Washington University in St. Louis
Washington University Open Scholarship

All Theses and Dissertations (ETDs)

5-24-2010

Genetic and Phenotypic Differentiation between Winemaking and Wild Strains of Saccharomyces cerevisiae

Katie Hyma
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/etd

Recommended Citation
Hyma, Katie, "Genetic and Phenotypic Differentiation between Winemaking and Wild Strains of Saccharomyces cerevisiae" (2010). *All Theses and Dissertations (ETDs)*. 888.
https://openscholarship.wustl.edu/etd/888

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Evolution, Ecology, and Population Biology

Dissertation Examination Committee:
Justin Fay, Chair
Jim Cheverud
Barak Cohen
Ken Olsen
Barbara Schaal
Heather True-Krob

GENETIC AND PHENOTYPIC DIFFERENTIATION BETWEEN WINEMAKING AND WILD STRAINS OF SACCHAROMYCES CEREVISIAE

by

Katie Elizabeth Hyma

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2010

Saint Louis, Missouri

copyright by

Katie Elizabeth Hyma

2010

Abstract

Traditionally, *Saccharomyces cerevisiae* has been associated with wine, beer and bread production, yet wild strains have also been isolated from natural habitats. While all strains of *S. cerevisiae* as well as other *Saccharomyces* species are capable of wine fermentation, a genetically distinct group of *S. cerevisiae* strains is primarily used to produce wine. These strains exhibit an apparent genetic bottleneck, which led to the hypothesis that wine stains have been domesticated from 'wild' natural strains. However, it is unknown whether the genetic bottleneck was accompanied by selection for phenotypic differences.

In this study we tested for phenotypes correlated with the genetic bottleneck observed for wine strains. First, growth and fitness parameters (e.g. growth rate) of yeast strains were evaluated on different media types that simulated winemaking and natural habitats. Results provided no evidence that 'wine' or 'wild' strains have greater fitness in their respective environments, and suggest that the putative domestication has not resulted in habitat specific growth adaptation. Second, we tested for phenotypes associated with human perception of wine aroma and flavor characteristics using discriminatory and descriptive sensory analysis. The results from this study established human perception as a selectable yeast phenotype, and demonstrated that divergence in wine aroma and flavor attributes is consistent with the domestication hypothesis.

The isolates used to infer domestication are geographically broad, but ecologically undersampled. We tested the relevance of global population genetic patterns in *S. cerevisiae* by conducting a population genetic study of *S. cerevisiae* isolated from vineyard and non-vineyard locations in North America. We used genome-wide single nucleotide markers to determine if the domestication hypothesis is supported at a local scale. Results demonstrate that two distinct populations of *S. cerevisiae* exist in North America, corresponding to European 'wine' and North

American ‘wild’ genotypes. We provide evidence for genetic exchange between populations, suggesting a lack of physical or temporal barriers to gene flow. While wine strains exhibit a population genetic pattern consistent with previous studies, we find that the wild population is dominated by a few clonal genotypes, identifying new questions regarding the domestication hypothesis and the genetic structure of other wild populations.

Acknowledgement

I gratefully acknowledge the financial support provided by the Division of Biology and Biomedical Sciences (Program in Ecology, Evolution and Population Biology), and the NIH Genome Analysis Training Fellowship.

Numerous faculty, staff and students have provided support for this research in what seems like infinitely many ways. I would like to especially thank my advisor, Justin Fay, for welcoming me in to his lab and encouraging me to pursue this research project. He provided me with intellectual support throughout the project, provided great insight into all aspects of my research, and was even willing to wield a pipette when I couldn't. In addition, I am indebted to my thesis committee members: Dr. Barbara Schaal, Dr. Ken Olsen, Dr. Heather True-Krob, Dr. Jim Cheverud, and Dr. Barak Cohen, all of whom were supportive of me and provided thoughtful comments throughout this project. My labmates Betsy Engle and Devjanee Swain demonstrated great patience while teaching me the ins and outs of the lab and the yeast system, making it a pleasure to work in the Fay lab. Justin Fay, Betsy Engle and Devi Swain also provided a tremendous amount of help during the 2009 sampling season, and also during sensory analysis experiments. I would also like to thank the other members of the Fay lab: Scott Doniger, Hyun Kim, Sung Chun, Katie Reed, Andrew Bergen, Sandeep Venkataram, Juyoung Huh, Adam Coster and Linda Riles for all of their help and support in the lab on many experiments, and for all of their useful comments and suggestions during the time we spent working together.

I learned a great deal from my fellow students, and many of them provided specific support in regard to this dissertation. I would especially like to thank a few in particular. Kim Lorenz designed restriction associated DNA sequencing DNA barcodes, shared detailed protocols and reagents for genotyping. Justin Gerke was a sounding board during early stages of this project, providing invaluable advice and instruction. Vitas Wagner accompanied me to a hot

mosquito infested vineyard to help collect samples while he was rotating in the Fay lab. Jeff Sabina fielded questions regarding enzymatic assays and cloning methods. Chris Hittinger shared *S. cerevisiae* and *S. kudriavzevii* strains, and patiently taught me how to do competition experiments using flow cytometry. I have had the pleasure of experiencing graduate school surrounded by students in the Evolution, Ecology and Population Biology program, all of whom are incredibly talented, providing lively discussions on interesting topics. In particular, Brad Oberle, Wayne Law, Jennifer Neuwald, James Beck, Thom Sanger, Michele Johnson, Brian Allan, Liam Revell, Josh Reece, Nick Griffin, and Nikki Miller all provided excellent academic and emotional support.

The support teams that keep research moving forward at Washington University are absolutely incredible. In particular, both Debbie Peterson and Alice Gleason helped me find solutions to any problems that I ran in to while conducting research on this project, and were always happy to help. Without the help of Jessica Hoisington-Lopez the genotyping I did for this project would not have been completed. My graduate coordinators, Dana Sterbenz and Melissa Torres were an invaluable resource, providing both administrative and emotional support. Without support like theirs, this research could not have been completed.

Although I cannot name them, I would like to thank all of the people who participated in the discriminatory sensory analysis of wines. They were incredibly generous with their time, and played the role of a truly invaluable measuring instrument.

I would also like to thank those people who welcomed me onto their property to collect yeast samples, or provided technical assistance on various aspects of these projects. Specifically, I would like to thank Tony Saballa of Charleville Vineyard (Ste. Genevieve, MO) and Hank Johnson of Chaumette Vineyard (Ste. Genevieve, MO), Mark Baehman of Mt. Pleasant Vineyard (Augusta, MO), Tom and Celeste Symonette of Whistling Dog Cellars (Polk Co., OR),

Marilee Buchanan of Tyee Vineyards (Benton Co, OR), Matt Gross of the Wine and Cheese Place (St. Louis, MO), Dave Deaton of St. Louis Wine and Beer Making (St. Louis, MO), Denise Gardner of Vinquiry, Inc., Richard DeScenzo of ETS Lab services, Lidia Watrud and Mike Bollman. Many conversations with these individuals brought up some of the most interesting questions regarding my research.

My undergraduate advisor, Tom Whittam, is the reason I entered graduate school, and although I can't share this moment with him, his memory provides a constant source of inspiration and encouragement.

My family and friends provided me with an incredible amount of support throughout this process. My Mom, Kathy Hyma, is always there when I need her and always has faith in me. She was ready to help whenever she could, and even got involved in my research by helping to wash and sterilize hundreds of wine bottles during an unfortunately timed visit. My Dad was thrilled to the point of embarrassing when I graduated the 8th grade, and I know that if he were able to have been a part of this process, he would have been equally proud. My wonderful friends Ana Jacobsen and Carrie Vodehnal, who have been a constant source of help and support, also generously donated their time to help with the execution of a post-experiment wine tasting for participants in the sensory analysis. There is one family member who deserves acknowledgement in almost area I've mentioned so far. Jason Londo has, with great love, patience and a seemingly limitless sense of humor been a constant source of support and companionship. He has played the roles of research assistant and manuscript editor to crisis counselor, housekeeper and personal chef.

Table of Contents

Abstract ii

Acknowledgement iv

Table of Contents vii

List of Tables and Figures xi

Introduction 2

The Natural History of *Saccharomyces* Yeasts 3

A Brief History of Winemaking and Viticulture 5

The Population Genetics of *Saccharomyces* Yeasts 6

Domestication 8

Chapter 1 : Fitness related phenotypes of *Saccharomyces cerevisiae* strains from diverse habitats. 11

Introduction 12

Materials and Methods 16

Growth of S. cerevisiae under simulated environmental pressures 16

Resistance to copper sulfate 17

Statistical Analysis 17

Results 18

Growth of S. cerevisiae under simulated environmental pressures 18

Differences in copper sulfate resistance between vineyard and oak strains of S. cerevisiae. 19

Discussion 21

Chapter 2 : Divergence in wine characteristics produced by wild and domesticated strains of *Saccharomyces cerevisiae* 42

Introduction 43

Materials and Methods 46

Yeast strains and fermentation 46

Discriminatory sensory evaluation 46

Quantitative descriptive sensory analysis 47

Chemical Analysis 49

Results 50

Human discrimination of wines fermented using wine yeast and non-wine yeast 50

Quantification of sensory attributes 52

Chemical Analysis 57

Chapter 3 : Isolation and Characterization of *Saccharomyces* species from Vineyard and Non-Vineyard locations 76

Introduction 77

Materials and Methods 79

Study Sites 79

Sampling and Enrichment 79

Isolate screening and species identification 80

Assimilation of carbon and nitrogen sources 82

Fermentation of carbon sources 83

Copper sulfate resistance 83
Statistical Analysis 83
Results 84
Evaluation of enrichment procedures for the recovery of Saccharomyces sensu stricto species 84
Variation in isolation rates of Saccharomyces species 85
Composition of a spontaneous fermentation 86
Carbon and nitrogen utilization 86
Copper sulfate resistance in yeast communities 87
Discussion 89
Chapter 4 : Population Genetics of Vineyard and Non-Vineyard Populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* 111
Introduction 112
Materials and Methods 115
Strains 115
Genotyping 115
Sequence Analysis 116
Statistical Analysis 119
Phylogenetic Analysis and Population Structure 119
Genetic Admixture and Introgression 120
Results 121
Distribution of Genetic Markers and Heterozygosity in S. cerevisiae 121

Phylogenetic Analysis of S. cerevisiae 121

The Distribution of S. cerevisiae Diversity 122

Allele frequencies in S. cerevisiae 124

Linkage Disequilibrium in S. cerevisiae 125

Population Structure within S. cerevisiae 125

Genetic Admixture and Introgression in S. cerevisiae 126

Phylogenetic Analysis and Distribution of Genetic Diversity in S. paradoxus 127

Population Structure in S. paradoxus 128

Discussion 129

Genetic Differentiation between Saccharomyces cerevisiae Populations 129

Genetic Variation and Nucleotide Diversity within Saccharomyces cerevisiae Populations 133

Differences in Population Genetic Structure between Saccharomyces cerevisiae and Saccharomyces paradoxus 134

Conclusions 137

Chapter 5 : Conclusions, Inferences and Future Directions 182

Domestication Phenotypes in S. cerevisiae 183

The Genetics Signatures of Domestication in S. cerevisiae 187

Literature Cited 191

List of Tables and Figures

Table 1-1. Media used for simulating environmental pressures. 24

Table 1-2. Strains used in this study. 25

Figure 1-1. Example of the copper sulfate resistance assay. 30

Figure 1-2. Variation in growth parameters between wine and oak strains in grape juice and synthetic oak exudate. 31

Figure 1-3. The effects of osmotic stress on growth parameters for wine and oak strains of *S. cerevisiae.* 32

Figure 1-4. Maximum growth rate in grape juice for wine and oak strains of *S. cerevisiae*. 33

Figure 1-5. Maximum growth rate in grape juice for other strains of *S. cerevisiae*. 34

Figure 1-6. The effects of a non-fermentable carbon source on growth parameters for wine and oak strains of *S. cerevisiae.* 35

Figure 1-7. The effect of nutrient limitation on growth parameters for wine and oak strains of *S. cerevisiae.* 36

Figure 1-8. Maximum density in Yeast Peptone Ethanol for wine and oak strains of *S. cerevisiae*. 37

Figure 1-9. Maximum density in Yeast Peptone Ethanol for other strains of *S. cerevisiae.* 38

Figure 1-10. Copper sulfate resistance by *S. cerevisiae* strain category. 39

Figure 1-11. Copper sulfate resistance in wine and oak strains of *S. cerevisiae*. 40

Figure 1-12. Copper sulfate resistance in other strains of *S. cerevisiae*. 41

Table 2-1. Yeast strains used in this study. 65

Figure 2-1. Humans can discriminate between wines fermented using different strains of *S. cerevisiae*. 66

Table 2-2. Discrimination of wines produced by *S. cerevisiae* and *S. paradoxus* strains. 67

