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 37235, USA

* Correspondence: antonis.rokas@vanderbilt.edu

Keywords

Structural variation, alcohol fermentation, sugar metabolism, gene duplication, gene loss, population genomics

Abbreviations

BIR: break-induced recombination

CN: copy number

HR: homologous recombination

NHR: non-homologous repair

SNPs: single nucleotide polymorphisms

Abstract

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 associated with the adaptation and domestication of 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, 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.

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 not only in yeast, but in any kind of organism (McNally et al., 2009; Sachidanandam et al., 2001; 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 copy number (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 (Gu et al., 2016; Rausch et al., 2012; Youings et al., 2004). In contrast, many CN

64 variants are submicroscopic and eschewed attention until the advent of whole genome
65 sequencing technologies (Feuk et al., 2006).
66
67 CN variants are defined as duplications or deletions that range from 50 base pairs to whole
68 chromosomes (Figure 1) and can significantly influence phenotypic diversity (Arlt et al., 2014;
69 Zhang et al., 2009). For example, in humans, the CN of the salivary amylase gene, *AMY1*, is
70 higher in populations with high-starch diets and correlated with salivary protein abundance
71 thereby improving digestion of starchy foods (Perry et al., 2007). Levels of CN variation have
72 been examined in diverse organisms across the tree of life, including animals (e.g., Humans;
73 *Homo sapiens*: Sudmant et al., 2015, House mouse; *Mus musculus*: Pezer et al., 2015), plants
74 (e.g., soybean; *Glycine max*: Cook et al., 2012, maize; *Zea mays*: Swanson-Wagner et al., 2010)
75 and fungi (e.g., *Cryptococcus neoformans*: Hu et al., 2011, *Batrachochytrium dendrobatidis*:
76 Farrer et al., 2013, *Zymoseptoria tritici:* Hartmann and Croll, 2017).
77
78 *S. cerevisiae* has been an important model for genetics, genomics and evolution (Botstein et al.,
79 1997; Goffeau et al., 1996; Winzeler et al., 1999). Much of what we know about the evolutionary
80 history of *S. cerevisiae* stems from investigating genome-wide patterns of SNPs among globally
81 distributed strains. Examination of genome-wide patterns of SNP variation has yielded valuable
82 insights into yeast function in the wine fermentation environment. For example, 13 SNPs in
83 *ABZ1*, a gene associated with nitrogen biosynthetic pathways, have been shown to modify the
84 rate of fermentation and nitrogen utilization during fermentation (Ambroset et al., 2011).
85

Interrogations of genome-wide patterns of SNPs have also shown that industrial clades – including those of beer, bread, cacao, sake, and wine – often mirror human history (Cromie et al., 2013; Gallone et al., 2016; Gonçalves et al., 2016; Schacherer et al., 2009; Sicard and Legras, 2011), 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 (Borneman et al., 2016; Cromie et al., 2013; Liti et al., 2009; Schacherer et al., 2009; Sicard and Legras, 2011), consistent with a historical population bottle-neck 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 clade (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 (CN) 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 (Arlt et al., 2014; Sener, 2014; Zhang et al., 2009). 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).

CN variants are commonly generated from aberrant DNA repair via three mechanisms: homologous recombination (HR), non-homologous repair (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 (Aylon and Kupiec, 2004; Hastings et al., 2009b; Khanna and Jackson, 2001). Improper repair by HR can result in duplication, deletion, or inversion of

128 genetic material (Reams and Roth, 2015). Non-allelic HR (also known as ectopic
129 recombination), defined as recombination between two different loci of the same or different
130 chromosomes that share sequence similarity and are ≥300 base pairs in length, is among the most
131 well-studied examples of improper repair (Kupiec and Petes, 1988; Prado et al., 2003). Most
132 evidence of non-allelic HR resulting in CN variation is directly associated with low copy repeats
133 or transposable elements (Hurles, 2005; Xu and Boeke, 1987). For example, a duplication and
134 deletion may result during unequal crossing over of homologous sequences (Figure 2a)
135 (Carvalho and Lupski, 2016). Improper HR may also occur at collapsed or broken replication
136 forks by break-induced replication (BIR) (Figure 2b). BIR requires 3' strand invasion at the
137 allelic site of stalled replication to properly restart DNA synthesis (Figure 2bi) (Llorente et al.,
138 2008), however, template switching, the non-allelic pairing of homologous sequences, in the
139 backward (Figure 2bii) or forward (Figure 2biii) direction can result in a duplication or deletion,
140 respectively (Morrow et al., 1997; Smith et al., 2007). Although HR occurs with high fidelity,
141 errors in the process, which are thought to increase in frequency during mitosis and meiosis, can
142 generate CN variants (Hastings et al., 2009b).
143
144 In contrast to HR, NHR utilizes microhomologies (typically defined as ~65% or more sequence
145 similarity of short sequences up to ten bases long) or does not require homology altogether, and
146 can too lead to CN variant formation (Daley et al., 2005; McVey and Lee, 2008). NHR can occur
147 by two mechanisms: non-replicative and replicative (Hastings et al., 2009b). Non-replicative
148 mechanisms include non-homologous end joining and microhomology-mediated end-joining
149 (Lieber, 2008; McVey and Lee, 2008). Non-homologous end-joining refers to the direct ligation
150 of sequences in a double-strand break (Daley et al., 2005). Prior to ligation, there may be a loss

151 of genetic material or the addition of free DNA (e.g., from transposable elements or
152 mitochondrial DNA) (Yu and Gabriel, 2003). Microhomology-mediated end joining is similar to
153 non-homologous end-joining but occurs more frequently, requires different enzymes, and
154 leverages homologies 1-10 base pairs in length to ensure more efficient annealing (Lieber, 2008;
155 Yu et al., 2004). Non-homologous end-joining and microhomology-mediated non-homologous
156 end-joining are primarily associated with small insertions and deletions and therefore are not
157 likely to be a major driver of CN variation (Gu et al., 2008; Yu and Gabriel, 2003). Replicative
158 mechanisms of CN variant formation include replication slippage, fork stalling, and
159 microhomology BIR. Replication slippage occurs along repetitive stretches of DNA resulting in
160 the duplication or deletion of sequence between repetitive regions (Hastings et al., 2009b). Fork
161 stalling is thought to cause large CNVs of 20 kb average length through template switching
162 between distal replication forks rather than within a replication fork (Slack et al., 2006).
163 However, fork stalling without distal template switching can also be highly mutagenic and
164 induce CN variants (Hull et al., 2017; Paul et al., 2013). Lastly, microhomology-mediated break-
165 induced replication occurs when the 3' end of a collapsed fork anneals with any single-stranded
166 template that it shares microhomology with to reinitiate DNA synthesis (Figure 2b) (Hastings et
167 al., 2009b). Annealing can occur in the backward (Figure 2bii) or forward (Figure 2biii) direction
168 of the allelic site causing a duplication or deletion, respectively, and is thought to be the primary
169 cause of low copy repeats (Hastings et al., 2009a).
170
171 The third mechanism is associated with an epigenetic mark that can stimulate the formation of
172 CN variants. Histone acetylation, specifically H3K56ac, is, in part, environmentally driven
173 (Turner, 2009), associated with highly transcribed loci, and can promote CN variant formation

