to CN variable loci unique to each region and the global diversity of wine yeast genomes.

Another major set of questions are associated with examining the impact of CN variable loci at the different stages of wine fermentation. Insights on how CN variable loci modify gene expression, protein abundance and in turn fermentation behavior and end-product would be immensely valuable. A complementary, perhaps more straightforward, approach would be focused on examining the phenotypic impact of single-gene or gene family CN variants, such as the ones discussed in previous sections (e.g., genes belonging to the *ADH*, *HXT*, *MAL*, and *VBA* families; **Table 1**) on fermentation outcome; this approach would also aid distinguishing adaptive and neutral CN variants. Such studies may provide an important bridge between scientist, oenologist, and wine-maker to enhance fermentation efficiency and consistency between batches or in the design of new wine flavor profiles.

Although this review focused solely on the contribution of *S. cerevisiae* CN variation, it is important to keep in mind that several other yeasts are also part of the wine fermentation environment. Members of many other wine yeast genera (e.g., *Hanseniaspora*, *Saccharomyces*, and *Torulaspora*) are known to modify properties wine fermentation end product (Ciani and Maccarelli, 1998). Furthermore, recent sequencing projects have made several non-conventional wine yeast genomes publicly available such as several *Hanseniaspora* species (Sternes et al., 2016; Seixas et al., 2017), *Starmerella bacillaris* (Lemos Junior et al., 2017), *Lachancea lanzarotensis* (Sarilar et al., 2015), and *Brettanomyces bruxellensis*, which has already been demonstrated to harbor CN variants (Curtin et al., 2012). In-depth sequencing of populations from these yeast species and others associated with wine will provide insight to niche specialization within the wine environment as well as greatly enhance our understanding of CN variation and its role in the ecology and evolution of fungal populations.

## AUTHOR CONTRIBUTIONS

JS and AR chose the topic of the review and identified the areas that it would cover and the figures that it would contain. JS wrote the first draft of the manuscript and designed the figures. AR provided several rounds of extensive feedback on both the manuscript and the figures.

## FUNDING

JS was supported by the Graduate Program in Biological Sciences at Vanderbilt University. Research in AR's laboratory was supported by the National Science Foundation (DEB-1442113), the Burroughs Wellcome Fund, and the March of Dimes Prematurity Research Center Ohio Collaborative.

## REFERENCES

- Aivazidis, S., Coughlan, C. M., Rauniyar, A. K., Jiang, H., Liggett, L. A., Maclean, K. N., et al. (2017). The burden of trisomy 21 disrupts the proteostasis network in down syndrome. *PLoS One* 12:e0176307. doi: 10.1371/journal.pone.0176307
- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., et al. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Mol. Ecol.* 24, 5412–5427. doi: 10.1111/mec.13341
- Ambroset, C., Petit, M., Brion, C., Sanchez, I., Delobel, P., Guérin, C., et al. (2011). Deciphering the molecular basis of wine yeast fermentation traits using a combined genetic and genomic approach. *G3* 1, 263–281. doi: 10.1534/g3.111.000422
- Arlt, M. F., Rajendran, S., Birkeland, S. R., Wilson, T. E., and Glover, T. W. (2014). Copy number variants are produced in response to low-dose ionizing radiation in cultured cells. *Environ. Mol. Mutagen.* 55, 103–113. doi: 10.1002/em.21840
- Aylon, Y., and Kupiec, M. (2004). New insights into the mechanism of homologous recombination in yeast. *Mutat. Res.* 566, 231–248. doi: 10.1016/j.mrrev.2003.10.001
- Bailey, J. A., and Eichler, E. E. (2006). Primate segmental duplications: crucibles of evolution, diversity and disease. *Nat. Rev. Genet.* 7, 552–564. doi: 10.1038/nrg1895