Table 2-3. Discrimination of wines produced by palm wine, sake, grape wine and wild strains of *S. cerevisiae*. 68

Figure 2-2. Grape wine and non-grape wine strains of *S. cerevisiae* are differentiated for wine flavor and aroma attributes. 69

Table 2-4. Significant correlations between descriptive and chemical attributes. 70

Table 2-5. PCA loadings and ANOVA p values for individual attributes. 73

Table 2-6. Tukey's HSD (Honestly Significant Difference) p values. 74

Figure 2-3. Grape wine strains produce desirable wine attributes and wild strains produced undesirable wine attributes. 75

Figure 3-1. Sampling locations for this study. 93

Figure 3-2. Sampling and isolation procedure for *Saccharomyces* spp. 94

Table 3-1. Samples collected in 2008. 95

Table 3-2. Samples collected in 2009. 96

Table 3-3. Multiplex PCR assay for *Saccharomyces* sensu stricto. 97

Table 3-4. Size of expected fragments for *Saccharomyces* sensu stricto ribotyping. 98

Table 3-5. The performance of different enrichment media on overall isolate recovery in 2008... 99

Table 3-6. The proportion of bacteria recovered by different enrichment media in 2008. 100

Table 3-7. The proportion of *Saccharomyces* species recovered by different enrichment media in 2008. 101

Table 3-8. The proportion of *Saccharomyces* species recovered from different substrates in 2008. 102

Figure 3-3. The number of *Saccharomyces* sensu stricto isolates recovered by location in 2008. 103

Figure 3-4. The proportion of *S. cerevisiae* and *S. paradoxus* isolates from Missouri and Oregon in 2008. 104

Figure 3-5. Variation in carbon and nitrogen use between *Saccharomyces* and non-*Saccharomyces* species from Missouri. 105
Figure 3-6. Variation in carbon and nitrogen use between oak and grape yeast communities in Missouri vineyards. 106
Figure 3-7. Variation in carbon and nitrogen use by non *Saccharomyces* sensu-stricto yeast species living on oak trees at three different locations. 107
Figure 3-8. Variation in carbon and nitrogen use between bark, twig and soil non *Saccharomyces* sensu stricto yeast isolates. 108
Figure 3-9. Copper sulfate resistance in vineyard and non-vineyard yeast communities. 109
Figure 3-10. Copper sulfate resistance in vineyard and non-vineyard strains of *S. cerevisiae.* 110
Table 4-1. RAD genotyped *S. cerevisiae* strains collected in this study. 139
Table 4-2. RAD genotyped *S. paradoxus* strains collected in this study. 141
Table 4-3. Additional RAD genotyped strains. 143
Table 4-4. Published *S. cerevisiae* genomes (SGRP) used in this study. 144
Table 4-5. Published *S. paradoxus* genomes (SGRP) used in this study. 146
Table 4-6. Unpublished genome sequences used in this study. 148
Table 4-7. Alignment and sequencing statistics for *S. cerevisiae.* 149
Table 4-8. Alignment and sequencing statistics for *S. paradoxus.* 152
Table 4-9. SNPS and heterozygosity for *S. cerevisiae.* 154
Table 4-10. SNPs and heterozygosity for *S. paradoxus.* 156
Table 4-11. False discovery rate estimates for solexa RAD genotyping. 158
Figure 4-1. SNP density in *S. cerevisiae.* 159
Figure 4-2. Neighbor-Joining phylogeny of *S. cerevisiae.* 160
Figure 4-3. Neighbor Joining phylogeny of *S. cerevisiae* wine strains. 161
Figure 4-4. Neighbor Joining phylogeny of *S. cerevisiae* wild strains. 162
Figure 4-5. Geographical distribution of *S. cerevisiae* genotypes within Chaumette Vineyards. 163

Table 4-12. Nucleotide diversity (π) within and between wine and wild strains of *S. cerevisiae*. 164

Figure 4-6. The proportion of private alleles at 4-fold degenerate and nondegenerate sites in wine and wild strains of *S. cerevisiae.* 165

Table 4-13. The proportion of private alleles in wine and wild strains of *S. cerevisiae.* 166

Figure 4-7. Expected and observed minor allele frequencies for *S. cerevisiae.* 167

Figure 4-8. Expected and observed minor allele frequencies in wine and wild strains of *S. cerevisiae.* 168

Figure 4-9. Linkage disequilibrium in *S. cerevisiae*. 169

Figure 4-10. Linkage disequilibrium in wine and wild *S. cerevisiae* populations. 170

Figure 4-11. Average likelihood and variance for population structure simulations of *S. cerevisiae.* 171

Figure 4-12. Inferred population structure of *S. cerevisiae*. 172

Figure 4-13. Admixture between wine and wild genotypes of *S. cerevisiae*. 173

Table 4-14. Inferred ancestry of admixed *S. cerevisiae* isolates. 174

Figure 4-14. Neighbor-joining phylogeny of *S. paradoxus*. 175

Figure 4-15. Neighbor-joining phylogeny of *S. paradoxus*. 176

Figure 4-16. Neighbor-joining phylogeny of American *S. paradoxus* isolates. 177

Figure 4-17. Neighbor-joining phylogeny of European *S. paradoxus.* 178

Table 4-15. Nucleotide diversity in *S. paradoxus.* 179

Figure 4-18. Minor allele frequencies in *S. paradoxus.* 180

Figure 4-19. Population structure in *S. paradoxus.* 181

“in vino veritas”

Pliny, the Elder

Historia Naturalis XIV, 141

Introduction

The budding yeast *Saccharomyces cerevisiae* is one of the most widely used and best understood model organisms in biological and biomedical research, and is important for many applications from the production of food and beverages to pharmaceuticals, biofuel production and bioremediation of toxins. Additionally, it was the first organism for which the entire nuclear genome was sequenced (Goffeau et al. 1996). Despite its importance in modern genetics, the population genetics of the species has not been investigated until recently (Winzeler et al. 2003; Fay & Benavides 2005; Aa et al. 2006; J. L. Legras et al. 2007; Liti et al. 2009; Schacherer et al. 2009; Diezmann & Dietrich 2009). Traditionally *S. cerevisiae* has been associated with wine, beer and bread production, but wild strains have also been identified and isolated from more natural habitats, specifically in association with oak trees (Naumov et al. 1998; Sniegowski et al. 2002; Sampaio & Gonçalves 2008). Although *S. cerevisiae* and many different yeast species are capable of alcohol fermentation (de Deken 1966), a genetically distinct group of *S. cerevisiae* strains are primarily used for wine production (Fay & Benavides 2005). Additionally, the strains associated with wine production are genetically differentiated from natural isolates. This genetic differentiation, combined with an observed reduction in genetic diversity within wine strains has led researchers to hypothesize that wine strains were domesticated from wild *S. cerevisiae* (Fay & Benavides 2005). As many previous studies of domesticated plant and animal species have demonstrated, domestication can drastically alter the genetic structure of a species, potentially leading to the fixation of maladaptive traits and genome wide changes in the level of diversity (Doebley et al. 2006). These types of changes could have significant implications for studies using *S. cerevisiae* as a model system and as well as for industrial applications.

The Natural History of *Saccharomyces* Yeasts

Saccharomyces cerevisiae is a member of a closely related group of yeast species referred to as *Saccharomyces* sensu stricto. *Saccharomyces* species within this group include *S. cerevisiae, S. paradoxus, S. cariocanus, S. bayanus, S. mikatae, and S. kudriavzevii,* along with

S. pastorianus, a sterile hybrid between *S. bayanus* and *S. cerevisiae* (Replansky et al. 2008). Of these species, *S. cerevisiae* and the closely related *S. paradoxus* have been the most extensively collected and described. *Saccharomcyes cerevisiae* is unique among the sensu stricto group due to its intimate association with humans and use as a model organism.

While *S. cerevisiae* can be isolated from fermentations and grapes in vineyards, wild strains of *S. cerevisiae* have also been isolated from a variety of natural sources and have been frequently found in association with oak tree exudates, bark and soil (Sniegowski et al. 2002; Naumov et al. 1998). In comparison, *S. paradoxus* is rarely found in association with vineyards but is frequently found in association with oak trees (Sniegowski et al. 2002; Naumov et al. 1998; Johnson et al. 2004; Naumov et al. 1997; Glushakova et al. 2007; Koufopanou et al. 2006; Redzepovic et al. 2002; Yurkov 2005). A number of the other *Saccharomyces* sensu stricto species have also been found in association with oak trees and soil, and in some instances occur in sympatry with *S. cerevisiae* and *S. paradoxus* (Sniegowski et al. 2002; Naumov et al. 1998; Sampaio & Gonçalves 2008; Naumov et al. 2003). Aside from the differences in habitat between *S. cerevisiae* and *S. paradoxus*, the species are nearly indistinguishable. However, there is some evidence that these species may differ in physiological traits. For example, the two species appear to have diverged in their thermal growth profiles (Sweeney et al. 2004), and it is hypothesized that this is a mechanism that allows for coexistence of sympatric *Saccharomyces* species on oak trees (Sampaio & Gonçalves 2008; Sweeney et al. 2004).

The dispersal mechanisms of *Saccharomyces* species are not well understood, but it has been hypothesized that insects (e.g. *Drosophila, Apies)* may be vectors for long distance migration (Mortimer & Polsinelli 1999; Goddard et al. 2010). *Saccharomyces* species have been found in the intestinal tract of wild *Drosophila* species (Phaff & Knapp 1956), though they were not detected at *Drosophila* breeding sites, and the strains found in association with juveniles differ from those found in association with adults (Shehata et al. 1955; Carson et al. 1956), indicating

that *S. cerevisiae* associations with insects need further examination. In addition to potential insect vectors, there is also evidence that *S. cerevisiae* may be dispersed in association with winemaking equipment including transport on wine barrels (Goddard et al. 2010), and that *S. paradoxus* may be dispersed via acorns (H. Zhang et al. 2010).

Rates of gene flow and the effects of differences in the mode of reproduction in *Saccharomyces* are poorly understood. *Saccharomyces* species can reproduce sexually if, after sporulation (meiotic cell division), two haploid strains of opposite mating types are brought into contact, though individual cells can also reproduce clonally via mitotic cell division. The rates of outcrossing under natural conditions have not been studied directly, although population genetic analyses have estimated outcrossing rates between 0.002% and 25% for *S. cerevisiae* (Ruderfer et al. 2006; Ezov et al. 2006; Goddard et al. 2010), and around 1% for *S. paradoxus* (Johnson et al. 2004; Tsai et al. 2008). A recent laboratory based study demonstrated a much higher rate of out-crossing among *S. cerevisiae* strains (up to 40%) (Murphy & Zeyl 2010). Interestingly, one study found that passage through the intestinal tract of *D. melanogaster* increased rates of outcrossing by tenfold (Reuter et al. 2007), possibly through the partial digestion of the ascus (Reuter et al. 2007). Thus, rates of out-crossing in nature could be directly influenced if dispersal via insect vectors is common. There is also evidence for variation in the level of sporulation between 'wine' and 'wild' strains of *S. cerevisiae* (Gerke et al. 2006), which implies that differences in outcrossing rates may also be likely. While certain aspects of the natural habitat of *S. cerevisiae* have begun to be examined, the demography and population genetics of this species remains relatively unknown.

A Brief History of Winemaking and Viticulture

The earliest evidence for wine fermentation comes from the molecular analysis of pottery jars that have been dated as far back as 7000 BC (McGovern et al. 2004), and extraction of DNA

from ancient wine containers is consistent with the presence of *S. cerevisiae* (Cavalieri et al. 2003). Formal domestication of the grapevine (*Vitis vinifera* L.), which is inextricably linked to winemaking, is believed to have occurred through horticultural practices employed by hunter-gatherers in the Fertile Crescent (McGovern 2003). The domesticated Eurasian grape vine, along with the practice of winemaking, appears to have spread to Egypt and Lower Mesopotamia around 3500-3000 B.C., and to Crete by 2200 B.C. (McGovern 2003; This et al. 2006).

Although the Eurasian grapevine *Vitis vinifera* L. subsp. *sylvestris* represents only one of hundreds of known grape species worldwide, it is responsible for 99% of the world's wine production (McGovern 2003). Despite a plethora of wild grape species available for domestication in Asia, North America and Europe, the Eurasian vine has been preferentially imported for the production of wine (McGovern 2003). Since its domestication, this single subspecies of grape has contributed to the proliferation of up to 10,000 cultivars (McGovern 2003; This et al. 2006). The wild *Vitis vinifera* species that gave rise to these domesticated cultivars still exists in Eurasia, although in a geographically limited range and only in areas with intact woodlands and sufficient water supply (McGovern 2003; This et al. 2006). Despite the rich anthropological and archaeological history of winemaking and viticulture, it wasn't until 1866 that *S. cerevisiae* was formally described as the organism responsible for fermentations (Pasteur 1866).