174 through repeated fork stalling or template switching (Figure 2c) (Hull et al., 2017). For example,
175 it has been shown that exposure to environmental copper stimulates the generation of CN
176 variation in *CUP1,* a gene that is associated with copper resistance when duplicated (Fogel and
177 Welch, 1982), thereby increasing the likelihood of favorable alleles that exhibit increased copper
178 resistance (Hull et al., 2017). Similarly, environmental formaldehyde exposure was shown to
179 stimulate CN variation (Hull et al., 2017) of the *SFA1* gene, which confers formaldehyde
180 resistance at higher CNs (Wehner et al., 1993). Altogether, these experiments provide insight to
181 how perturbations of an environmental parameter may stimulate CN variation at a locus
182 important to adaptation in the new environment (Hull et al., 2017).
183

184 Copy number variation as a source of phenotypic diversity

185 CN variants can have multiple effects on gene activity, such as changing gene dosage (i.e., gene
186 CN; Figure 3) and interrupting coding sequences (Itsara et al., 2009; Sener, 2014). These effects
187 can be substantial; for example, 17.7% of gene expression variation in human populations can be
188 attributed to CN variants (Stranger et al., 2007). Furthermore, changes in human gene expression
189 attributed to CN variants have little overlap with changes in gene expression caused by SNPs,
190 suggesting the two types of variation independently affect gene expression (Stranger et al.,
191 2007). Additionally, gene CN tends to correlate with levels of both gene expression and protein
192 abundance (Henrichsen et al., 2009; Perry et al., 2007; Stranger et al., 2007). For example,
193 changes in gene expression and therefore protein abundance caused by chromosomal CN
194 variation in human chromosome 21 are thought to contribute to Down syndrome (Aivazidis et
195 al., 2017; Kahlem et al., 2004).

196

197 Copy number variation as a source of genetic and phenotypic
198 diversity in fungal populations

199 CN variant loci contribute to population genetic and phenotypic diversity (Box 1), such as
200 virulence (Farrer et al., 2013; Hu et al., 2011b), in diverse fungal species, including as the
201 baker's yeast *Saccharomyces cerevisiae* (ASCOMYCOTA, Saccharomycetes) (Gallone et al.,
202 2016; Gonçalves et al., 2016; Steenwyk and Rokas, 2017), the fission yeast
203 *Schizosaccharomyces pombe* (ASCOMYCOTA, Schizosaccharomycetes) (Jeffares et al., 2017),
204 the human fungal pathogen *Cryptococcus deuterogattii* (BASIDIOMYCOTA, Tremellomycetes)
205 (previously known as *Cryptococcus gattii* VGII; Steenwyk et al., 2016) and *C. neoformans* (Hu
206 et al., 2011b), the amphibian pathogen *Batrachochytrium dendrobatidis*
207 (CHYTRIDIOMYCOTA, Chytridiomycetes) (Farrer et al., 2013), and the wheat pathogen
208 *Zymoseptoria tritici* (ASCOMYCOTA, Dothideomycetes) (Hartmann and Croll, 2017).

209

210 Importantly, the degree of CN variation (which can be represented by CN variable base pairs per
211 kilobase) in fungal populations is not always correlated to the degree of SNP variation (which
212 can be represented by SNPs per kilobase) (Figure 4a). For example, there is no correlation
213 between CN variable base pairs per kilobase and SNPs per kilobase among *S. cerevisiae* wine
214 strains (Steenwyk and Rokas, 2017) and a population of *Cryptococcus deuterogattii* (Steenwyk
215 et al., 2016). Interestingly, both populations harbor low levels of SNP diversity; for *S. cerevisiae*
216 wine strains this is due to a single domestication-associated bottleneck event (Cromie et al.,
217 2013; Liti et al., 2009; Schacherer et al., 2009; Sicard and Legras, 2011), 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*, followed by *C. deuterogattii,* and then *S. pombe*. In contrast, the *S. cerevisiae* population has fewer SNPs per kilobase compared to *S. pombe* but more SNPs per kilobase compared to *C. deuterogattii* (Figure 4b).

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).

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. 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* (Gallone et al., 2016; Gonçalves et al., 2016; Ibáñez et al., 2014; 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). 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 (Almeida et al., 2015; Fay et al., 2004).

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 (Fidalgo et al., 2006; Martínez et al., 1997). 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). Some of this variation may be adaptive. For example, 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 among "flor" or "sherry" yeasts (Fidalgo et al., 2006). Furthermore, the same partial duplications have also been observed in the more general wine population (Steenwyk and Rokas, 2017), suggesting that the benefits associated with this phenotype may not be unique to "flor" yeasts.

287 CN variation is also observed in genes related to stuck (incomplete) or sluggish (delayed)
288 fermentations. Stuck fermentations are caused by a multitude of factors including nitrogen
289 availability, nutrient transport, and decreased resistance to starvation (Salmon, 1989; Thomsson
290 et al., 2005). Two genes associated with decrease resistance to starvation, *ADH7* and *AAD3*, are
291 sometimes duplicated or deleted among wine yeast strains (Steenwyk and Rokas, 2017). Diverse
292 CN profiles of *ADH7*, an alcohol dehydrogenase that reduces acetaldehyde to ethanol during
293 glucose fermentation, and *AAD3,* an aryl-alcohol dehydrogenase whose null mutant displays
294 greater starvation sensitivity (Walker et al., 2014), suggest variable degrees of starvation
295 sensitivity and therefore fermentation performance. Additionally, wine yeasts are enriched for
296 duplication in *PDR18* (Gallone et al., 2016), a transporter that aids in resistance to ethanol stress,
297 one of the traits that differentiates wine from other industrial strains. Another gene associated
298 with decreased resistance to starvation that also exhibits CN variation is *IMA1* (Steenwyk and
299 Rokas, 2017), a major isomaltase with glucosidase activity (Teste et al., 2010).
300
301 CN variable genes related to metabolism
302 Nutrient availability and acquisition is a major driving factor of wine fermentation outcome.
303 Among the most important nutrients dictating the pace and success of wine fermentation is sugar
304 availability (Marsit and Dequin, 2015). The most abundant fermentable hexose sugars in the
305 wine environment include glucose and fructose (Marques et al., 2015), whose transport is largely
306 carried out by genes from the hexose transporter (*HXT*) family (Boles and Hollenberg, 1997). A
307 reproducible evolutionary outcome of yeasts exposed to glucose-limited environments, which are
308 reflective of late wine fermentation, is duplication in hexose transporters, such as *HXT6* and
309 *HXT7* (Brown et al., 1998; Dunham et al., 2002; Gresham et al., 2008, 2010), suggesting that

changes in transporter CN are adaptive. Interestingly, the genes from the *HXT* gene family are highly CN variable among wine yeast strains (Dunn et al., 2012; Steenwyk and Rokas, 2017). For example, *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) and therefore 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, the *MAL1* locus is deleted across many wine yeasts (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 (M.Victoria and M. 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).

CN variation is also observed in genes of the *THI* family, which are involved in thiamine, or vitamin B_1, metabolism (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).