- Bartra, E., Casado, M., Carro, D., Campamà, C., and Piña, B. (2010). Differential expression of thiamine biosynthetic genes in yeast strains with high and low production of hydrogen sulfide during wine fermentation. *J. Appl. Microbiol.* 109, 272–281. doi: 10.1111/j.1365-2672.2009.04652.x
- Bergstrom, A., Simpson, J. T., Salinas, F., Barre, B., Parts, L., Zia, A., et al. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Mol. Biol. Evol.* 31, 872–888. doi: 10.1093/molbev/msu037
- Boles, E., and Hollenberg, C. P. (1997). The molecular genetics of hexose transport in yeasts. *FEMS Microbiol. Rev.* 21, 85–111. doi: 10.1016/S0168-6445(97)00052-1
- Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., and Schmidt, S. A. (2016). Whole genome comparison reveals high levels of inbreeding and strain redundancy across the spectrum of commercial wine strains of *Saccharomyces cerevisiae*. *G3* 6, 957–971. doi: 10.1534/g3.115.025692
- Botstein, D., Chervitz, S. A., and Cherry, J. M. (1997). Yeast as a model organism. *Science* 227, 1259–1260. doi: 10.1126/science.277.5330.1259
- Brauer, M. J., Christianson, C. M., Pai, D. A., and Dunham, M. J. (2006). Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. *Genetics* 173, 1813–1816. doi: 10.1534/genetics.106.057927
- Brion, C., Ambroset, C., Delobel, P., Sanchez, I., and Blondin, B. (2014). Deciphering regulatory variation of THI genes in alcoholic fermentation indicate an impact of Thi3p on PDC1 expression. *BMC Genomics* 15:1085. doi: 10.1186/1471-2164-15-1085
- Brown, C. A., Murray, A. W., and Verstrepen, K. J. (2010). Rapid expansion and functional divergence of subtelomeric gene families in yeasts. *Curr. Biol.* 20, 895–903. doi: 10.1016/j.cub.2010.04.027
- Brown, C. J., Todd, K. M., and Rosenzweig, R. F. (1998). Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* 15, 931–942. doi: 10.1093/oxfordjournals.molbev.a026009
- Carvalho, C. M. B., and Lupski, J. R. (2016). Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* 17, 224–238. doi: 10.1038/nrg.2015.25
- Chakravarti, A. (1999). Population genetics - Making sense out of sequence. *Nat. Genet.* 21, 56–60. doi: 10.1038/4482
- Chikashige, Y., Tsutsumi, C., Okamura, K., Yamane, M., Nakayama, J., Niwa, O., et al. (2007). Gene expression and distribution of Swi6 in partial aneuploids of the fission yeast *Schizosaccharomyces pombe*. *Cell Struct. Funct.* 32, 149–161. doi: 10.1247/csf.07036
- Ciani, M., and Maccarelli, F. (1998). Oenological properties of non-*Saccharomyces* yeasts associated with wine-making. *World J. Microbiol. Biotechnol.* 14, 199–203. doi: 10.1023/A:1008825928354
- Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., et al. (2012). Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* 338, 1206–1209. doi: 10.1126/science.1228746
- Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., et al. (2013). Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3* 3, 2163–2171. doi: 10.1534/g3.113.007492
- Curtin, C. D., Borneman, A. R., Chambers, P. J., and Pretorius, I. S. (2012). De-novo assembly and analysis of the heterozygous triploid genome of the wine spoilage yeast *dekkera bruxellensis* AWRI1499. *PLoS One* 7:e33840. doi: 10.1371/journal.pone.0033840
- Daley, J. M., Palmos, P. L., Wu, D., and Wilson, T. E. (2005). Nonhomologous end joining in yeast. *Annu. Rev. Genet.* 39, 431–451. doi: 10.1146/annurev.genet.39.073003.113340
- Dunham, M. J., Badrane, H., Ferea, T., Adams, J., Brown, P. O., Rosenzweig, F., et al. (2002). Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 16144–16149. doi: 10.1073/pnas.242624799
- Dunn, B., Richter, C., Kvitik, D. J., Pugh, T., and Sherlock, G. (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, 908–924. doi: 10.1101/gr.130310.111
- Engelthaler, D. M., Hicks, N. D., Gillece, J. D., Roe, C. C., Schupp, J. M., Driebe, E. M., et al. (2014). *Cryptococcus gattii* in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. *mBio* 5:e01464-14. doi: 10.1128/mBio.01464-14