The Population Genetics of *Saccharomyces* Yeasts

Recent studies of the population genetics of *S. cerevisiae* have revealed that population structure is primarily determined by ecological habitat, rather than by geography (Liti et al. 2009; Schacherer et al. 2009; Fay & Benavides 2005; Legras et al. 2007; Aa et al. 2006). Strains isolated from and associated with wine and vineyards represent the most thoroughly sampled group, and genetic analysis between wine strains and other strains indicate a genetic bottleneck that has resulted in a 30% reduction in genetic diversity in wine strains compared to the species

as a whole (Fay & Benavides 2005). Additional fermentation associated strains for sake (Fay & Benavides 2005) and palm wine (Ezeronye & Legras 2009), although not well represented in population genetics studies, also show evidence of population bottlenecks within *S. cerevisiae*. Studies examining genetic data from complete genome sequencing of a large sampling of *S. cerevisiae* strains from many diverse habitats have revealed the presence of five distinct genetic lineages, including wine strains, wild North American strains, sake strains, Malaysian strains and West African strains (Liti et al. 2009). A large proportion of the sequenced strains, however, showed evidence for mixed genetic backgrounds (Liti et al. 2009). Additionally, this genome sequence data supports previous evidence for a bottleneck associated with wine strains, as well as the correlation of genotypic variation with ecological habitat, rather than geographical location.

Although *S. cerevisiae* and *S. paradoxus* are nearly indistinguishable phenotypically, the collection and description of isolates from several locations reveal very different population genetic patterns. Isolates of *S. paradoxus* exhibits a structure consistent with geography, and show a pattern of isolation by distance (Koufopanou et al. 2006; Liti et al. 2009).Three main genetic groups have been discovered within *S. paradoxus* that represent populations from the Far East, Europe and the Americas, along with a single strain from Hawaii. These groups are clearly differentiated; the majority of genetic variation in the species is due to differences between rather than within populations (Liti et al. 2009; Koufopanou et al. 2006). The amount of genetic variation that has been recovered from *S. paradoxus* is also substantially greater than within *S. cerevisiae* (Liti et al. 2009). In fact, the amount of variation within a single *S. paradoxus* population is nearly equal to that within the entire *S. cerevisiae* species (Liti et al. 2009). One potential explanation for the major differences in population genetic structure between *S. cerevisiae* and *S. paradoxus* is the association of *S. cerevisiae* with human activity (Liti et al. 2009; Legras et al. 2007). Although these species are clearly being shaped by different processes, further characterization of both

species, both at the genetic and phenotypic level, are needed to determine what those processes are and to examine the hypothesis of wine strain domestication.

Domestication

Domesticated species frequently exhibit a suite of phenotypes referred to as the 'domestication syndrome.' Such phenotypes can be maladaptive, and in some cases cause the organism to become completely dependent on humans (Doebley et al. 2006). In plants these phenotypes can include larger fruits or grains, more robust plants, determinate growth, loss of seed dispersal mechanisms, and physiological changes such as loss of seed dormancy, decrease in bitter secondary products, photoperiod changes and flowering synchrony (Doebley et al. 2006). In animals the domestication syndrome typically involves behavioral phenotypes such as reduction in responsiveness to stimuli, reduced activity, increased social compatibility and intensified sexual behavior (Zeder 2006). Morphological changes in domesticated animals such as reduced horn length and changes in pelt coloration are typically thought to represent secondary traits selected for after initial domestication that correlate with behavioral selection (Zeder 2006).

Genetic bottlenecks associated with the domestication of crops are common, as only a select number of individuals are repeatedly propagated (Doebley et al. 2006). The extent of genetic reduction in diversity is dependent on the size of the population during domestication and the length of time of domestication (Eyre-Walker et al. 1998). This process has also been associated with the relaxation of selective constraints, resulting in an excess of nonsynonymous substitutions in the domesticated lineage (Lu et al. 2006). Although putatively domesticated strains of *S. cerevisiae* exhibit both a population bottleneck (Fay & Benavides 2005; Liti et al. 2009a) as well as a slight increase in the ratio of nonsynonymous to synonymous mutations

(Doniger et al. 2008), it is unknown whether the genetic signatures of domestication in fungi are similar to those in plants and animals.

There are very few studies of yeast or other fungi that have been domesticated by humans, although humans use many species of fungi in food production (Hesseltine 1965). The only example of a study on the genetics of domestication in a fungal species comes from *Aspergillus oryzae,* a fungus used to prepare soy sauce, sake and miso that is thought to have been domesticated from wild populations of *Aspergillus flavus* (Geiser et al. 1998). These two species are phenotypically nearly indistinguishable aside from the production of the secondary metabolite aflatoxin (Geiser et al. 1998). Although the domestication is dated at approximately 2,000 years ago, there seems to be no genetic signature associated with the domestication (Rokas 2009). It may be that bottlenecks in fungi are less severe because entire populations of cells rather than a few individuals are selected on every generation, or that persistent asexual reproduction decreases the efficacy of selection (Rokas 2009).

For *S. cerevisiae,* the genetic differentiation between wine and wild strains combined with the observed reduction in genetic diversity within wine strains suggests that wine strains may have been domesticated from wild strains (Fay & Benavides 2005). However, it is not clear whether genetic differentiation of wine and wild strains is a consequence of divergence driven by selection on specific phenotypes relevant to wine making, or a consequence of restricted gene flow following an initial founder event without local adaptation or selection for desirable phenotypes. Additionally, the evidence for domestication thus far is based on genetic patterns, yet the isolates used to generate this hypothesis span decades and sampling of any one ecological habitat is limited, potentially confounding the signature of domestication with cryptic population structure or demography.

In this study we address the domestication hypothesis for *S. cerevisiae* by first testing for phenotypic correlates with the genetic bottleneck for wine strains. In the first chapter we evaluate the potential for local adaption through fitness differences by measuring growth rate of *S. cerevisiae* on various media intended to simulate natural environments. In the second chapter we test the hypothesis that humans have selected for differences in wine aroma and flavor phenotypes of *Saccharomyces* strains through the use of sensory analyses. Secondly we test the relevance of global population genetic patterns within *S. cerevisiae* and *S. paradoxus* populations through a population genetics study of *S. cerevisiae* and *S. paradoxus* isolated from vineyard and non-vineyard locations in North America. In the third chapter we describe the isolation and characterization of *S. cerevisiae* and *S. paradoxus* isolates from grapes and oak trees in North American vineyards and nearby non-vineyard locations. Finally, in the fourth chapter we examine the population genetic structure of *S. cerevisiae* and *S. paradoxus* from these habitats to determine if the genetic signatures of domestication that have been observed on a global scale can be recapitulated at the local scale.

Chapter 1 : Fitness related phenotypes of *Saccharomyces cerevisiae* strains from diverse habitats.

Introduction

The preferential fermentation of sugar into ethanol in the presence of oxygen, referred to as the crabtree effect (de Deken 1966), is a major evolutionary transition associated with gene duplication in yeast (Thomson et al. 2005). Based on the inferred ancestral sequences of duplicated genes, this phenotype arose about 80 million years ago, during the Cretaceous period, and may coincide with the origin of fleshy fruits (Thomson et al. 2005). The duplicated genes, and the resulting crabtree effect mechanism are present in many yeast species (de Deken 1966; Thomson et al. 2005) and are thought to have arisen as an adaptive mechanism through competitive exclusion of other species through both ethanol toxicity and resource competition (Boulton et al. 1996). Although many yeast species are capable of producing alcohol, *S. cerevisiae* is the dominant species responsible for wine fermentations worldwide (Mortimer 2000). Furthermore, only a small genetic subset of the species is used by humans in association with wine production (Johnston 1990), and subsequent genetic analysis has contributed to the hypothesis that these strains of *S. cerevisiae* have been domesticated by human selection from natural populations for wine making (Fay & Benavides 2005). Support for domestication, however, is limited to the genetic evidence of a population bottleneck associated with wine making strains (Fay & Benavides 2005) without additional correlative phenotypic evidence.

Domesticated species commonly exhibit a suite of phenotypes that are collectively referred to as the ‘domestication syndrome’. For domesticated plant species these phenotypes typically include larger fruits or grains, more robust plants, determinate growth, loss of seed dispersal mechanisms, and physiological changes such as loss of seed dormancy, decrease in bitter secondary products, photoperiod changes and flowering synchrony (Doebley et al. 2006). Such phenotypes can be maladaptive under wild conditions, and in extreme cases cause the organism to become completely dependent on humans (e.g. corn) for survival (Doebley et al. 2006). In animals the domestication syndrome typically involves behavioral rather than

morphological phenotypes such as reduction in responsiveness to stimuli, reduced activity, increased social compatibility and intensified sexual behavior (Zeder 2006). Morphological changes in domesticated animals (e.g. horn size reduction, pelt coloring) are typically thought to correlate with behavioral selection and relate to secondary phenotypes (Zeder 2006).

In contrast to numerous studies of plant and animal domestications and the genotypic and phenotypic results of human selection, there are very few studies of domestication for yeast or other fungi despite the use of many species of fungi in food production (Hesseltine 1965). The noted exception is *Aspergillus oryzae,* a fungus used to prepare soy sauce, sake and miso that is thought to have been domesticated from wild populations of *Aspergillus flavus* (Geiser et al. 1998). Genetic and phenotypic differentiation between the species is minimal (Rokas 2009) aside from the production of the secondary metabolite aflatoxin (Geiser et al. 1998), the carcinogen and human allergen associated with peanut allergy. With the lack of studies regarding domesticated fungi and minimal information from the *Aspergillus* system to draw from, the phenotypic and genotypic signatures of domestication in fungi remain unknown.

Vineyard and grape fermentation strains ('domesticated' wine strains) and oak tree and soil strains (oak strains) have been frequently sampled in previous studies and form ecologically cohesive groups according to habitat. Additionally, oak strains are hypothesized to represent the ancestral wild populations from which wine strains were domesticated (Fay & Benavides 2005). It is possible that the genetic differentiation observed between 'wild' and 'domesticated' strains of *S. cerevisiae* has resulted from adaptations to specific environmental pressures (e.g. high osmotic pressure). The habitat of 'domesticated' strains, grape juice and wine, is typified by a number of different environmental stresses that could represent selective pressures contributing to the genetic differentiation of winemaking strains of *S. cerevisiae* from wild populations. The habitat conditions associated with wine making include high osmotic stress, low pH, rapid environmental changes and low nitrogen, as well as exposure to copper sulfate, sulfite, tannins and various

other chemical compounds. The 'wild' oak tree habitat, on the other hand, has low nutrients, rapid alterations in temperature, desiccation and freezing stress as well as many other variable environmental stresses. Similar to reduced fitness of feral crop species that escape cultivated fields (Gressel 2005), it is possible that the domesticated strains of *S. cerevisiae* are less fit in the 'wild' habitat than natural populations. In addition to these differences in habitat, previous studies have shown that strains of *S. cerevisiae* are variable in many of their growth and fermentation-related phenotypes, and have classified *S. cerevisiae* strains into two categories, 'ants' and 'grasshoppers' (Spor et al. 2008; Spor et al. 2009), The 'grasshoppers', which include wine strains, are characterized by increased glucose consumption, increased cell size, reduced carrying capacity, reduced reproduction rate in fermentation and increased reproduction under respiration conditions. In contrast, 'ants' include wild and lab strains and are characterized by the opposite traits. However, variation within the species is distributed between both extremes (Spor et al. 2008). The wine and oak tree habitats are likely to impart additional stresses for which *S. cerevisiae* strains vary in their fitness response, and may further drive phenotypic differences between wine and oak strains. Previous studies have demonstrated significant variation within *S. cerevisiae* in response to natural and human-associated environmental stresses (Kvitek et al. 2008), including copper sulfate resistance (Fay et al. 2004).

In addition to wine fermentations and oak trees, *S. cerevisiae* has been successfully isolated from various other habitats, including non-wine fermentations, various plant species (e.g. cactus, palm), and other natural sources (Shehata et al. 1955; Ezeronye & Legras 2009; Goddard et al. 2010). Strains from other types of fermentations (e.g. Sake, ragi, palm wine) also form genetically distinct groups (Fay & Benavides 2005; Liti et al. 2009), but are relatively undersampled in comparison to wine and oak strains and require further study.

In this study we tested the hypothesis that the domestication of wine strains from wild strains has resulted in the respective adaptation to the habitats associated with winemaking and

oak trees. To determine if wine and oak strains differ in their growth and fermentation characteristics, we used six different media types chosen to simulate the stresses that strains would be exposed to under “wild” conditions or during wine making (e.g. osmotic stress, decreased or limiting nutrient conditions, and non-fermentable carbon sources). We measured maximal growth rate, maximal density, and maximal growth interval in an effort to quantify differences in fitness between strains. We also tested strains for resistance to copper sulfate, a known human-applied selective pressure associated with the control of microbial growth in vineyards (Mortimer 2000). While we focused on comparisons between wine strains and oak strains, we attempted to place them into context with additional strains obtained from other habitat types (e.g. lab, clinical).