Conclusions and perspectives

An emerging body of work suggests that CN variation is an important, largely underappreciated, dimension of fungal genome biology and evolution (Farrer et al., 2013; Gallone et al., 2016; Gonçalves et al., 2016; Hartmann and Croll, 2017; Hu et al., 2011a; Steenwyk et al., 2016; 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

355 copies typically found in the same genomic neighborhood or are they dispersed? Similarly, what
356 percentage of phenotypic differences among fungal strains is explained by CN variation?
357
358 The same can be said about the role of CN variation in yeast adaptation to the wine fermentation
359 environment. Comparison of genome-wide patterns of CN variation among yeast populations
360 responsible for the fermentation of different wines (e.g., white and red) would provide insight to
361 how human activity has shaped the genome of yeasts associated with particular types of wine.
362 Additionally, most sequenced wine strains originate from Italy, Australia, or France. Genome
363 sequencing of yeasts from underrepresented regions (e.g., Africa and the Americas) may provide
364 further insight to CN variable loci unique to each region and the global diversity of wine yeast
365 genomes.
366
367 Another major set of questions are associated with examining the impact of CN variable loci at
368 the different stages of wine fermentation. Insights on how CN variable loci modify gene
369 expression, protein abundance and in turn fermentation behavior and end-product would be
370 immensely valuable. A complementary, perhaps more straightforward, approach would be
371 focused on examining the phenotypic impact of single-gene or gene family CN variants, such as
372 the ones discussed in previous sections (e.g. genes belonging to the *ADH*, *HXT*, *MAL*, and *VBA*
373 families; Table 1) on fermentation outcome. Such studies may provide an important bridge
374 between scientist, oenologist, and wine-maker to enhance fermentation efficiency and
375 consistency between batches or in the design of new wine flavor profiles.
376

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

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 (Bethesda).* 1, 263–81. 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, 552–64. 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.

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.

411 Borneman, A. R., Forgan, A. H., Kolouchova, R., Fraser, J. A., and Schmidt, S. A. (2016).
412 Whole Genome Comparison Reveals High Levels of Inbreeding and Strain Redundancy
413 Across the Spectrum of Commercial Wine Strains of Saccharomyces cerevisiae. *G3*
414 *Genes|Genomes|Genetics*, g3.115.025692. doi:10.1534/g3.115.025692.
415 Botstein, D., Chervitz, S. a., and Cherry, J. M. (1997). Yeast as a Model Organism. *Science (80-.*
416 *).* 227, 1259–1260. doi:10.1126/science.277.5330.1259.
417 Brauer, M. J., Christianson, C. M., Pai, D. A., and Dunham, M. J. (2006). Mapping novel traits
418 by array-assisted bulk segregant analysis in Saccharomyces cerevisiae. *Genetics* 173, 1813–
419 1816. doi:10.1534/genetics.106.057927.
420 Brion, C., Ambroset, C., Delobel, P., Sanchez, I., and Blondin, B. (2014). Deciphering regulatory
421 variation of THI genes in alcoholic fermentation indicate an impact of Thi3p on PDC1
422 expression. *BMC Genomics* 15, 1085. doi:10.1186/1471-2164-15-1085.
423 Brown, C. A., Murray, A. W., and Verstrepen, K. J. (2010). Rapid Expansion and Functional
424 Divergence of Subtelomeric Gene Families in Yeasts. *Curr. Biol.* 20, 895–903.
425 doi:10.1016/j.cub.2010.04.027.
426 Brown, C. J., Todd, K. M., and Rosenzweig, R. F. (1998). Multiple duplications of yeast hexose
427 transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.*
428 15, 931–42. doi:10.1093/oxfordjournals.molbev.a026009.
429 Carvalho, C. M. B., and Lupski, J. R. (2016). Mechanisms underlying structural variant
430 formation in genomic disorders. *Nat. Rev. Genet.* 17, 224–238. doi:10.1038/nrg.2015.25.
431 Chakravarti, A. (1999). Population genetics - Making sense out of sequence. *Nat. Genet.* 21, 56–
432 60. doi:10.1038/4482.
433 Chikashige, Y., Tsutsumi, C., Okamasa, 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, 149–61. 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 (80-.).* 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 (Bethesda).* 3, 2163–71. doi:10.1534/g3.113.007492.

Daley, J. M., Palmbos, 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., Kvitek, 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.

457 doi:10.1128/mBio.01464-14.

458 Estivill, X., and Armengol, L. (2007). Copy number variants and common disorders: Filling the
459 gaps and exploring complexity in genome-wide association studies. *PLoS Genet.* 3, 1787–
460 1799. doi:10.1371/journal.pgen.0030190.

461 Farrer, R. A., Henk, D. A., Garner, T. W., Balloux, F., Woodhams, D. C., and Fisher, M. C.
462 (2013). Chromosomal copy number variation, selection and uneven rates of recombination
463 reveal cryptic genome diversity linked to pathogenicity. *PLoS Genet* 9, e1003703.
464 doi:10.1371/journal.pgen.1003703.

465 Fay, J. C., McCullough, H. L., Sniegowski, P. D., and Eisen, M. B. (2004). Population genetic
466 variation in gene expression is associated with phenotypic variation in Saccharomyces
467 cerevisiae. *Genome Biol.* 5, R26. doi:10.1186/gb-2004-5-4-r26.

468 Felsenstein, J. (1976). The Theoretical Population Genetics of Variable Selection and Migration.
469 *Annu. Rev. Genet.* 10, 253–280. doi:10.1146/annurev.ge.10.120176.001345.

470 Feuk, L., Carson, A. R., and Scherer, S. W. (2006). Structural variation in the human genome.
471 *Nat Rev Genet* 7, 85–97. doi:10.1038/nrg1767.

472 Fidalgo, M., Barrales, R. R., Ibeas, J. I., and Jimenez, J. (2006). Adaptive evolution by mutations
473 in the FLO11 gene. *Proc. Natl. Acad. Sci. U. S. A.* 103, 11228–33.
474 doi:10.1073/pnas.0601713103.

475 Fogel, S., and Welch, J. W. (1982). Tandem gene amplification mediates copper resistance in
476 yeast. *Proc. Natl. Acad. Sci. USA* 79, 5342–5346. doi:10.1073/pnas.79.17.5342.

477 Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., et al. (2016).
478 Domestication and Divergence of Saccharomyces cerevisiae Beer Yeasts. *Cell* 166, 1397–
479 1410.e16. doi:10.1016/j.cell.2016.08.020.