- Estivill, X., and Armengol, L. (2007). Copy number variants and common disorders: filling the gaps and exploring complexity in genome-wide association studies. *PLoS Genet.* 3, 1787–1799. doi: 10.1371/journal.pgen.0030190
- Farrer, R. A., Henk, D. A., Garner, T. W., Balloux, F., Woodhams, D. C., and Fisher, M. C. (2013). Chromosomal copy number variation, selection and uneven rates of recombination reveal cryptic genome diversity linked to pathogenicity. *PLoS Genet.* 9:e1003703. doi: 10.1371/journal.pgen.1003703
- Fay, J. C., McCullough, H. L., Sniegowski, P. D., and Eisen, M. B. (2004). Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biol.* 5:R26. doi: 10.1186/gb-2004-5-4-r26
- Felsenstein, J. (1976). The theoretical population genetics of variable selection and migration. *Annu. Rev. Genet.* 10, 253–280. doi: 10.1146/annurev.ge.10.120176.001345
- Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome. *Nat. Rev. Genet.* 7, 85–97. doi: 10.1038/nrg1767
- Fidalgo, M., Barrales, R. R., Ibeas, J. I., and Jimenez, J. (2006). Adaptive evolution by mutations in the FLO11 gene. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11228–11233. doi: 10.1073/pnas.0601713103
- Fogel, S., and Welch, J. W. (1982). Tandem gene amplification mediates copper resistance in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 79, 5342–5346. doi: 10.1073/pnas.79.17.5342
- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* 166, 1397–1410.e16. doi: 10.1016/j.cell.2016.08.020
- Gibbons, J. G., and Rinker, D. C. (2015). The genomics of microbial domestication in the fermented food environment. *Curr. Opin. Genet. Dev.* 35, 1–8. doi: 10.1016/j.gde.2015.07.003
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996). Life with 6000 genes. *Science* 274, 546–567. doi: 10.1126/science.274.5287.546
- Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., et al. (2016). Distinct domestication trajectories in top-fermenting beer yeasts and wine yeasts. *Curr. Biol.* 26, 2750–2761. doi: 10.1016/j.cub.2016.08.040
- Govender, P., Domingo, J. L., Bester, M. C., Pretorius, I. S., and Bauer, F. F. (2008). Controlled expression of the dominant flocculation genes FLO1, FLO5, and FLO11 in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 74, 6041–6052. doi: 10.1128/AEM.00394-08
- Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A., et al. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* 4:e1000303. doi: 10.1371/journal.pgen.1000303
- Gresham, D., Usaite, R., Germann, S. M., Lisby, M., Botstein, D., and Regenberg, B. (2010). Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal element formation of the GAPI locus. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18551–18556. doi: 10.1073/pnas.1014023107

- Gu, S., Szafranski, P., Akdemir, Z. C., Yuan, B., Cooper, M. L., Magriñá, M. A., et al. (2016). Mechanisms for complex chromosomal insertions. *PLoS Genet.* 12:e1006446. doi: 10.1371/journal.pgen.1006446
- Gu, W., Zhang, F., and Lupski, J. R. (2008). Mechanisms for human genomic rearrangements. *Pathogenetics* 1:4. doi: 10.1186/1755-8417-1-4
- Hartmann, F. E., and Croll, D. (2017). Distinct trajectories of massive recent gene gains and losses in populations of a microbial eukaryotic pathogen. *Mol. Biol. Evol.* 34, 2808–2822. doi: 10.1093/molbev/msx208
- Hastings, P. J., Ira, G., and Lupski, J. R. (2009a). A microhomology-mediated break-induced replication model for the origin of human copy number variation. *PLoS Genet.* 5:e1000327. doi: 10.1371/journal.pgen.1000327
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. (2009b). Mechanisms of change in gene copy number. *Nat. Rev. Genet.* 10, 551–564. doi: 10.1038/nrg2593
- Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., et al. (2009). Segmental copy number variation shapes tissue transcriptomes. *Nat. Genet.* 41, 424–429. doi: 10.1038/ng.345