Materials and Methods

Growth of S. cerevisiae under simulated environmental pressures

In order to test the effects of simulated environmental pressures on *S. cerevisiae* strains, we examined growth in six different media types designed to simulate the domesticated and wild environments, and to test for the effect of specific environmental pressures found in those environments (Table 1-1). Media types include: YPD (10 g/L yeast extract, 20 g/L peptone, 2% w/v dextrose), YPD14 (10 g/L yeast extract, 20 g/L peptone, 14% w/v dextrose), YPE (10 g/L yeast extract, 20 g/L peptone, 3% v/v ethanol), Minimal Media (MM) (6.7 g/L yeast nitrogen base, 2% w/v glucose), Synthetic oak exudates (SOE) (Murphy et al. 2006) (1% sucrose, 0.5% fructose, 0.1% yeast extract 0.15% peptone), and chardonnay grape juice (GJ) from Vintners Reserve Chardonnay kits (Winexpert, Port Coquitlam, B.C., Canada). A total of 88 *S. cerevisiae* strains from a variety of sources were included in this study. Each strain is classified according to the source from which it was isolated (e.g. wine, oak, nature, fermentation, clinical, other) (Table 1-2). Strains were grown in 2 ml of YPD overnight and then diluted 1:1000 into 800 µl of each media type, which was contained in a deep well 96 well plate. Plates were covered with adhesive foil, pierced with a pinhole over each well, covered with sterile 96 well plate covers, then incubated at 30ºC without shaking. At 0, 12, 16, 24, 48 and 72 hours, cultures were mixed, 100 µl of media was removed, for which optical density (OD) was measured using a microplate spectrophotometer (BioTek, Winooski, MD). Growth was characterized by three different parameters: 1) the maximum OD, 2) the maximum growth rate, and 3) the interval during which the maximum growth rate occurred. Growth rate was measured as the maximum change in OD between two timepoints, and growth intervals 1 through 5 correspond to the time between hours 0-12, 12-16, 16-24, 24-48, and 48-72 hours. Two replicate experiments were performed for each strain.

Resistance to copper sulfate

Strains were grown overnight in YPD. From the overnight cultures, 200 µl was removed, and centrifuged at 3,000 rpm for 5 minutes. The resulting supernatant was decanted, the cells were resuspended in 1 ml of water and adjusted to OD=1. From the adjusted culture, a serial dilution series in water was created with the following steps: 1:10 (100 µl -> 900 µl), 1:10 (100 µl -> 900 µl), 1:3 (100 µl -> 200 µl), 1:3 (100 µl -> 200 µl), 1:10 (100 µl -> 900 µl). After the dilution was prepared, 5 µl of the OD=1 culture and 5 µl each of the 5 dilutions in the series were dispensed onto YPD agar plates with 2.5 mM Copper Sulfate. Six strains were measured per plate, along with two control strains (M22* and S288c). Three replicate assays were performed for each strain. Plates were photographed and scored by three independent judges on a scale from 0 to 3, where 0 = no growth/ complete inhibition, 1 = low growth/ almost complete inhibition, 2 = high growth/ very little inhibition, 3 = healthy growth at all dilutions with no inhibition. Half scores were allowed (e.g. 0.5, 1.5, 2.5). See Figure 1-1 for an example photograph of the assay.

Statistical Analysis

Statistical significance was examined using Chi-square tests and Fisher's exact test. When the number of levels for the test was greater than 2, a Chi-square test was performed. If the Chi-square test was significant ($p < 0.05$), fishers exact tests for each contrast was performed independently. Statistical analysis was carried out using R (R Development Core Team 2009).

Results

Growth of S. cerevisiae under simulated environmental pressures

To determine if there are any differences in the inherent growth and fermentation abilities of wine and wild strains, we measured growth rate in several media types intended to simulate environmental pressures that correspond to the natural habitats of these strains (Table 1-1). We tested for significant differences between wine and oak strains, and for comparisons that were significantly different between wine and oak strains, we also show the variation within other fermentation strains and other natural strains of *S. cerevisiae*.

First, we examine growth of wine and oak strains in grape juice and in synthetic oak exudates, intended to represent the natural habitats of these strains. We find a significant difference in maximum growth rate between wine and oak strains in grape juice, but no significant differences in synthetic oak exudate (Figure 1-2). Interestingly, it is the oak tree strains that have a higher maximum growth rate in grape juice ($p < 0.001$), although there is no significant difference in the maximum density which they ultimately reach, and they both reach the maximum growth rate in the 16-24 hour interval (Figure 1-2). However, the lower growth rate in wine strains appears to be due to a handful of strains, as the variation within wine strains is quite high compared to within oak strains (Figure 1-4). Only one oak strain, DBVPG1373, showed a relatively low growth rate in grape juice. Other non-wine fermentation strains and non-oak natural strains also exhibit a high degree of variation in growth rate (Figure 1-5).

One of the major environmental stresses in grape juice is a very high concentration of glucose (osmotic stress). In order to determine if the growth differences between wine and oak strains that we observed in grape juice can be explained by differences in osmotic stress, we compared the growth of wine and oak strains in YPD media with 14% dextrose, which is similar to the concentration of sugar found in unfermented grape juice. We found that there were no

significant differences between wine and oak strains in this media (Figure 1-3), suggesting that the differences in growth rate we found in grape juice cannot simply be explained by different responses to osmotic stress.

Wine and oak strains may also show different capacity for respiratory growth rate, which could affect fitness in either habitat. We used a medium with ethanol as the carbon source (YPE), which is non-fermentable, thereby forcing the strains to grow through respiration. We compared growth on YPE to growth on YPD, which are identical aside from the carbon source. In YPD, the carbon source, dextrose, can be used for either fermentative or respiratory growth. We found no significant differences between wine and oak strains when grown on YPD, but found that oak strains grew to a significantly higher density on YPE than did wine strains, although there were no differences in their growth rate (Figure 1-6). Other natural strains also exhibit a high degree of variation in the density to which they grow in YPE (Figure 1-8, Figure 1-9).

One of the major differences between the wine habitat and the oak habitat is the difference in the quantity of nutrients available, although there might also be a difference in biological availability of certain types of nutrients. To test the effects of low nutrient availability versus specific nutrient limitation, we compared growth in minimal media (nutrient limiting) to growth in synthetic oak exudates (low nutrient). We found no significant differences between wine and oak strains for either of these types of media.

Differences in copper sulfate resistance between vineyard and oak strains of S. cerevisiae.

Copper sulfate resistance has been shown to be variable between in *S. cerevisiae* and is thought to represent an adaptation to vineyard life (Fay et al. 2004; Mortimer 2000). We tested for differences in resistance to copper sulfate between wine and oak strains, and we also placed this

variation into the context of other types of *S. cerevisiae* strains including baking and clinical isolates, as well as other fermentation and natural strains.

We found that wine strains are significantly more resistant to copper sulfate than are oak strains (Figure 1-10), and that most other types of strains are resistant. Aside from the oak strains, there is a large amount of variation within groups. Oak strains, with the exception of a few outliers (DBVPG1373 and DBVPG1788), demonstrate no resistance to copper sulfate, whereas many wine strains only show partial resistance (Figure 1-11). Other fermentation and natural strains also demonstrate a wide range of variation (Figure 1-12), although many of the other natural strains exhibit little to no resistance (Figure 1-12).

Discussion

In this study we tested the hypothesis that specific growth parameters (fitness) of *S. cerevisiae* correlate between habitat types (wine or oak). The comparisons of several different simulated environmental pressures did not support this hypothesis. Wine strains are no more fit in the simulated domestication environment (grape juice), nor are oak strains more fit in the simulated natural habitat (synthetic oak exudates). In contrast, the oak strains have a significantly greater growth rate in grape juice compared to wine strains, while no significant difference was measured between strains in synthetic oak exudate. Although in some cases domestication phenotypes (artificial selection) can mimic local adaptation through natural selection (i.e. flowering time in plants (Izawa 2007)), in other cases domestication phenotypes can be considered maladaptive (i.e. loss of seed dispersal mechanisms in corn (Doebley et al. 2006)). Although we did not measured fitness directly, if the growth phenotypes that we measured correlate to true fitness differences, wine strains are less fit than oak strains in grape juice, the 'domesticated' habitat, which may be evidence for a maladaptive domestication phenotype in *S. cerevisiae.*

The greater growth rate observed for oak strains relative to wine strains in grape juice was not observed for growth on high sugar media, and thus differences in grape juice are not likely due to osmotic stress but rather to some other aspect of grape juice chemistry, such as pH or nitrogen concentration. Although wild strains show a significant increase in growth rate in grape juice, both wine and wild strains ultimately reach the same density, suggesting that growth rate may not be related to maximal cell density. Previous studies have demonstrated variation in growth rate and density, but also in cell size (Spor et al. 2008; Spor et al. 2009). Our results may suggest similar differentiation in growth phenotypes, but as we used optical density as a proxy for cell density, differences in cell size were not measured, and could also have obscured differences in cell density. Additionally, the resolution provided by measuring optical density of strains grown

independently of one another cannot detect small differences in growth rate, which may have a large effect when strains are competing for resources. It remains to be seen whether wine strains and wild strains exhibit fitness differences when grown in competition with each other.

While previous studies suggested growth phenotype differences between 'wine' and 'wild' strains under low resource conditions (Spor et al. 2009), we find no difference between wine and wild strains in their ability to grow in low resource (SOE) or resource limiting (MM) environments. We do, however, find some evidence for differentiation in respiratory growth, although our results indicate that oak strains (ants) reach a higher density under respiratory conditions than wine strains (grasshoppers), in contrast to previous studies (Spor et al. 2009).

We also found differences between wine and oak strains in the maximum density reached when grown on a nonfermentable carbon source (YPE). It should be noted, however, that the maximum density and growth rate in YPE were very low, and difference between the maximum density for oak strains (0.078) and that for wine strains (0.059) is slight. It is possible that growth continues at low levels over a long period of time in YPE, and if grown over a period longer than 72 hours they would reach equal densities.

In contrast to the growth associated traits, the phenotype of copper sulfate resistance shows a definitive pattern; wine strains demonstrate greater resistance to copper sulfate than oak strains. With the exception of two strains (DBVPG1373 and DBVPG1788), no oak strains were able to grow in the presence of copper sulfate. In congruence with the pattern observed, the two aberrant strains that demonstrate resistance, although isolated from natural habitats, have been recently examined and were found to be genotypically similar to wine strains (Liti et al. 2009). Although copper sulfate resistance clearly differentiates wine and oak strains of *S. cerevisiae,* this trait is believed to have arisen from selective pressures applied within vineyards after the development of viticulture, representing a secondary trait associated with the domestication of *S.*

cerevisiae (Fay et al. 2004; Mortimer 2000). Variation in copper sulfate resistance within wine strains as well as in other *S. cerevisiae* strains is variable but high. On average, baking, clinical and fermentation strains exhibit high levels of resistance, whereas other naturally associated strains exhibit lower levels of resistance. High resistance in other human associated strains may indicate shared genetic heritage with wine strains or use of and selection from copper sulfate exposure beyond vineyards (e.g. general agricultural production).

To conclude, differences in growth-related phenotypes, measured as a proxy for fitness, were unable to differentiate between putatively domesticated 'wine' strains and 'wild' oak tree strains of *S. cerevisiae.* Specifically, we did not observe any correlation between the substrate from which strains were isolated and their simulated habitat, suggesting that growth rate phenotypes have not significantly diverged between strain types as a result of selection in different environments. However, it is possible that growth phenotypes have developed in response to the 'domesticated' environment of grape juice but the actual phenotypic differences may only be observed under conditions of competition. Strains were evaluated alone in media conditions without competition from additional microbes and most importantly, without competition from the contrasting yeast strain type. Future studies designed to incorporate resource competition between strains may refine the ability to detect phenotypic differences between strains.

Table 1-1. Media used for simulating environmental pressures.

Media Types		Simulated Environmental Pressure
Chardonnay grape juice	Yeast peptone dextrose (14% dextrose)	Osmotic stress
Yeast peptone dextrose (2% dextrose)	Yeast peptone ethanol (3% ethanol	Non-fermentable carbon source (respiratory growth)
Synthetic oak exudate	Minimal media	Low nutrients vs. limiting nutrients

Table 1-2. Strains used in this study.