480 Gibbons, J. G., and Rinker, D. C. (2015). The genomics of microbial domestication in the
481 fermented food environment. *Curr. Opin. Genet. Dev.* 35, 1–8.
482 doi:10.1016/j.gde.2015.07.003.
483 Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., et al. (1996).
484 Life with 6000 Genes. *Science (80-.).* 274, 546–567. doi:10.1126/science.274.5287.546.
485 Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., et al. (2016).
486 Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts. *Curr.*
487 *Biol.* 26, 2750–2761. doi:http://dx.doi.org/10.1016/j.cub.2016.08.040.
488 Govender, P., Domingo, J. L., Bester, M. C., Pretorius, I. S., and Bauer, F. F. (2008). Controlled
489 expression of the dominant flocculation genes FLO1, FLO5, and FLO11 in Saccharomyces
490 cerevisiae. *Appl. Environ. Microbiol.* 74, 6041–6052. doi:10.1128/AEM.00394-08.
491 Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A., et al. (2008). The
492 repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited
493 environments in yeast. *PLoS Genet.* 4. doi:10.1371/journal.pgen.1000303.
494 Gresham, D., Usaite, R., Germann, S. M., Lisby, M., Botstein, D., and Regenberg, B. (2010).
495 Adaptation to diverse nitrogen-limited environments by deletion or extrachromosomal
496 element formation of the GAP1 locus. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18551–18556.
497 doi:10.1073/pnas.1014023107.
498 Gu, S., Szafranski, P., Akdemir, Z. C., Yuan, B., Cooper, M. L., Magriñá, M. A., et al. (2016).
499 Mechanisms for Complex Chromosomal Insertions. *PLoS Genet.* 12.
500 doi:10.1371/journal.pgen.1006446.
501 Gu, W., Zhang, F., and Lupski, J. R. (2008). Mechanisms for human genomic rearrangements.
502 *Pathogenetics* 1, 4. doi:10.1186/1755-8417-1-4.

503 Hartmann, F. E., and Croll, D. (2017). Distinct Trajectories of Massive Recent Gene Gains and
504 Losses in Populations of a Microbial Eukaryotic Pathogen. *Mol. Biol. Evol.*
505 doi:10.1093/molbev/msx208.
506 Hastings, P. J., Ira, G., and Lupski, J. R. (2009a). A microhomology-mediated break-induced
507 replication model for the origin of human copy number variation. *PLoS Genet.* 5.
508 doi:10.1371/journal.pgen.1000327.
509 Hastings, P. J., Lupski, J. R., Rosenberg, S. M., and Ira, G. (2009b). Mechanisms of change in
510 gene copy number. *Nat. Rev. Genet.* 10, 551–564. doi:10.1038/nrg2593.
511 Henrichsen, C. N., Vinckenbosch, N., Zöllner, S., Chaignat, E., Pradervand, S., Schütz, F., et al.
512 (2009). Segmental copy number variation shapes tissue transcriptomes. *Nat. Genet.* 41,
513 424–429. doi:10.1038/ng.345.
514 Hu, G., Wang, J., Choi, J., Jung, W. H., Liu, I., Litvintseva, A. P., et al. (2011a). Variation in
515 chromosome copy number influences the virulence of Cryptococcus neoformans and occurs
516 in isolates from AIDS patients. *BMC Genomics* 12, 526. doi:10.1186/1471-2164-12-526.
517 Hu, G., Wang, J., Choi, J., Jung, W. H., Liu, I., Litvintseva, A. P., et al. (2011b). Variation in
518 chromosome copy number influences the virulence of Cryptococcus neoformans and occurs
519 in isolates from AIDS patients. *BMC Genomics* 12, 526. doi:10.1186/1471-2164-12-526.
520 Hua, S., Qiu, M., Chan, E., Zhu, L., and Luo, Y. (1997). Minimum Length of Sequence
521 Homology Required forin VivoCloning by Homologous Recombination in Yeast. *Plasmid*
522 38, 91–96. doi:http://dx.doi.org/10.1006/plas.1997.1305.
523 Hull, R. M., Cruz, C., Jack, C. V., Houseley, J., Noronha, M., and Calderon, L. (2017).
524 Environmental change drives accelerated adaptation through stimulated copy number
525 variation. *PLOS Biol.* 15, e2001333. doi:10.1371/journal.pbio.2001333.

526 Hurles, M. (2005). How homologous recombination generates a mutable genome. *Hum.*
527 *Genomics* 2, 179. doi:10.1186/1479-7364-2-3-179.
528 Ibáñez, C., Pérez-Torrado, R., Chiva, R., Guillamón, J. M., Barrio, E., and Querol, A. (2014).
529 Comparative genomic analysis of Saccharomyces cerevisiae yeasts isolated from
530 fermentations of traditional beverages unveils different adaptive strategies. *Int. J. Food*
531 *Microbiol.* 171, 129–135. doi:10.1016/j.ijfoodmicro.2013.10.023.
532 Itsara, A., Cooper, G. M., Baker, C., Girirajan, S., Li, J., Absher, D., et al. (2009). Population
533 Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease. *Am. J.*
534 *Hum. Genet.* 84, 148–161. doi:10.1016/j.ajhg.2008.12.014.
535 Jeffares, D. C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., et al. (2017). Transient
536 structural variations have strong effects on quantitative traits and reproductive isolation in
537 fission yeast. *Nat. Commun.* 8, 14061. doi:10.1038/ncomms14061.
538 Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Převorovský, M., Mourier, T., et al. (2015). The
539 genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nat. Genet.* 47, 235–41.
540 doi:10.1038/ng.3215.
541 Kahlem, P., Sultan, M., Herwig, R., Steinfath, M., Balzereit, D., Eppens, B., et al. (2004).
542 Transcript level alterations reflect gene dosage effects across multiple tissues in a mouse
543 model of Down syndrome. *Genome Res.* 14, 1258–1267. doi:10.1101/gr.1951304.
544 Khanna, K. K., and Jackson, S. P. (2001). DNA double-strand breaks: signaling, repair and the
545 cancer connection. *Nat. Genet.* 27, 247–54. doi:10.1038/85798.
546 Kupiec, M., and Petes, T. D. (1988). Allelic and ectopic recombination between Ty elements in
547 yeast. *Genetics* 119, 549–559.
548 Lemos Junior, W. J. F., Treu, L., da Silva Duarte, V., Carlot, M., Nadai, C., Campanaro, S., et al.

549 (2017). Whole-Genome Sequence of Starmerella bacillaris PAS13, a Nonconventional
550 Enological Yeast with Antifungal Activity. *Genome Announc.* 5, e00788-17.
551 doi:10.1128/genomeA.00788-17.
552 Li, M., Petteys, B. J., McClure, J. M., Valsakumar, V., Bekiranov, S., Frank, E. L., et al. (2010).
553 Thiamine biosynthesis in Saccharomyces cerevisiae is regulated by the NAD+-dependent
554 histone deacetylase Hst1. *Mol. Cell. Biol.* 30, 3329–41. doi:10.1128/MCB.01590-09.
555 Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. *J. Biol.*
556 *Chem.* 283, 1–5. doi:10.1074/jbc.R700039200.
557 Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., et al. (2009).
558 Population genomics of domestic and wild yeasts. *Nature* 458, 337–41.
559 doi:10.1038/nature07743.
560 Llorente, B., Smith, C. E., and Symington, L. S. (2008). Break-induced replication: What is it
561 and what is it for? *Cell Cycle* 7, 859–864. doi:10.4161/cc.7.7.5613.
562 M.Victoria, M.-A., and M. Carmen, P. (2013). Wine Chemistry and Biochemistry. *J. Chem. Inf.*
563 *Model.* 53, 1689–1699. doi:10.1017/CBO9781107415324.004.
564 Marques, W. L., Raghavendran, V., Stambuk, B. U., and Gombert, A. K. (2015). Sucrose and
565 Saccharomyces cerevisiae: A relationship most sweet. *FEMS Yeast Res.* 16.
566 doi:10.1093/femsyr/fov107.
567 Marsit, S., and Dequin, S. (2015). Diversity and adaptive evolution of Saccharomyces wine
568 yeast: a review. *FEMS Yeast Res.*, 1–12. doi:10.1093/femsyr/fov067.
569 Martínez, P., Pérez Rodríguez, L., and Benítez, T. (1997). Velum Formation by Flor Yeasts
570 Isolated From Sherry Wine. *Am. J. Enol. Vitic.* 48.
571 McNally, K. L., Childs, K. L., Bohnert, R., Davidson, R. M., Zhao, K., Ulat, V. J., et al. (2009).