- Hu, G., Wang, J., Choi, J., Jung, W. H., Liu, I., Litvintseva, A. P., et al. (2011). Variation in chromosome copy number influences the virulence of *Cryptococcus neoformans* and occurs in isolates from AIDS patients. *BMC Genomics* 12:526. doi: 10.1186/1471-2164-12-526
- Hua, S., Qiu, M., Chan, E., Zhu, L., and Luo, Y. (1997). Minimum length of sequence homology required for *in vivo* cloning by homologous recombination in yeast. *Plasmid* 38, 91–96. doi: 10.1006/plas.1997.1305
- Hull, R. M., Cruz, C., Jack, C. V., Houseley, J., Noronha, M., and Calderon, L. (2017). Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol.* 15:e2001333. doi: 10.1371/journal.pbio.2001333
- Hurles, M. (2005). How homologous recombination generates a mutable genome. *Hum. Genomics* 2, 179–186. doi: 10.1186/1479-7364-2-3-179
- Ibáñez, C., Pérez-Torrado, R., Chiva, R., Guillamón, J. M., Barrio, E., and Querol, A. (2014). Comparative genomic analysis of *Saccharomyces cerevisiae* yeasts isolated from fermentations of traditional beverages unveils different adaptive strategies. *Int. J. Food Microbiol.* 171, 129–135. doi: 10.1016/j.ijfoodmicro.2013.10.023
- Itsara, A., Cooper, G. M., Baker, C., Girirajan, S., Li, J., Absher, D., et al. (2009). Population analysis of large copy number variants and hotspots of human genetic disease. *Am. J. Hum. Genet.* 84, 148–161. doi: 10.1016/j.ajhg.2008.12.014
- Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., et al. (2017). Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* 8:14061. doi: 10.1038/ncomms14061
- Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Převedrovský, M., Mourier, T., et al. (2015). The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nat. Genet.* 47, 235–241. doi: 10.1038/ng.3215
- Kahlem, P., Sultan, M., Herwig, R., Steinfath, M., Balzereit, D., Eppens, B., et al. (2004). Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse model of down syndrome. *Genome Res.* 14, 1258–1267. doi: 10.1101/gr.1951304
- Khanna, K. K., and Jackson, S. P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet.* 27, 247–254. doi: 10.1038/85798
- Kupiec, M., and Petes, T. D. (1988). Allelic and ectopic recombination between Ty elements in yeast. *Genetics* 119, 549–559.
- Lemos Junior, W. J. F., Treu, L., da Silva Duarte, V., Carlot, M., Nadai, C., Campanaro, S., et al. (2017). Whole-genome sequence of *Starmerella bacillaris* PAS13, a nonconventional ecological yeast with antifungal activity. *Genome Announc.* 5:e00788-17. doi: 10.1128/genomeA.00788-17
- Li, M., Petteys, B. J., McClure, J. M., Valsakumar, V., Bekiranov, S., Frank, E. L., et al. (2010). Thiamine biosynthesis in *Saccharomyces cerevisiae* is regulated by the NAD<sup>+</sup>-dependent histone deacetylase Hst1. *Mol. Cell. Biol.* 30, 3329–3341. doi: 10.1128/MCB.01590-09
- Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *J. Biol. Chem.* 283, 1–5. doi: 10.1074/jbc.R700039200
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009). Population genomics of domestic and wild yeasts. *Nature* 458, 337–341. doi: 10.1038/nature07743
- Llorente, B., Smith, C. E., and Symington, L. S. (2008). Break-induced replication: what is it and what is it for? *Cell Cycle* 7, 859–864. doi: 10.4161/cc.7.7.5613
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Lynch, V. J. (2007). Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol. Biol.* 7:2. doi: 10.1186/1471-2148-7-2
- Marques, W. L., Raghavendran, V., Stambuk, B. U., and Gombert, A. K. (2015). Sucrose and *Saccharomyces cerevisiae*: a relationship most sweet. *FEMS Yeast Res.* 16:fov107. doi: 10.1093/femsyr/fov107
- Marsit, S., and Dequin, S. (2015). Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. *FEMS Yeast Res.* 15:fov067. doi: 10.1093/femsyr/fov067
- Martínez, P., Pérez Rodríguez, L., and Benítez, T. (1997). Velum formation by flor yeasts isolated from sherry wine. *Am. J. Enol. Vitic.* 48, 55–62.
- McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12273–12278. doi: 10.1073/pnas.0900992106
- McVey, M., and Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet.* 24, 529–538. doi: 10.1016/j.tig.2008.08.007
- Michels, C. A., Read, E., Nat, K., and Charron, M. J. (1992). The telomere-associated MAL3 locus of *Saccharomyces* is a tandem array of repeated genes. *Yeast* 8, 655–665. doi: 10.1002/yea.320080809
- Moritz, C. (1994). Defining “evolutionarily significant units” for conservation. *Trends Ecol. Evol.* 9, 373–375. doi: 10.1016/0169-5347(94)90057-4
- Morrow, D. M., Connelly, C., and Hieter, P. (1997). “Break copy” duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. *Genetics* 147, 371–382.
- Naumov, G. I., Naumova, E. S., and Michels, C. A. (1994). Genetic variation of the repeated MAL loci in natural populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genetics* 136, 803–812. doi: 10.1002/yea.320080809
- Ortiz, J. F., and Rokas, A. (2017). CTDGFinder: a novel homology-based algorithm for identifying clusters of closely spaced clusters of tandemly duplicated genes. *Mol. Biol. Evol.* 34, 215–229. doi: 10.1093/molbev/msw227
- Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E., and Merrikh, H. (2013). Accelerated gene evolution through replication–transcription conflicts. *Nature* 495, 512–515. doi: 10.1038/nature11989
- Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., et al. (2007). Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* 39, 1256–1260. doi: 10.1038/ng2123
- Petukhova, G., Stratton, S., and Sung, P. (1998). Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* 393, 91–94. doi: 10.1038/30037
- Pezer, Z., Harr, B., Teschke, M., Babiker, H., and Tautz, D. (2015). Divergence patterns of genic copy number variation in natural populations of the house mouse (*Mus musculus domesticus*) reveal three conserved genes with major population-specific expansions. *Genome Res.* 25, 1114–1124. doi: 10.1101/gr.187187.114
- Prado, F., Cortés-Ledesma, F., Huertas, P., and Aguilera, A. (2003). Mitotic recombination in *Saccharomyces cerevisiae*. *Curr. Genet.* 42, 185–198. doi: 10.1007/s00294-002-0346-3
- Pretorius, I. S. (2000). Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* 16, 675–729. doi: 10.1002/1097-0061(20000615)16:8<675::AID-YEA585>3.0.CO;2-B
- Rahim, N. G., Harismendy, O., Topol, E. J., and Frazer, K. A. (2008). Genetic determinants of phenotypic diversity in humans. *Genome Biol.* 9:215. doi: 10.1186/gb-2008-9-4-215
- Rausch, T., Jones, D. T. W., Zapatka, M., Stütz, A. M., Zichner, T., Weischenfeldt, J., et al. (2012). Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* 148, 59–71. doi: 10.1016/j.cell.2011.12.013
- Reams, A. B., and Roth, J. R. (2015). Mechanisms of gene duplication and amplification. *Cold Spring Harb. Perspect. Biol.* 7:a016592. doi: 10.1101/cshperspect.a016592
- Riethman, H. (2009). Human subtelomeric copy number variations. *Cytogenet. Genome Res.* 123, 244–252. doi: 10.1159/000184714
- Rodríguez-Navarro, S., Llorente, B., Rodríguez-Manzanque, M. T., Ramne, A., Uber, G., Marchesan, D., et al. (2002). Functional analysis of yeast gene

- families involved in metabolism of vitamins B1 and B6. *Yeast* 19, 1261–1276. doi: 10.1002/yea.916
- Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933. doi: 10.1038/35057149
- Salmon, J. M. (1989). Effect of sugar transport inactivation in *Saccharomyces cerevisiae* on sluggish and stuck enological fermentations. *Appl. Environ. Microbiol.* 55, 953–958.