Strain	Category	Location	Source	Notes
CLIB215-1	Baker	New Zealand	Baker	Monosporic clone
CLIB324-2	Baker	Vietnam	Baker	Haploid
YS2	Baker	Australia	Baker strain	SGRP
YS4	Baker	Netherlands	Baker strain	SGRP
YS9	Baker	Singapore	Baker strain	SGRP
273614X	Clinical	RVI, Newcastle, UK	Clinical isolate	SGRP
322134S	Clinical	RVI, Newcastle, UK	Clinical isolate	SGRP
378604X	Clinical	RVI, Newcastle, UK	Clinical isolate	SGRP
YJM280	Clinical	USA	n.a.	Monosporic clone
YJM320	Clinical	USA	n.a.	Monosporic clone
YJM326	Clinical	USA	n.a.	Monosporic clone
YJM421	Clinical	USA	n.a.	Monosporic clone
YJM428	Clinical	USA	n.a.	Monosporic clone
YJM436-1	Clinical	Europe	n.a.	Monosporic clone
YJM653-1	Clinical	n.a.	n.a.	Monosporic clone
YJM975	Clinical	USA	Clinical isolate	SGRP
YJM978	Clinical	USA	Clinical isolate	SGRP
YJM981	Clinical	USA	Clinical isolate	SGRP
CLIB382-1	Fermentation	n.a.	n.a.	Monosporic clone
DBVPG1853	Fermentation	Ethiopia	White Tecc	SGRP
DBVPG6040	Fermentation	Netherlands	Fermenting fruit juice	SGRP

Strain	Category	Location	Source	Notes
DBVPG6044	Fermentation	West Africa	Bili wine	SGRP
K11	Fermentation	Japan	Shochu Sake strain	SGRP
NCYC110	Fermentation	West Africa	Ginger beer from Z. officinale	SGRP
PW5	Fermentation	Nigeria	Palm Wine	Monosporic clone
UC5.1	Fermentation	Japan	Sake	Monosporic clone
UC5.2	Fermentation	Japan	Sake	Monosporic clone
Y12	Fermentation	Africa	Palm wine	SGRP
Y12-3	Fermentation	Africa	Palm Wine	Monosporic clone
Y9	Fermentation	Japan	Ragi (similar to sake wine)	SGRP
Y9-4	Fermentation	Indonesia	Ragi	Monosporic clone
Y9-7	Fermentation	Indonesia	Ragi	Monosporic clone
YJM269-1	Fermentation	n.a.	Apple Juice fermentation	Monosporic clone
CBS7960-2	Nature	South Africa	Sugar Cane	Haploid
DBVPG1106	Nature	Australia	Grapes	SGRP
NCYC361	Nature	Ireland	Beer spoilage strain from wort	SGRP
UWOPS03-461.4	Nature	Malaysia	Nectar, Bertam palm	SGRP
UWOPS05-217.3	Nature	Malaysia	Nectar, Bertam palm	SGRP
UWOPS05-227.2	Nature	Malaysia	Trigona, Bertam palm	SGRP
UWOPS83-787.3	Nature	Bahamas	Fruit, Opuntia stricta	SGRP
UWOPS87-2421	Nature	Hawaii	Cladode, Opuntia megacantha	SGRP
Y10-2	Nature	Philippines	Coconut	Monosporic clone
DBVPG1373	Oak	Netherlands	Soil	SGRP

Strain	Category	Location	Source	Notes
DBVPG1788	Oak	Finland	Soil	SGRP
IL-01	Oak	USA	Soil	Monosporic clone
NC-02	Oak	USA	Forest	Monosporic clone
T7	Oak	USA	Soil	Monosporic clone
YPS1009	Oak	USA	Oak tree	Monosporic clone
YPS128	Oak	Pennsylvania, USA	Oak	SGRP
YPS129	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS142	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS143	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS163	Oak	Pennsylvania, USA	Oak Exudate	Monosporic clone
YPS2052	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS2056	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS2057	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS2060	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS2066	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS2067	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS2070	Oak	Pennsylvania, USA	n.a.	Monosporic clone
YPS606	Oak	Pennsylvania, USA	Oak	SGRP
DBVPG6765	Other	Unknown	Unknown	SGRP
FL100	Other	n.a.	n.a.	Haploid, lab strain
S288c	Other	California, USA	Rotting fig	SGRP, lab strain
SK1	Other	USA	Soil	SGRP, lab strain

Strain	Category	Location	Source	Notes
W303	Other	Unknown	Unknown	SGRP
B5	Wine	n.a.	Italy	Monosporic clone
BC187	Wine	Napa Valley, USA	Barrel fermentation	SGRP
I14	Wine	Italy	Vineyard	Monosporic clone
I14-1	Wine	Italy	Vineyard	Monosporic clone
L-1374	Wine	Chile	Wine	SGRP
L-1528	Wine	Chile	Wine	SGRP
M15	Wine	Italy	Vineyard	Monosporic clone
M22	Wine	Italy	Vineyard	Monosporic clone
M22*	Wine	Italy	Vineyard	Monosporic clone
M29	Wine	Italy	Vineyard	Monosporic clone
M30	Wine	Italy	Vineyard	Monosporic clone
M33	Wine	Italy	Vineyard	Monosporic clone
M34	Wine	Italy	Vineyard	Monosporic clone
M7	Wine	Italy	Vineyard	Monosporic clone
PR	Wine	Paris, France	Red Star Wine Yeast	Monosporic clone
T73-1	Wine	Spain	Wine	Monosporic clone
WE372-1	Wine	South Africa	n.a.	Monosporic clone
Y55	Wine	France	Wine	SGRP
Y8	Wine	n.a.	Turkey	NRRL- y2411
YIIC17-E5	Wine	Sauternes, France	Wine	SGRP
YJM269	Wine	n.a.	Grape	Monosporic clone

SGRP indicates strains described in (Liti et al. 2009)

Strain names followed by -1, -2 etc. indicate different monosporic clones derived from the same parental strain.

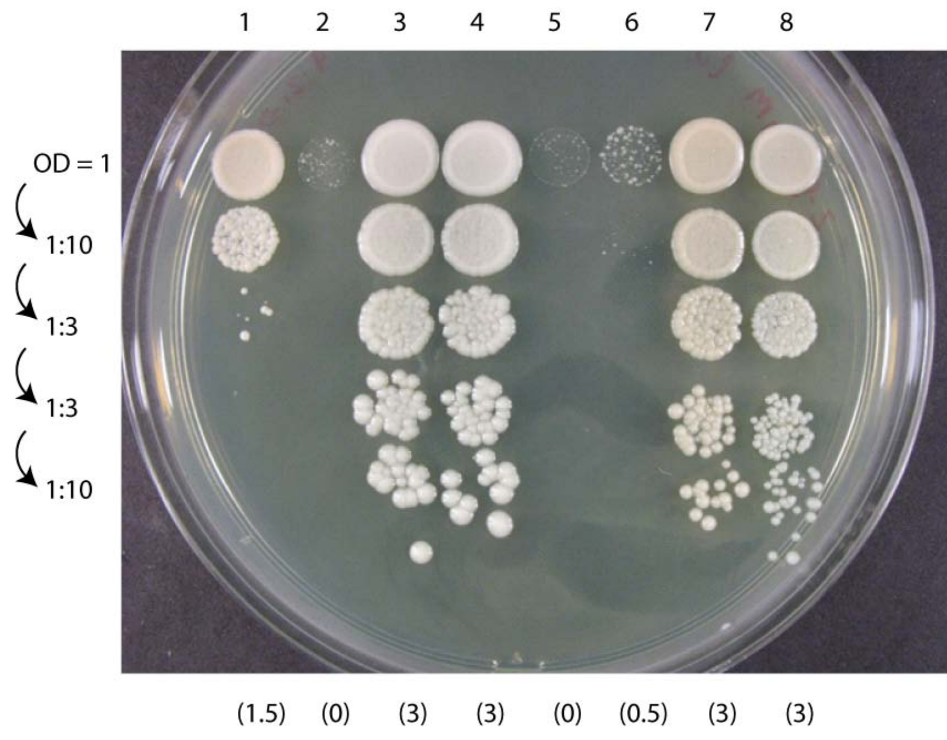


Figure 1-1. Example of the copper sulfate resistance assay.

Strains are in columns and the series increases in dilution from row one to row five. Judges scored photographs from 0 (no growth, column 5), to 3 (no inhibition, column 8).

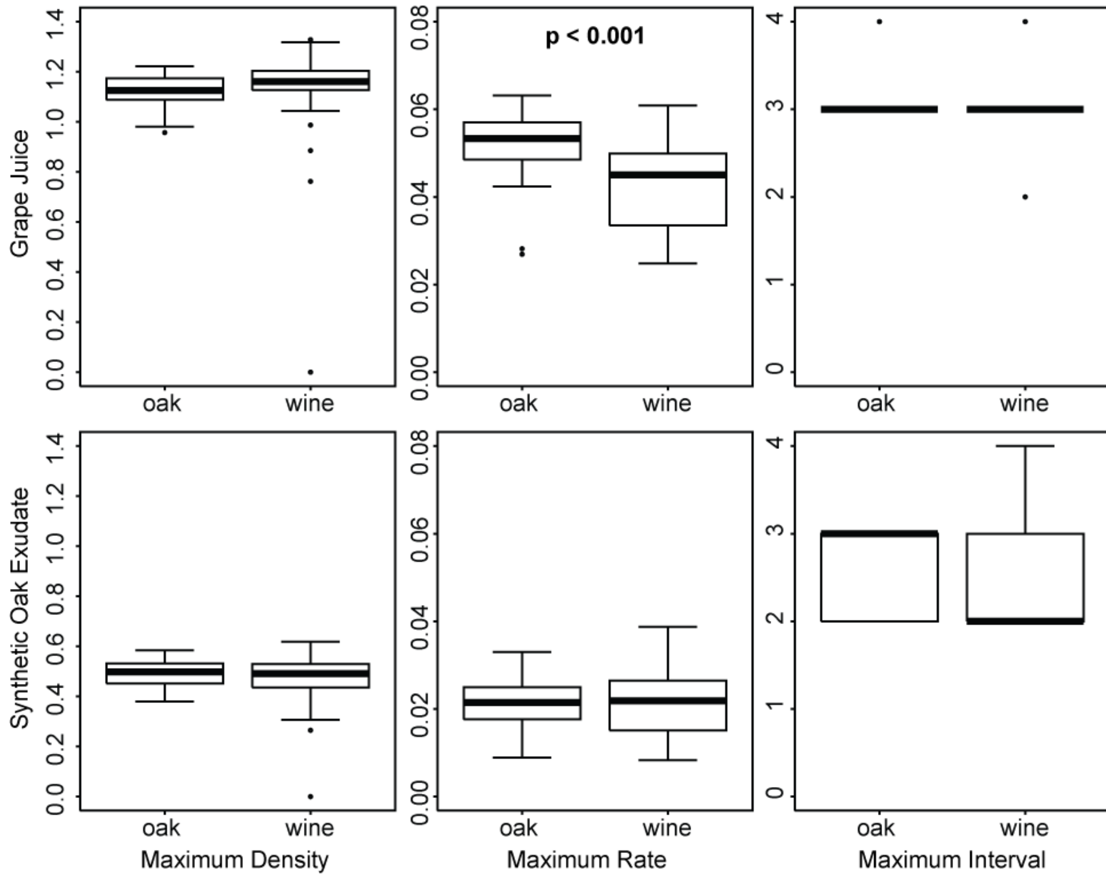


Figure 1-2. Variation in growth parameters between wine and oak strains in grape juice and synthetic oak exudate.

The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots.

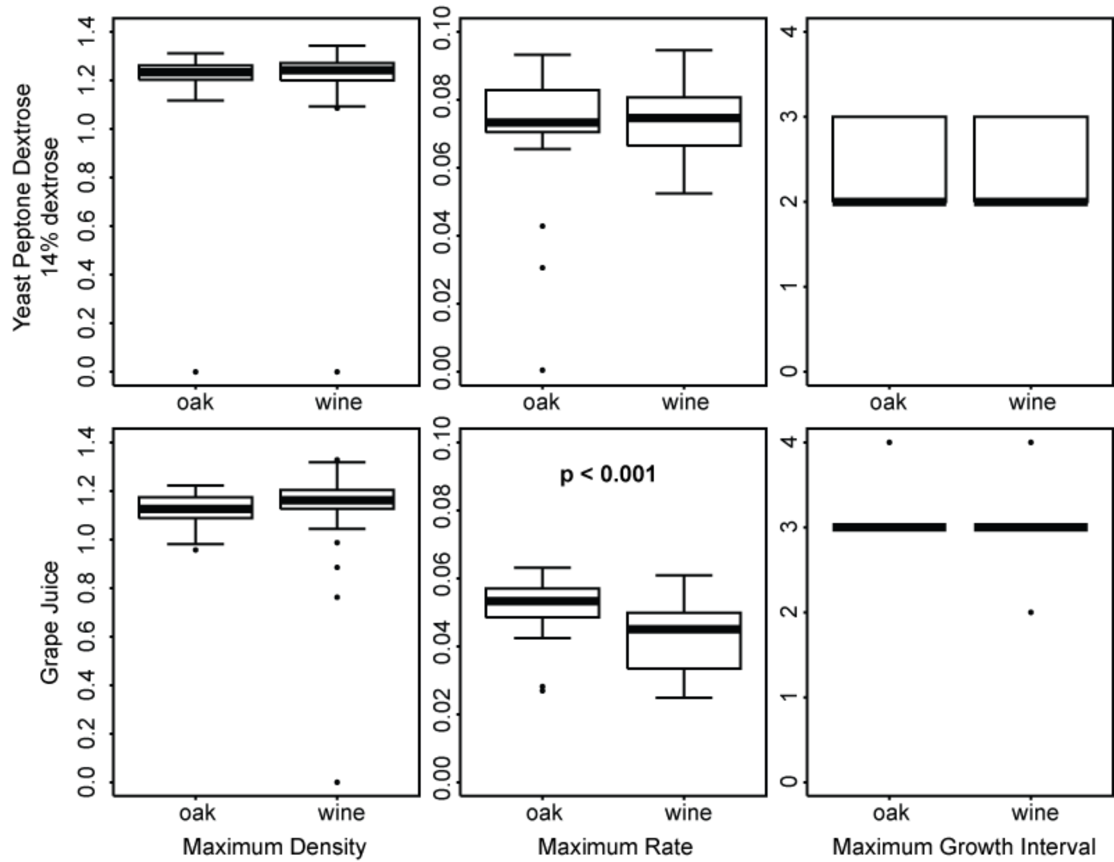


Figure 1-3. The effects of osmotic stress on growth parameters for wine and oak strains of *S. cerevisiae.*

The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots.