572 Genomewide SNP variation reveals relationships among landraces and modern varieties of
573 rice. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12273–8. doi:10.1073/pnas.0900992106.
574 McVey, M., and Lee, S. E. (2008). MMEJ repair of double-strand breaks (director's cut): deleted
575 sequences and alternative endings. *Trends Genet.* 24, 529–538.
576 doi:10.1016/j.tig.2008.08.007.
577 Michels, C. A., Read, E., Nat, K., and Charron, M. J. (1992). The telomere-associated MAL3
578 locus of Saccharomyces is a tandem array of repeated genes. *Yeast* 8, 655–665. Available
579 at:
580 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation
581 &list_uids=1441745.
582 Moritz, C. (1994). Defining "Evolutionarily Significant Units" for conservation. *Trends Ecol.*
583 *Evol.* 9, 373–375. doi:10.1016/0169-5347(94)90057-4.
584 Morrow, D. M., Connelly, C., and Hieter, P. (1997). "Break copy" duplication: A model for
585 chromosome fragment formation in saccharomyces cerevisiae. *Genetics* 147, 371–382.
586 Naumov, G. I., Naumova, E. S., and Michels, C. A. (1994). Genetic variation of the repeated
587 MAL loci in natural populations of Saccharomyces cerevisiae and Saccharomyces
588 paradoxus. *Genetics* 136, 803–12. Available at:
589 http://www.ncbi.nlm.nih.gov/pubmed/8005435 [Accessed September 19, 2016].
590 Paul, S., Million-Weaver, S., Chattopadhyay, S., Sokurenko, E., and Merrikh, H. (2013).
591 Accelerated gene evolution through replication–transcription conflicts. *Nature* 495, 512–
592 515. doi:10.1038/nature11989.
593 Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., et al. (2007). Diet
594 and the evolution of human amylase gene copy number variation. *Nat Genet* 39, 1256–

595 1260. doi:10.1038/ng2123.

596 Petukhova, G., Stratton, S., and Sung, P. (1998). Catalysis of homologous DNA pairing by yeast
597 Rad51 and Rad54 proteins. *Nature* 393, 91–94. doi:10.1038/30037.

598 Pezer, Z., Harr, B., Teschke, M., Babiker, H., and Tautz, D. (2015). Divergence patterns of genic
599 copy number variation in natural populations of the house mouse (Mus musculus
600 domesticus) reveal three conserved genes with major population-specific expansions.
601 *Genome Res* 25, 1114–1124. doi:10.1101/gr.187187.114.

602 Prado, F., Cortés-Ledesma, F., Huertas, P., and Aguilera, A. (2003). Mitotic recombination in
603 Saccharomyces cerevisiae. *Curr. Genet.* 42, 185–198. doi:10.1007/s00294-002-0346-3.

604 Pretorius, I. S. (2000). Tailoring wine yeast for the new millennium: Novel approaches to the
605 ancient art of winemaking. *Yeast* 16, 675–729. doi:10.1002/1097-
606 0061(20000615)16:8<675::AID-YEA585>3.0.CO;2-B.

607 Rahim, N. G., Harismendy, O., Topol, E. J., and Frazer, K. A. (2008). Genetic determinants of
608 phenotypic diversity in humans. *Genome Biol.* 9, 215. doi:10.1186/gb-2008-9-4-215.

609 Rausch, T., Jones, D. T. W., Zapatka, M., Stütz, A. M., Zichner, T., Weischenfeldt, J., et al.
610 (2012). Genome sequencing of pediatric medulloblastoma links catastrophic DNA
611 rearrangements with TP53 mutations. *Cell* 148, 59–71. doi:10.1016/j.cell.2011.12.013.

612 Reams, A. B., and Roth, J. R. (2015). Mechanisms of gene duplication and amplification. *Cold*
613 *Spring Harb. Perspect. Biol.* 7. doi:10.1101/cshperspect.a016592.

614 Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al.
615 (2001). A map of human genome sequence variation containing 1.42 million single
616 nucleotide polymorphisms. *Nature* 409, 928–933. doi:10.1038/35057149.

617 Salmon, J. M. (1989). Effect of sugar transport inactivation in Saccharomyces cerevisiae on

618 sluggish and stuck enological fermentations. *Appl. Environ. Microbiol.* 55, 953–958.

619 Sarilar, V., Devillers, H., Freel, K. C., Schacherer, J., and Neuvéglise, C. (2015). Draft Genome
620 Sequence of Lachancea lanzarotensis CBS 12615 T , an Ascomycetous Yeast Isolated from
621 Grapes. *Genome Announc.* 3, e00292-15. doi:10.1128/genomeA.00292-15.

622 Schacherer, J., Shapiro, J. a, Ruderfer, D. M., and Kruglyak, L. (2009). Comprehensive
623 polymorphism survey elucidates population structure of Saccharomyces cerevisiae. *Nature*
624 458, 342–5. doi:10.1038/nature07670.

625 Seixas, I., Barbosa, C., Salazar, S. B., Mendes-Faia, A., Wang, Y., Güldener, U., et al. (2017).
626 Genome Sequence of the Nonconventional Wine Yeast Hanseniaspora guilliermondii
627 UTAD222. *Genome Announc.* 5, e01515-16. doi:10.1128/genomeA.01515-16.

628 Sener, E. F. (2014). Association of Copy Number Variations in Autism Spectrum Disorders: A
629 Systematic Review. *Chinese J. Biol.* 2014, 1–9. doi:10.1155/2014/713109.

630 Sicard, D., and Legras, J. L. (2011). Bread, beer and wine: Yeast domestication in the
631 Saccharomyces sensu stricto complex. *Comptes Rendus - Biol.* 334, 229–236.
632 doi:10.1016/j.crvi.2010.12.016.

633 Slack, A., Thornton, P. C., Magner, D. B., Rosenberg, S. M., and Hastings, P. J. (2006). On the
634 mechanism of gene amplification induced under stress in Escherichia coli. *PLoS Genet.* 2,
635 385–398. doi:10.1371/journal.pgen.0020048.

636 Smith, C. E., Llorente, B., and Symington, L. S. (2007). Template switching during break-
637 induced replication. *Nature* 447, 102–105. doi:10.1038/nature05723.

638 Soares, E. V. (2011). Flocculation in Saccharomyces cerevisiae: A review. *J. Appl. Microbiol.*
639 110, 1–18. doi:10.1111/j.1365-2672.2010.04897.x.