- Sarilar, V., Devillers, H., Freel, K. C., Schacherer, J., and Neuvéglise, C. (2015). Draft genome sequence of *Lachancea lanzarotensis* CBS 12615 T, an ascomycetous yeast isolated from grapes. *Genome Announc.* 3:e00292-15. doi: 10.1128/genomeA.00292-15
- Schacherer, J., Shapiro, J. A., Ruderfer, D. M., and Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458, 342–345. doi: 10.1038/nature07670
- Seixas, I., Barbosa, C., Salazar, S. B., Mendes-Faia, A., Wang, Y., Güldener, U., et al. (2017). Genome sequence of the nonconventional wine yeast *Hanseniaspora guilliermondii* UTAD222. *Genome Announc.* 5:e01515-16. doi: 10.1128/genomeA.01515-16
- Sener, E. F. (2014). Association of copy number variations in Autism spectrum disorders: a systematic review. *Chinese J. Biol.* 2014:713109. doi: 10.1155/2014/713109
- Sicard, D., and Legras, J. L. (2011). Bread, beer and wine: yeast domestication in the *Saccharomyces sensu stricto* complex. *C. R. Biol.* 334, 229–236. doi: 10.1016/j.crvi.2010.12.016
- Slack, A., Thornton, P. C., Magner, D. B., Rosenberg, S. M., and Hastings, P. J. (2006). On the mechanism of gene amplification induced under stress in *Escherichia coli*. *PLoS Genet.* 2:e48. doi: 10.1371/journal.pgen.0020048
- Smith, C. E., Llorente, B., and Symington, L. S. (2007). Template switching during break-induced replication. *Nature* 447, 102–105. doi: 10.1038/nature05723
- Soares, E. V. (2011). Flocculation in *Saccharomyces cerevisiae*: a review. *J. Appl. Microbiol.* 110, 1–18. doi: 10.1111/j.1365-2672.2010.04897.x
- Soria, P. S., McGary, K. L., and Rokas, A. (2014). Functional divergence for every paralog. *Mol. Biol. Evol.* 31, 984–992. doi: 10.1093/molbev/msu050
- Steenwyk, J., and Rokas, A. (2017). Extensive copy number variation in fermentation-related genes among *Saccharomyces cerevisiae* wine strains. *G3* 7, 1475–1485. doi: 10.1534/g3.117.040105
- Steenwyk, J. L., Soghigian, J. S., Perfect, J. R., and Gibbons, J. G. (2016). Copy number variation contributes to cryptic genetic variation in outbreak lineages of *Cryptococcus gattii* from the North American Pacific Northwest. *BMC Genomics* 17:700. doi: 10.1186/s12864-016-3044-0
- Sternes, P. R., Lee, D., Kutyna, D. R., and Borneman, A. R. (2016). Genome sequences of three species of *Hanseniaspora* isolated from spontaneous wine fermentations. *Genome Announc.* 4:e01287-16. doi: 10.1128/genomeA.01287-16
- Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., et al. (2007). Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315, 848–853. doi: 10.1126/science.1136678
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., 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, 762–774. doi: 10.1101/gr.185538.114
- Sudmant, P. H., Mallick, S., Nelson, B. J., Hormozdiari, F., Krumm, N., Huddleston, J., et al. (2015). Global diversity, population stratification, and selection of human copy-number variation. *Science* 349:aab3761. doi: 10.1126/science.aab3761
- Swanson-Wagner, R. A., Eichten, S. R., Kumari, S., Tiffin, P., Stein, J. C., Ware, D., et al. (2010). Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* 20, 1689–1699. doi: 10.1101/gr.109165.110
- Taylor, M. B., Phan, J., Lee, J. T., McCadden, M., and Ehrenreich, I. M. (2016). Diverse genetic architectures lead to the same cryptic phenotype in a yeast cross. *Nat. Commun.* 7:11669. doi: 10.1038/ncomms11669
- Teste, M. A., Marie François, J., and Parrou, J. L. (2010). Characterization of a new multigene family encoding isomaltases in the yeast *Saccharomyces cerevisiae*, the IMA family. *J. Biol. Chem.* 285, 26815–26824. doi: 10.1074/jbc.M110.145946