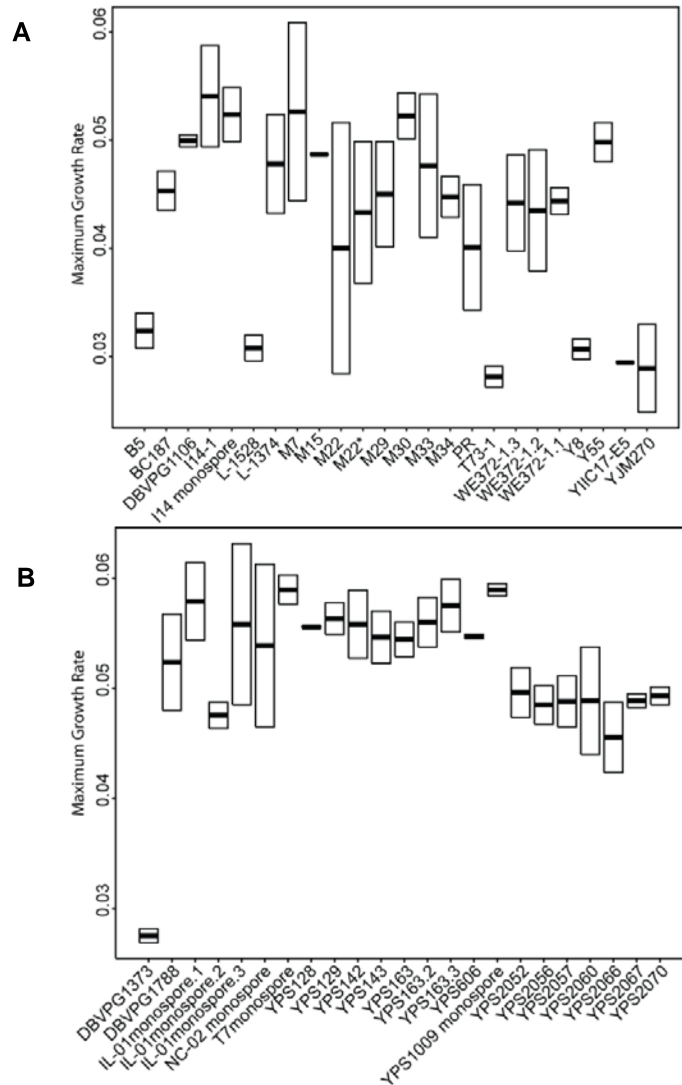


Figure 1-4. Maximum growth rate in grape juice for wine and oak strains of *S. cerevisiae*.

Maximum optical growth rate in (A) wine and (B) oak strains of S*. cerevisiae*. The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots.

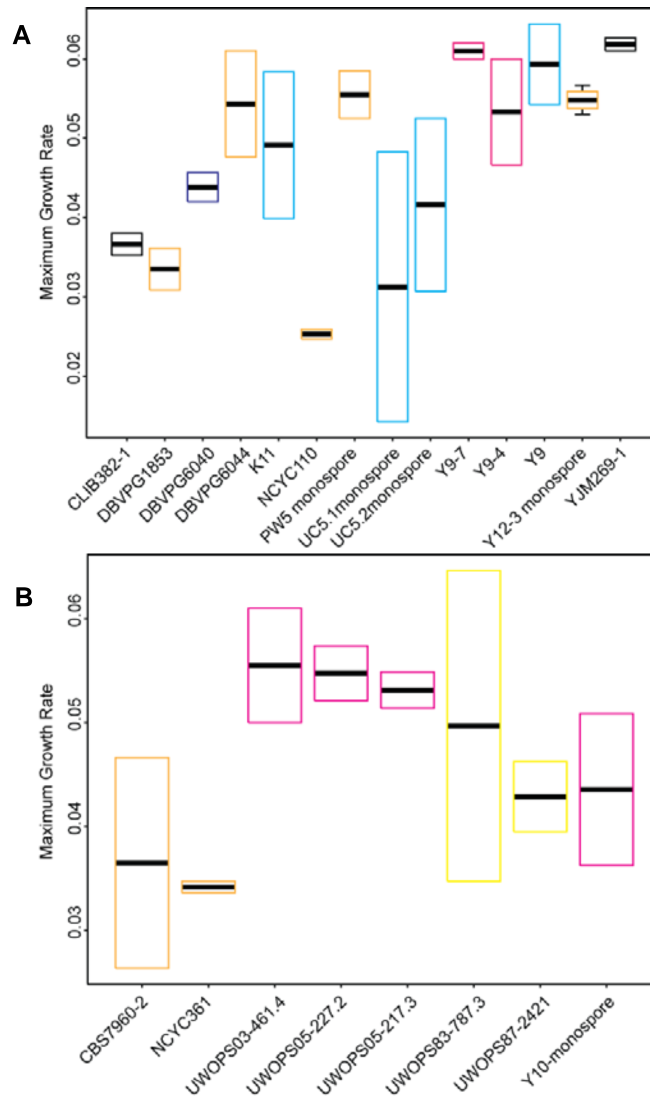


Figure 1-5. Maximum growth rate in grape juice for other strains of *S. cerevisiae.*

Maximum optical density for (A) non-wine fermentation strains and (B) other natural strains of S. cerevisiae. The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots. Strains from Africa are shown in orange, strains from Europe in blue, strains from Malaysia and Indonesia in pink, and strains from Hawaii and the Bahamas in yellow.

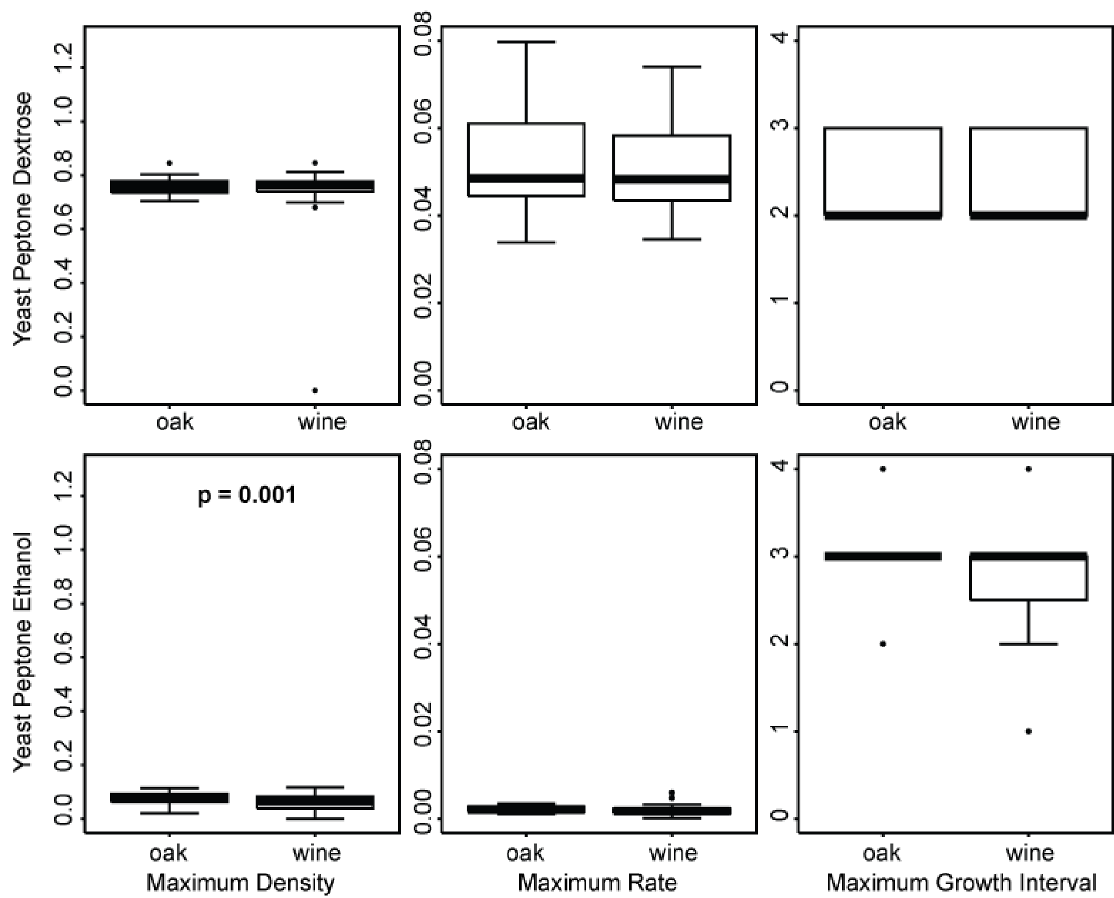


Figure 1-6. The effects of a non-fermentable carbon source on growth parameters for wine and oak strains of *S. cerevisiae.*

The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots.

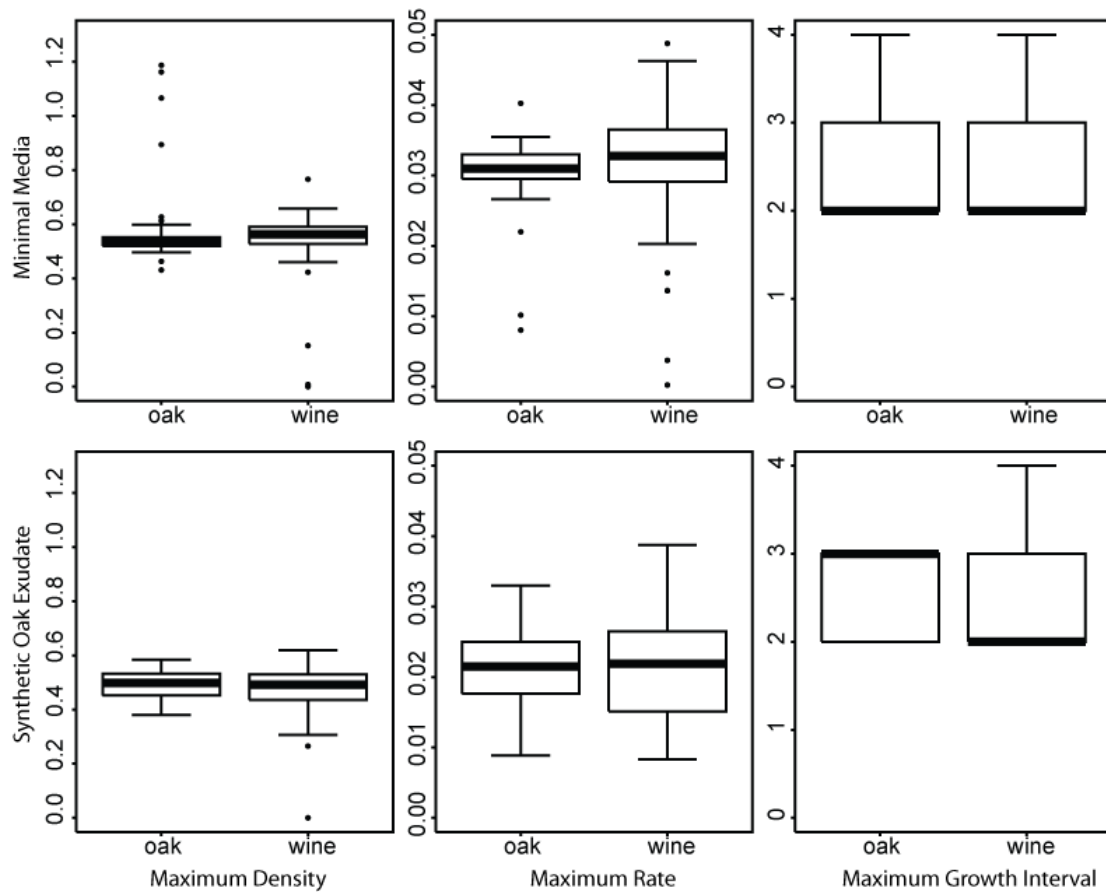


Figure 1-7. The effect of nutrient limitation on growth parameters for wine and oak strains of *S. cerevisiae.*

The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots.