640 Steenwyk, J. L., Soghigian, J. S., Perfect, J. R., and Gibbons, J. G. (2016). Copy number

641 variation contributes to cryptic genetic variation in outbreak lineages of Cryptococcus gattii
642 from the North American Pacific Northwest. *BMC Genomics* 17, 700. doi:10.1186/s12864-
643 016-3044-0.
644 Steenwyk, J., and Rokas, A. (2017). Extensive Copy Number Variation in Fermentation-Related
645 Genes Among Saccharomyces cerevisiae Wine Strains. *G3 Genes, Genomes, Genet.* 7.
646 Available at: http://www.g3journal.org/content/7/5/1475#ref-25 [Accessed July 3, 2017].
647 Sternes, P. R., Lee, D., Kutyna, D. R., and Borneman, A. R. (2016). Genome Sequences of Three
648 Species of Hanseniaspora Isolated from Spontaneous Wine Fermentations. *Genome*
649 *Announc.* 4, e01287-16. doi:10.1128/genomeA.01287-16.
650 Stranger, B. E., Forrest, M. S., Dunning, M., Ingle, C. E., Beazley, C., Thorne, N., et al. (2007).
651 Relative impact of nucleotide and copy number variation on gene expression phenotypes.
652 *Science* 315, 848–853. doi:10.1126/science.1136678.
653 Sudmant, P. H., Mallick, S., Nelson, B. J., Hormozdiari, F., Krumm, N., Huddleston, J., et al.
654 (2015). Global diversity, population stratification, and selection of human copy-number
655 variation. *Science (80-.).* 349, aab3761. doi:10.1126/science.aab3761.
656 Swanson-Wagner, R. A., Eichten, S. R., Kumari, S., Tiffin, P., Stein, J. C., Ware, D., et al.
657 (2010). Pervasive gene content variation and copy number variation in maize and its
658 undomesticated progenitor. *Genome Res.* 20, 1689–1699. doi:10.1101/gr.109165.110.
659 Taylor, M. B., Phan, J., Lee, J. T., McCadden, M., and Ehrenreich, I. M. (2016). Diverse genetic
660 architectures lead to the same cryptic phenotype in a yeast cross. *Nat. Commun.* 7.
661 doi:10.1038/ncomms11669.
662 Teste, M. A., Marie François, J., and Parrou, J. L. (2010). Characterization of a new multigene
663 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.

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, 351–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8483449.

Wightman, R., and Meacock, P. A. (2003). The THI5 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.

687 doi:10.1099/mic.0.26194-0.

688 Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B., et al.
689 (1999). Functional characterization of the S-cerevisiae genome by gene deletion and parallel
690 analysis. *Science (80-.).* 285, 901–906. doi:10.1126/science.285.5429.901.

691 Xu, H., and Boeke, J. D. (1987). High-frequency deletion between homologous sequences during
692 retrotransposition of Ty elements in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S.*
693 *A.* 84, 8553–8557. doi:10.1073/pnas.84.23.8553.

694 Youings, S., Ellis, K., Ennis, S., Barber, J., and Jacobs, P. (2004). A study of reciprocal
695 translocations and inversions detected by light microscopy with special reference to origin,
696 segregation, and recurrent abnormalities. *Am. J. Med. Genet. A* 126A, 46–60.
697 doi:10.1002/ajmg.a.20553.

698 Yu, J., Marshall, K., Yamaguchi, M., Haber, J. E., and Weil, C. F. (2004). Microhomology-
699 Dependent End Joining and Repair of Transposon-Induced DNA Hairpins by Host Factors
700 in Saccharomyces cerevisiae. *Mol. Cell. Biol.* 24, 1351–1364. doi:10.1128/MCB.24.3.1351-
701 1364.2004.

702 Yu, X., and Gabriel, A. (2003). Ku-dependent and Ku-independent end-joining pathways lead to
703 chromosomal rearrangements during double-strand break repair in Saccharomyces
704 cerevisiae. *Genetics* 163, 843–856.

705 Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., et al. (2017). Contrasting
706 evolutionary genome dynamics between domesticated and wild yeasts. *Nat. Genet.* 49, 913–
707 924. doi:10.1038/ng.3847.

708 Zhang, F., Gu, W., Hurles, M. E., and Lupski, J. R. (2009). Copy number variation in human
709 health, disease, and evolution. *Annu Rev Genomics Hum Genet* 10, 451–481.

710 doi:10.1146/annurev.genom.9.081307.164217.

711

712 **Table 1. Genes associated with fermentation-related processes that**

713 **exhibit CN variation among *S. cerevisiae* wine strains**

Process (organized alphabetically)	**Gene**	**Primarily duplicated, deleted, or both**	**Reference (organized alphabetically)**
Amino acid & nitrogen utilization	*VBA3, VBA5, PUT1*	Duplicated	Gallone *et al.* 2016; Ibáñez *et al.* 2013
Cu and Fe homeostasis	*CUP1, CUP2*	Both	Almeida *et al.* 2015; Fay *et al.* 2004; Steenwyk & Rokas, 2017; Warringer *et al.* 2011
	FIT2, FIT3, FRE3	Duplicated	Gallone *et al.* 2016
Ethanol resistance and production	*PDR18*		Gallone *et al.* 2016
	ADH7	Both	Steenwyk & Rokas, 2017
Flocculation	*FLO11*	Duplicated	Steenwyk & Rokas, 2017
	FLO1, FLO5, FLO9, FLO10	Both	Gallone *et al.* 2016; Steenwyk & Rokas, 2017
Hexose transport	*HXT1, HXT4, HXT6, HXT7, HXT16*	Duplicated	Gallone *et al.* 2016; Steenwyk & Rokas, 2017
	HXT9, HXT11	Deleted	Gallone *et al.* 2016; Steenwyk & Rokas, 2017
	HXT13, HXT15,	Both	Gallone *et al.* 2016;

	HXT17		Steenwyk & Rokas, 2017
Maltose metabolism	*MAL3x, MPH3, YPR196W*	Duplicated	Gallone *et al.* 2016; Gonçalves *et al.* 2016; Steenwyk & Rokas, 2017
	MAL1x, IMA2, IMA4, IMA5	Deleted	Gallone *et al.* 2016; Gonçalves *et al.* 2016; Steenwyk & Rokas, 2017
	MPH2, IMA1, IMA3	Both	Gallone *et al.* 2016; Steenwyk & Rokas, 2017
Thiamine metabolism	*THI13*	Duplicated	Steenwyk & Rokas, 2017
	THI5, THI12	Deleted	Steenwyk & Rokas, 2017

714

Figure legends

Figure 1. The different types of CN variation. CN variants range in size (50 base pairs or greater) to whole chromosomes, and are identified through comparison to a reference genome. In this cartoon, a reference chromosome containing two highlighted loci, in blue and orange, is shown on top. The second chromosome illustrates an example of a segmental duplication CN, in which there are two copies of the blue locus. The third chromosome illustrates an example of a multiallelic CN variant, where the duplicated locus contains 3 or more copies. The fourth pair of chromosomes illustrates a CN variant associated with the duplication of an entire chromosome. Finally, the last two chromosomes illustrate deletion and complex CN variants, respectively; deletion CN variants are associated with loci that are not present relative to the reference, and complex CN variants refer to a combination of duplications, deletions, insertions, and/or inversions relative to the reference.