- Thomsson, E., Gustafsson, L., and Larsson, C. (2005). Starvation response of *Saccharomyces cerevisiae* grown in anaerobic nitrogen- or carbon-limited chemostat cultures. *Appl. Environ. Microbiol.* 71, 3007–3013. doi: 10.1128/AEM.71.6.3007-3013.2005
- Turner, B. M. (2009). Epigenetic responses to environmental change and their evolutionary implications. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 3403–3418. doi: 10.1098/rstb.2009.0125
- Usher, C. L., and McCarroll, S. A. (2015). Complex and multi-allelic copy number variation in human disease. *Brief. Funct. Genomics* 14, 329–338. doi: 10.1093/bfgp/elv028
- Victoria Moreno-Arribas, M., and Carmen, P. M. (2013). Wine chemistry and biochemistry. *J. Chem. Inf. Model.* 53, 1689–1699. doi: 10.1017/CBO9781107415324.004
- Vidgren, V., Ruohonen, L., and Londesborough, J. (2005). Characterization and functional analysis of the MAL and MPH loci for maltose utilization in some ale and lager yeast strains. *Appl. Environ. Microbiol.* 71, 7846–7857. doi: 10.1128/AEM.71.12.7846-7857.2005
- Walker, M. E., Nguyen, T. D., Liccioli, T., Schmid, F., Kalatzis, N., Sundstrom, J. F., et al. (2014). Genome-wide identification of the Fermentome; genes required for successful and timely completion of wine-like fermentation by *Saccharomyces cerevisiae*. *BMC Genomics* 15:552. doi: 10.1186/1471-2164-15-552
- Warringer, J., Zörgö, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., et al. (2011). Trait variation in yeast is defined by population history. *PLoS Genet.* 7:e1002111. doi: 10.1371/journal.pgen.1002111
- Wehner, E. P., Rao, E., and Brendel, M. (1993). Molecular structure and genetic regulation of SFA, a gene responsible for resistance to formaldehyde in *Saccharomyces cerevisiae*, and characterization of its protein product. *Mol. Gen. Genet.* 237, 351–358. doi: 10.1007/BF00279438
- Wightman, R., and Meacock, P. A. (2003). The TH15 gene family of *Saccharomyces cerevisiae*: distribution of homologues among the hemiascomycetes and functional redundancy in the aerobic biosynthesis of thiamin from pyridoxine. *Microbiology* 149, 1447–1460. doi: 10.1099/mic.0.26194-0
- Winzler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., et al. (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285, 901–906. doi: 10.1126/science.285.5429.901
- Xu, H., and Boeke, J. D. (1987). High-frequency deletion between homologous sequences during retrotransposition of Ty elements in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 84, 8553–8557. doi: 10.1073/pnas.84.23.8553
- Youngs, S., Ellis, K., Ennis, S., Barber, J., and Jacobs, P. (2004). A study of reciprocal translocations and inversions detected by light microscopy with special reference to origin, segregation, and recurrent abnormalities. *Am. J. Med. Genet. A* 126A, 46–60. doi: 10.1002/ajmg.a.20553
- Yu, J., Marshall, K., Yamaguchi, M., Haber, J. E., and Weil, C. F. (2004). Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 24, 1351–1364. doi: 10.1128/MCB.24.3.1351-1364.2004
- Yu, X., and Gabriel, A. (2003). Ku-dependent and Ku-independent end-joining pathways lead to chromosomal rearrangements during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* 163, 843–856.
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., et al. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* 49, 913–924. doi: 10.1038/ng.3847
- Zhang, F., Gu, W., Hurles, M. E., and Lupski, J. R. (2009). Copy number variation in human health, disease, and evolution. *Annu. Rev. Genomics Hum. Genet.* 10, 451–481. doi: 10.1146/annurev.genom.9.081307.164217

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Steenwyk and Rokas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.