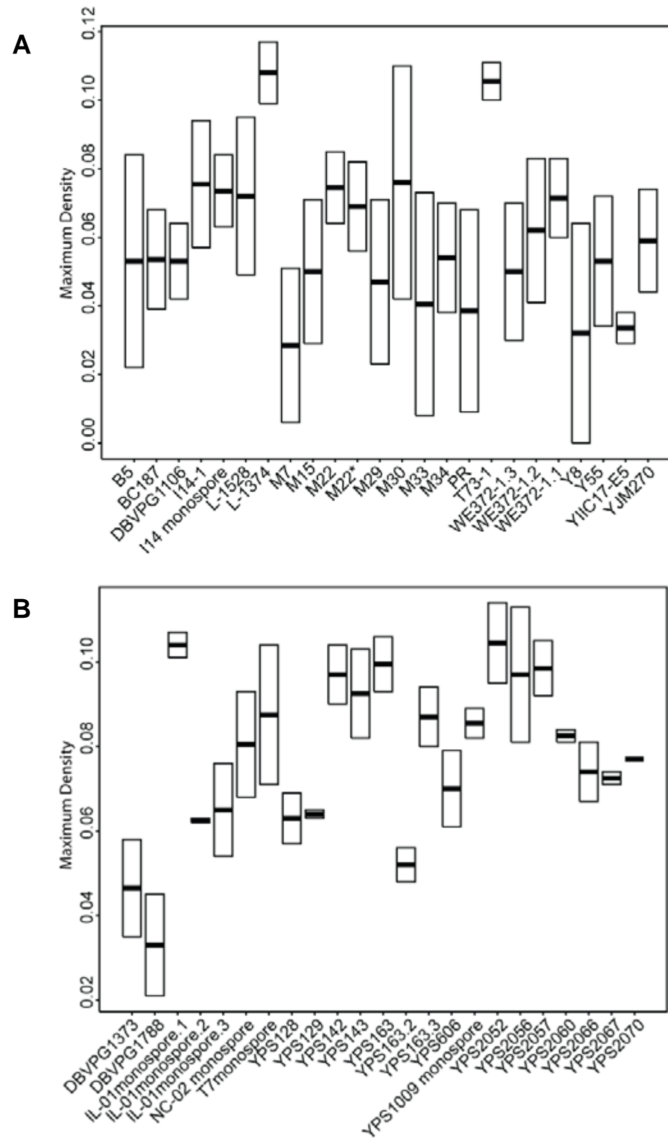


Figure 1-8. Maximum density in Yeast Peptone Ethanol for wine and oak strains of *S. cerevisiae*.

Maximum optical density in (A) wine and (B) oak strains of S*. cerevisiae*. The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots.

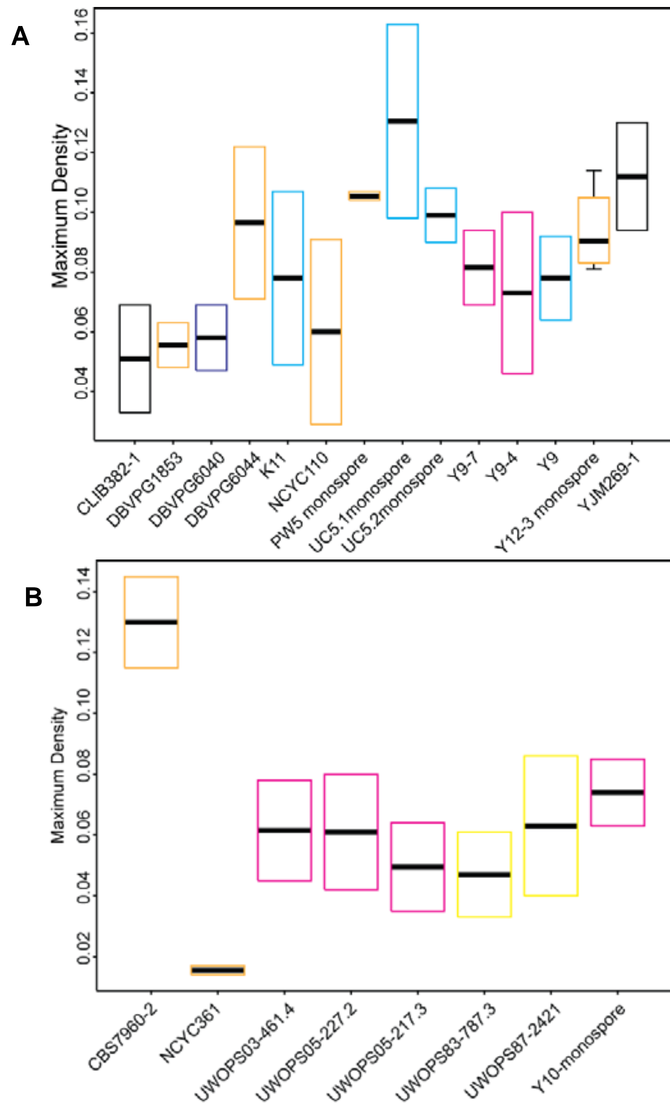


Figure 1-9. Maximum density in Yeast Peptone Ethanol for other strains of *S. cerevisiae.*

Maximum optical density for (A) non-wine fermentation strains and (B) other natural strains of S. cerevisiae. The mean score is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots. Strains from Africa are shown in orange, strains from Europe in blue, strains from Malaysia and Indonesia in pink, and strains from Hawaii and the Bahamas in yellow.

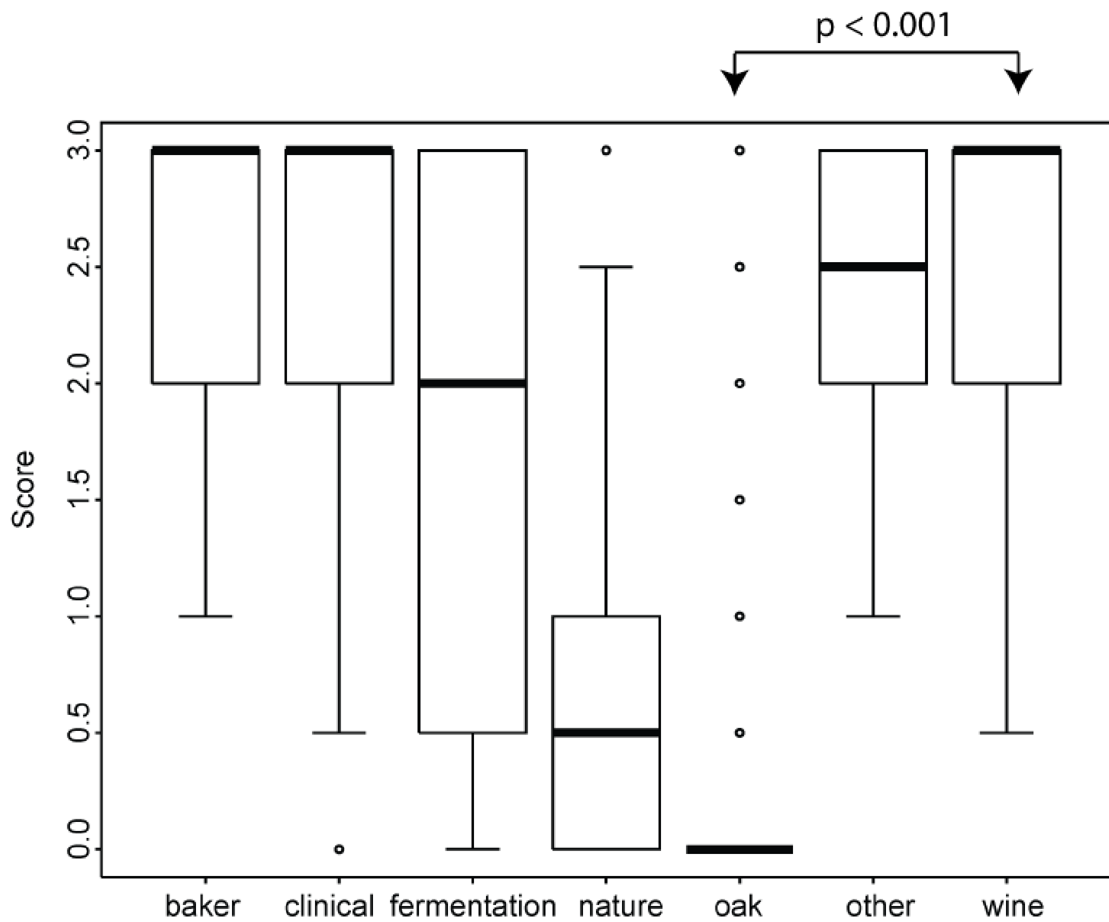


Figure 1-10. Copper sulfate resistance by *S. cerevisiae* strain category.

The mean score for copper sulfate resistance is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots. Scores range from 0 to 3, where 0 is complete inhibition of growth and 3 is no inhibition of growth. Significance was tested for the wine v. oak comparison only.

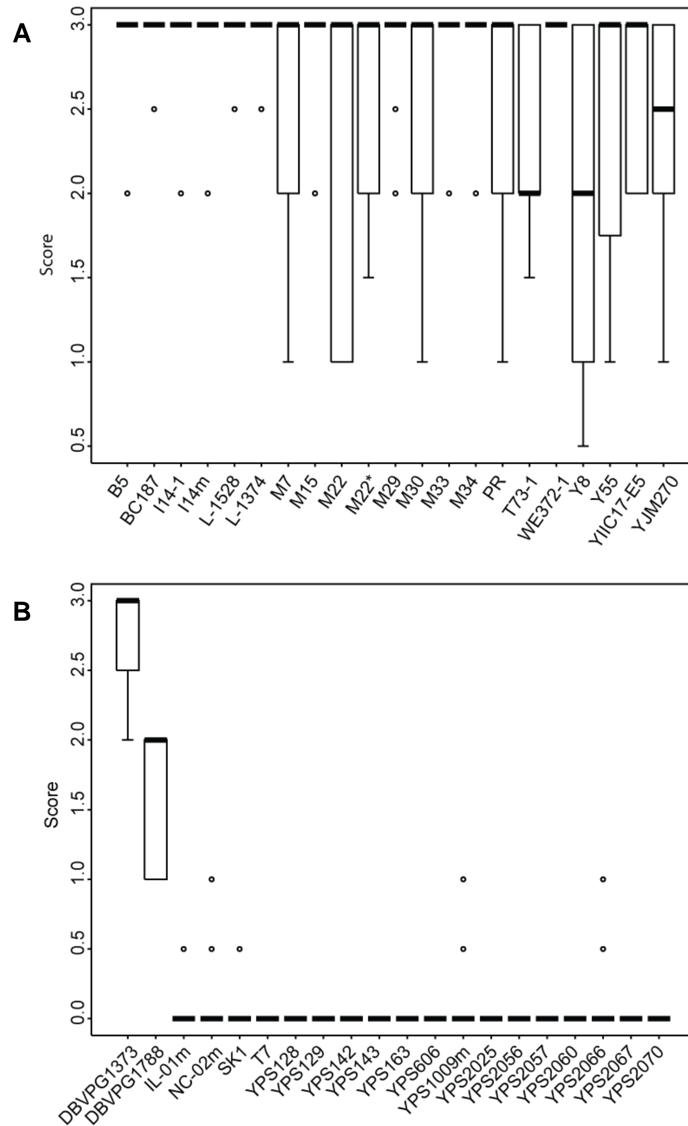


Figure 1-11. Copper sulfate resistance in wine and oak strains of *S. cerevisiae*.

Copper sulfate resistance in (A) wine and (B) oak strains of S*. cerevisiae*. The mean score for copper sulfate resistance is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots. Scores range from 0 to 3, where 0 is complete inhibition of growth and 3 is no inhibition of growth.