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 and 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). (bi) Proper BIR starts with strand invasion of a homologous or microhomologous sequence (shown in red) to allow for proper fork restart. (bii) If template

738 switching occurs in the backward direction, a segment of DNA will have been replicated twice
739 resulting in a duplication; (biii) in contrast, template switching in the forward direction results in
740 a deletion represented by a dashed line in the DNA sequence. Erroneous BIR may be mediated
741 by microhomologies as well. (c) CN variants may be stimulated near genes that are highly
742 expressed due to an increased chance of fork stalling. (ci) If a replication fork breaks down near
743 a gene that is not expressed (grey) and restarts once (represented by one black arrow), no
744 mutation will occur. (cii) If a replication fork breaks down near a gene that is expressed (green)
745 with cryptic unstable transcripts (red) then there may be two outcomes dependent on the degree
746 of the H3K56ac acetylation mark. If there are low levels of H3K56ac, it is more likely that there
747 will be proper fork restart by BIR (represented by one black arrow). If there are high levels of
748 H3K56ac, it is more likely that there will be repeated fork stalling (represented by three black
749 arrows) (see figure 8 from Hull *et al.* 2017).
750
751 **Figure 3. CN variation can alter gene expression.** (a) Consider a gene whose CN ranges from
752 0 to 4 (blue to black to red) among individuals (represented by dots) in a population (middle
753 gene). (b) Generally, CN and gene expression (represented as arbitrary units or a.u.) correlate
754 with one another such that individuals with lower CN values will have lower levels of gene
755 expression of that gene while those with higher CN values will have higher levels of gene
756 expression.
757
758 **Figure 4. Comparison of genomic content affected by CN variants and SNPs in 3 fungal**
759 **species.** (A) SNPs per kb is not significantly correlated with CN variable base pairs per kb in *S.*
760 *cerevisiae* wine strains (blue; $r_s = 0.02$; $p = 0.78$; Spearman rank correlation) and *C. deuterogattii*

761 (red; $r_s = 0.06$; $p = 0.62$; Spearman rank correlation); the reverse is true in *S. pombe* (green; $r_s =$
762 0.67; $p < 0.01$; Spearman rank correlation). (b, left) CN variable base pairs per kb in wine strains
763 of *S. cerevisiae* is greater than *C. deuterogattii* and *S. pombe* ($p < 0.01$; Kruskal-Wallis and $p <$
764 0.01 for all Dunn's test pairwise comparisons with Benjamini-Hochberg multi-test correction).
765 (b, right) SNPs per kb is low among *S. cerevisiae* wine strains (*Scer*) compared to *S. pombe*
766 (*Spom*) but greater than a clonally expanded population of *C. deuterogattii* (*Cdeu*) ($p < 0.01$;
767 Kruskal-Wallis and $p < 0.01$ for all Dunn's test pairwise comparisons with Benjamini-Hochberg
768 multi-test correction). Data from Jeffares et al., 2015, 2017 (*Spom*); Steenwyk et al., 2016
769 (*Cdeu*); Steenwyk and Rokas, 2017 (*Scer*).
770
771 **Figure 5. CN variable genes that affect functions important to wine making.** Functional
772 categories (e.g., Cu and Fe homeostasis, maltose metabolism, etc.) are shown in black font.
773 Genes of interest are shown proximal to the category described and are colored blue, red, or
774 purple to represent a gene observed to be primarily deleted, duplicated, or both across
775 populations and studies investigating *S. cerevisiae* wine strains. Genes found to be both
776 duplicated and deleted present an opportunity for oenologists to capitalize on standing genetic
777 diversity to select for particular flavor profiles or yeast performance.
778
779 **Box 1.** Standard population genetic principles of shifts in allele frequencies (Felsenstein, 1976;
780 Moritz, 1994) can be applied to CN variants. To illustrate the case, we provide an example using
781 the *CUP1* locus, where high CN provides protection against copper poisoning (Fogel and Welch,
782 1982), of how the allele frequency of a CN variant can increase through its phenotypic effect.
783 Suppose that in a yeast population exposed to copper that all individuals do not harbor CN

784 variation at the *CUP1* locus. Through a mutational event, a beneficial *CUP1* allele that contains
785 two or more copies of the locus may appear in the population. (a) Yeast with two or more copies
786 of *CUP1*, which in turn lead to higher *CUP1* protein levels, will be better and more efficient at
787 copper sequesteration unlike the parental allele and therefore avoiding copper poisoning (Fogel
788 and Welch, 1982). (b) Assuming a large population size and strong positive selection, changes in
789 allele frequency will occur in the population due to changes in yeast survivability and ability to
790 propagate. More specifically, the frequency of the beneficial allele (i.e., *CUP1* duplications) will
791 increase depending on the strength of selection, which increases as the concentration of
792 environmental copper increases, and the parental allele will decrease.

Figure 1

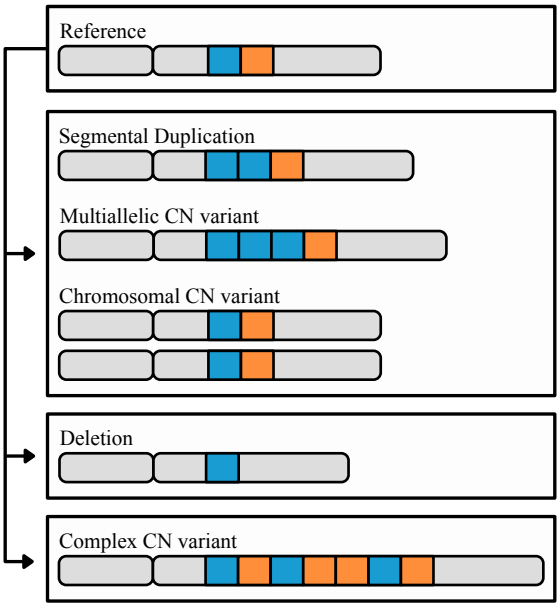

Figure 2

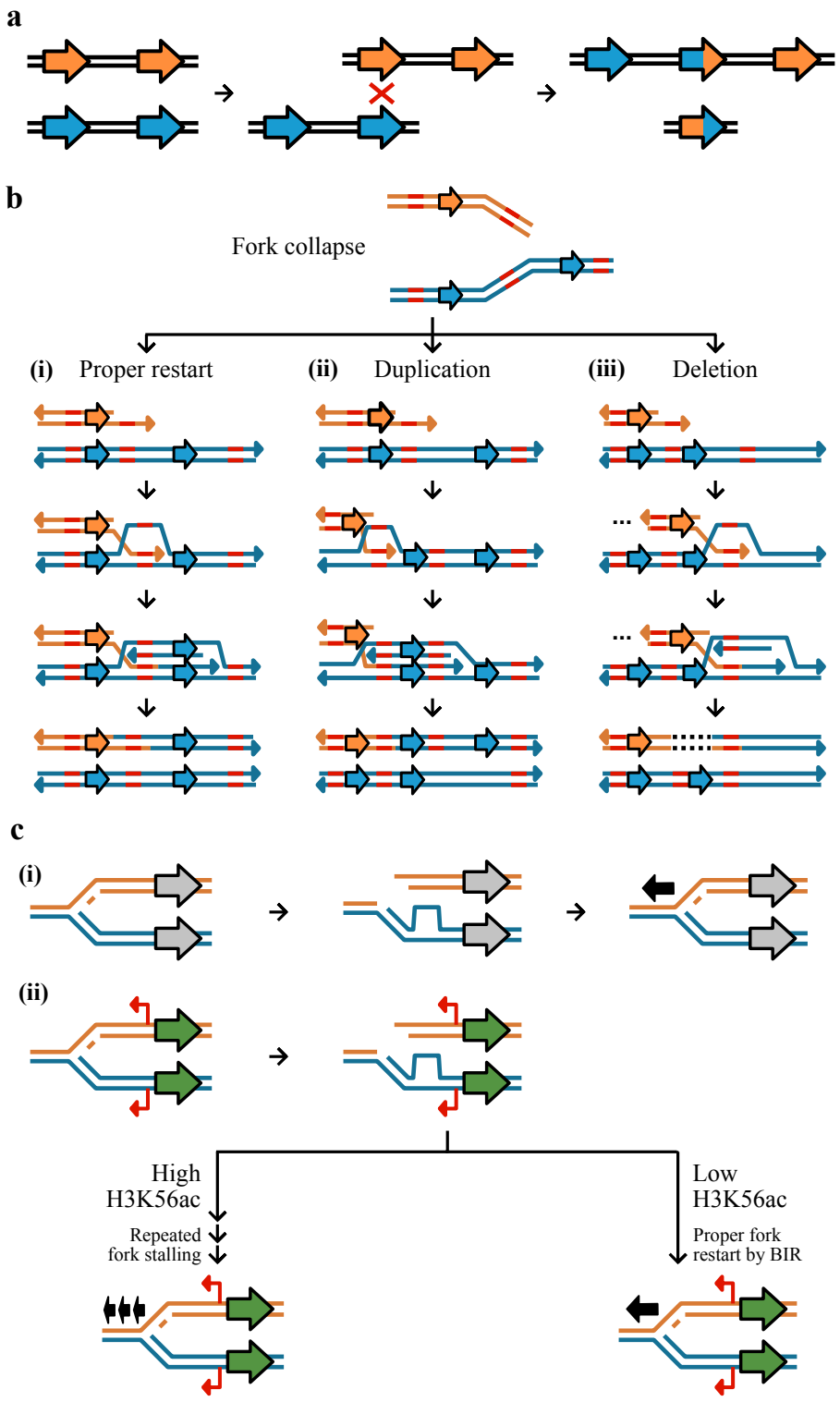

Figure 3

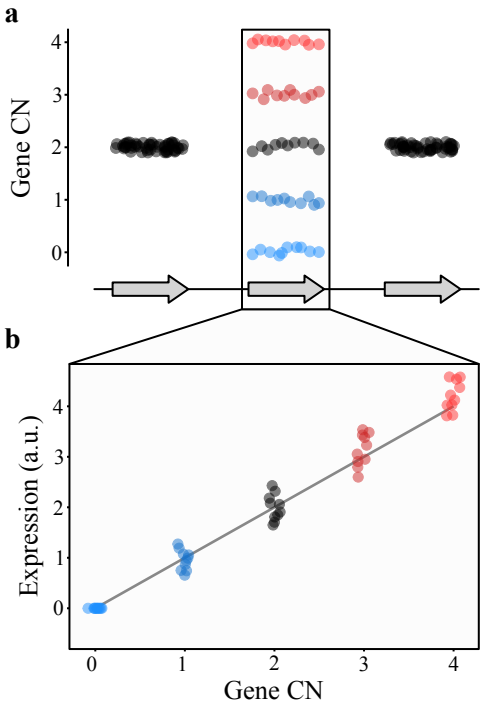

Figure 4

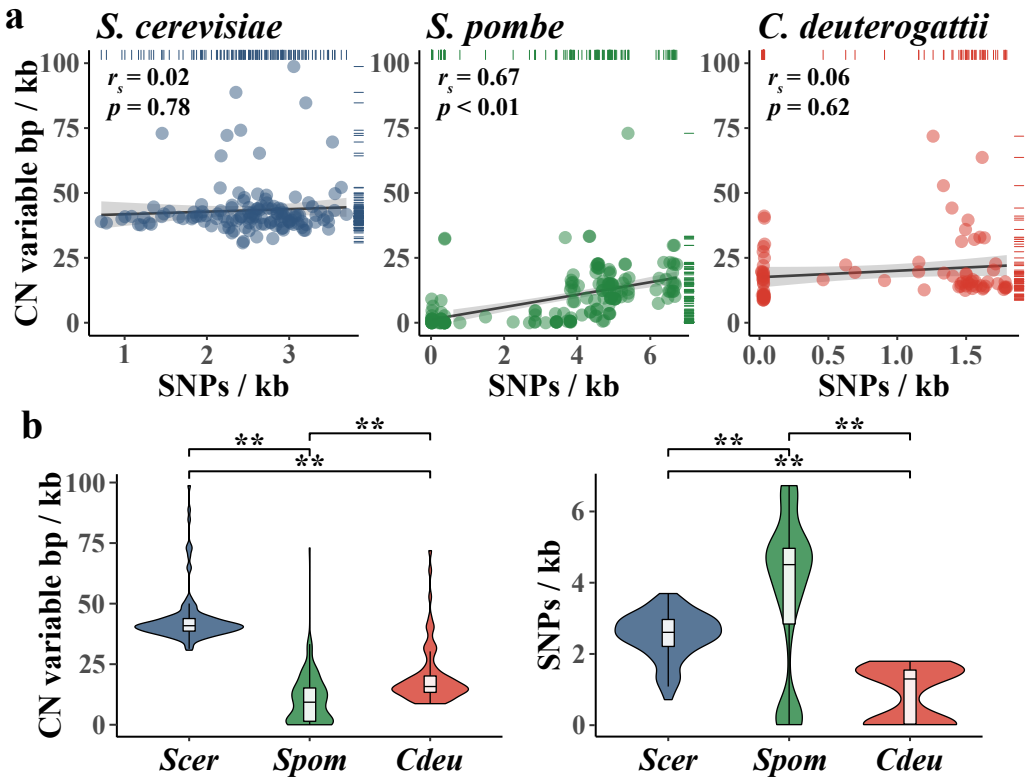

Figure 5

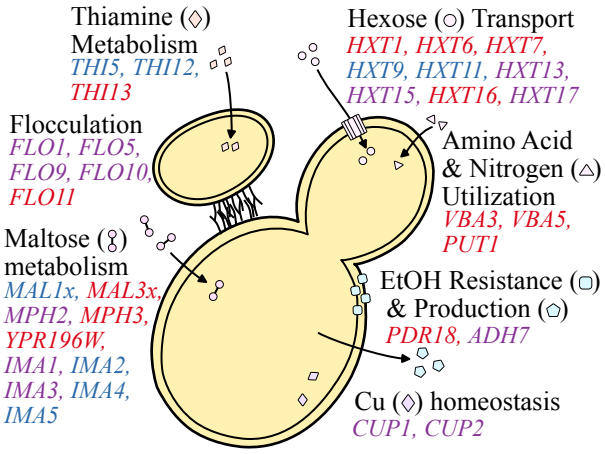

Figure within Box 1