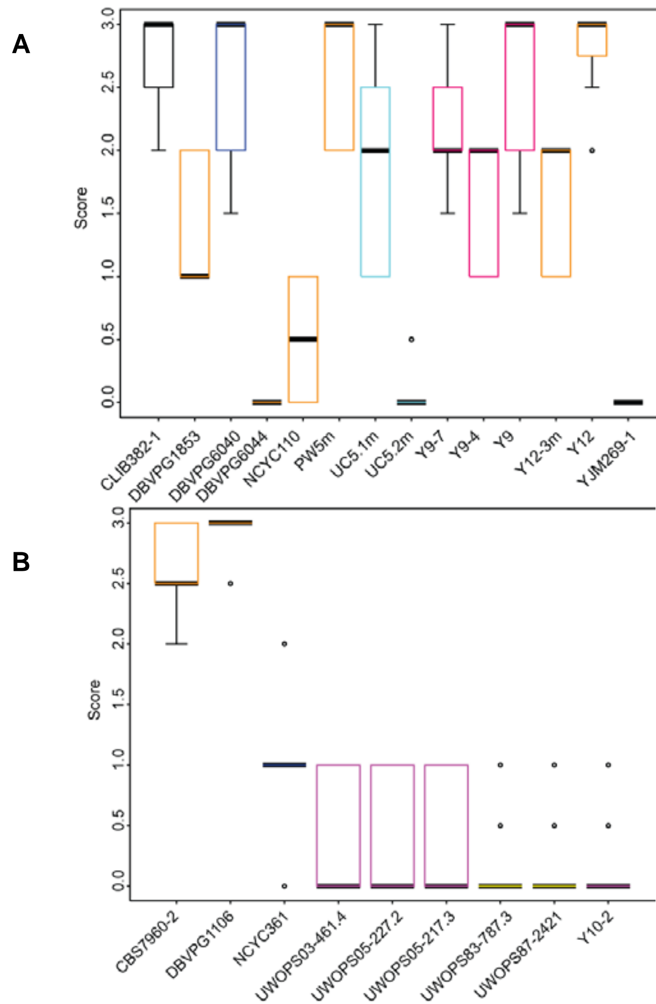


Figure 1-12. Copper sulfate resistance in other strains of *S. cerevisiae*.

Copper sulfate resistance in (A) non-wine fermentation strains (B) other natural strains of S. *cerevisiae*. The mean score for copper sulfate resistance is indicated by a bold line, confidence intervals are represented by boxes. Whiskers indicate the most extreme score no more than 1.5 times the interquartile range, and outliers are represented by dots. Scores range from 0 to 3, where 0 is complete inhibition of growth and 3 is no inhibition of growth. Strains from Africa are shown in orange, strains from Europe in blue, strains from Malaysia and Indonesia in pink, and strains from Hawaii and the Bahamas in yellow.

Chapter 2 : Divergence in wine characteristics produced by wild and domesticated strains of *Saccharomyces cerevisiae*

Introduction

Fermentation of the juices of fruits and starchy vegetables for the production of alcoholic beverages permeates cultures worldwide. Whether for ceremonial, religious, food safety or nutritional reasons, the production of alcohol is embedded in human history (McGovern 2003). The earliest evidence for wine fermentation comes from the molecular analysis of pottery jars that have been dated as far back as 7000 BC (McGovern et al. 2004), and extraction of DNA from ancient wine containers is consistent with the presence of the budding yeast *Saccharomyces cerevisiae* (Cavalieri et al. 2003). The use of *S. cerevisae* for wine production is likely to have occurred for thousands of years and to have preceded its use for bread and beer (Mortimer 2000; McGovern 2003). While *S. cerevisiae* is the dominant species used for wine, beer and bread production worldwide (Mortimer 2000), other *Saccharomyces* species have similar fermentative capabilities but are not as commonly used. For example, two closely related species, *S. bayanus* and *S. paradoxus,* are occasionally associated with wine production (Redzepovic et al. 2002; Naumov et al. 2002; Naumov et al. 2000). Additionally, *S. pastorianus,* a hybrid between *S. cerevisiae* and *S. bayanus,* is used for lager beer fermentation (Nguyen & Gaillardin 2005), and a number of other naturally occurring *Saccharomyces* hybrids have been associated with fermentations (González et al. 2006; Lopandic et al. 2007; Naumova et al. 2005; Groth et al. 1999; de Barros Lopes et al. 2002).

Not all strains of *S. cerevisiae* have been found in association with the production of wine, beer and bread. Wild strains of *S. cerevisiae* have been isolated from a variety of natural sources and have been frequently found in association with oak tree exudates, bark and soil (Sniegowski et al. 2002; Naumov et al. 1998). In comparison, *S. paradoxus*, the sibling species of *S. cerevisiae*, is rarely found in association with vineyards but is frequently found in association with oak trees (Sniegowski et al. 2002; Naumov et al. 1998; Johnson et al. 2004; Naumov et al. 1997; Glushakova et al. 2007; Koufopanou et al. 2006; Redzepovic et al. 2002; Yurkov 2005). A

number of other *Saccharomyces* species have also been found in association with oak trees and soil, and in some instances occur in sympatry with *S. cerevisiae* and *S. paradoxus* (Sniegowski et al. 2002; Naumov et al. 1998; Sampaio & Gonçalves 2008; Naumov et al. 2003).

Strains of *S. cerevisiae* collected from ecologically and geographically diverse sources typically demonstrate genetic differentiation between strains associated with wine production and wild strains associated with natural habitats (Fay & Benavides 2005; Liti et al. 2009; Schacherer et al. 2009; J. L. Legras et al. 2007) . Wine strains form two genetically distinct groups, one associated with vineyards and grape wine production, and the other associated with the production of sake and other rice wines, hereafter simply referred to as 'grape wine' and 'sake' strains, respectively (Fay & Benavides 2005; Liti et al. 2009; Schacherer et al. 2009; Legras et al. 2007). The genetic differentiation between wine and wild strains combined with the observed reduction in genetic diversity within wine strains suggests that wine strains were domesticated from wild *S. cerevisiae* (Fay & Benavides 2005). However, it is not clear whether genetic differentiation of wine and wild strains is a consequence of divergence driven by selection on specific phenotypes relevant to wine making, or a consequence of restricted gene flow following an initial founder event without local adaptation or selection for desirable phenotypes.

One potential explanation for genetic differentiation between grape wine, sake and wild strains is that wine strains were selected for particular flavor characteristics during the development of wine making. Yeast metabolites are known to influence the sensory attributes of wine through the production of esters, higher alcohols, carbonyl compounds, volatile acids, volatile phenols and sulfur compounds (Swiegers & Pretorius 2005). In some cases, it has also been shown that humans can differentiate between wines fermented using different strains of *S. cerevisiae* (Molina et al. 2009; Swiegers et al. 2009; Carrau et al. 2008; Callejon et al. 2010; Wondra & Berovic 2001). Commercial wine yeasts vary in their production of metabolites, often depending on grape juice and fermentation conditions (Molina et al. 2009; Swiegers et al. 2009;

Carrau et al. 2008; Barbosa et al. 2009; Vilela-Moura et al. 2010; Mendes-Ferreira et al. 2009; Loscos et al. 2007; Mateos et al. 2006; Masneuf-Pomarède et al. 2006; Howell et al. 2004; Estevez et al. 2004; Herjavec et al. 2003). Although the contribution of wild *S. cerevisiae* strains to wine aroma and flavor is largely unknown, studies of indigenous vineyard strains of *S. cerevisiae* have revealed variation in their production of wine aroma and flavor metabolites (Callejon et al. 2010; Orlić et al. 2010; Romano et al. 2003; Nurgel et al. 2002; Wondra & Berovic 2001).

The objectives of this study were to determine whether wine and wild yeasts produce wine characteristics that are perceptively different to humans. We tested our hypotheses using both discriminatory and quantitative descriptive sensory analysis of grape wines fermented using wine and wild yeast strains. Our results indicate that humans can distinguish between wines fermented using wine and wild yeast strains and suggest that the elimination of several undesirable wine characteristics produced by wild strains may have played an important role in the phenotypic differentiation between wine and wild strains of *S. cerevisiae*.

Materials and Methods

Yeast strains and fermentation

The *S. cerevisiae* and *S. paradoxus* strains used in this study are described in Table 2-1. Strains W303, N17 and YPS138 were kindly provided by Ed Louis and Gianni Liti (Liti et al. 2009a). Strain PW5 (NPA07) was kindly provided by O. Ezeronye (O U Ezeronye & J-L Legras 2009), and the remainder were described in a previous study (Fay & Benavides 2005). Sterile concentrated grape juice from Vintners Reserve Chardonnay kits (Winexpert, Port Coquitlam, B.C., Canada) was used for all experimental wines. Juice was distributed into sterilized two gallon food grade plastic buckets fitted with airlocks for primary fermentation. Yeast starter cultures were grown individually in 150 ml of sterile juice and used to inoculate 1.25 gallons of juice at a density of 2-5x 10^6 cells/ml. When specific gravity reached 1.010, juice was transferred into 1 gallon glass carboys for secondary fermentation. After fermentation was complete, as measured by absence of CO_2 release and glucose concentrations less than 0.5%, the wine was stabilized using metabisulphite and sorbate, cleared with isinglass and bottled in 375 or 750 ml glass wine bottles with synthetic cork closures. Between two to ten replicate wine fermentations were generated per strain.

Discriminatory sensory evaluation

Discriminatory sensory evaluation was performed through the use of triangle tests (Stone & Sidel 2004) to assess the significance of perceived sensory differences between wines. Trays with three samples of wine were served to participants. Two of the samples on each tray were identical while one was different. Samples were labeled with a randomized three digit number, and participants were asked to circle the number corresponding to the sample that was different. Four to six trays (triangles) were served to each participant during each session. Tests were performed using a balanced block design: triangles, serving orders and positions were balanced to allow for the detection of positional effects.

To test the validity of the method and the performance of participants, we carried out an initial discrimination test using a wine strain (CDB) and a lab strain (W303). Sixty five participants evaluated this triangle six times each (N=390). Participants distinguished between these two wines 42% of the time, significantly more often than the 33% expected by chance (Binomial test, $p < 0.001$). The distribution of the participants' individual scores approximated the binomial distribution, suggesting that judges were equally as skilled at detecting differences. Power analysis was used to determine that 100 evaluations were needed to detect differences. We found no significant difference between serving order (e.g. tray one through six), triangle (e.g. two CDB with one W303 or vice versa), position of the outlier on the tray (e.g. outlier in the first, second or third position left to right), fermentation replicate, or bottle using a Chi-square test. These effects were also not significant during any discriminatory evaluations, with the following exception: during the second experiment (wine and wild *S. cerevisiae* compared to *S. paradoxus*), the proportion of correct decisions for the second and fourth trays were significantly different (Chi-square $p = 0.003$). To test for outliers, each strain was compared to the rest of the strains within the same class (e.g. wine, wild, and *S. paradoxus*) using a Chi-square test.

Quantitative descriptive sensory analysis

A preliminary flavor/taste development session was conducted by Vinquiry, Inc (Sonoma, CA, USA) using six wine experts to evaluate a subset of the wines for aroma and flavor (W303, YPS1000, PW5, N17, K12, and CDB). From this evaluation 28 aromas and 5 flavors, representing eleven classes from the wine aroma wheel were found including: chemical, pungent, floral, fruity, vegetative, caramelized, woody, earthy, microbiological, oxidized and nutty. The results were filtered according to the number of wines in which the attribute was present, the number of panelists who reported the attribute for a given wine, and to ensure adequate representation of different classes of aroma and flavor. Based on these criteria, we chose 12 attributes for descriptive analysis: cabbage (sulfur), wet dog (sulfur), floral, citrus (fruity), tree fruit (fruity), oxidized (acetaldehyde), hay/straw (vegetative), mushroom (earthy), butterscotch

(caramel), acidity (taste), astringency (taste), and trueness to style (taste and aroma). Style trueness was measured relative to a traditional, commercial un-oaked chardonnay. A quantitative descriptive analysis of all 12 attributes was conducted for each wine using an independent panel of six expert judges. Judges underwent three training sessions to review properties of aroma and taste identification as well as variation in aroma/flavor intensity using standard references. Judges scored aroma/flavor attributes based on a numerical scale of 0 to 9 in duplicate for each wine.

Statistical analysis of wine characteristics was carried out using R (R Development Core Team 2009). Each judge's scores for each attribute were centered on the judge's mean score for that attribute and scaled to a standard deviation of 1. Principal component analysis (PCA) and linear descriptive function analysis (LDA) was performed on the transformed data. A stepwise selection criterion was employed to determine which combination of attributes optimized the predictive value for grape wine, oak and *S. paradoxus* strains. Analysis of variance (ANOVA) was conducted on the values for the first two principal components as well as on the transformed scores for each of the twelve aroma/flavor attributes with the model:

$$Z_i = \mu + class_i + strain_i + \varepsilon_i$$

where Z is the quantitative variable (value for the first or second principal component or attribute score), class is the type of strain (grape wine, wild, palm wine, sake, *S. paradoxus* and laboratory), strain is the random strain effect within each class, and Ɛ is the residual error. For attributes that were significantly different for the class or strain term by univariate ANOVA, a post-hoc Tukey's HSD test was performed to determine which classes and/or strains were significantly different from each other. No significant effects were found for tasting session or wine replicate using ANOVA.

For all ANOVAs, the normality of the residual distribution was examined using the Shaprio-Wilks normality test. When residuals were not normally distributed, data transformations

were applied as determined using a boxcox power transformation. The following transformations were applied: for the first principal component scores $y= x+10^{-0.8383}$, for oxidized $y= x+10^{-0.8686}$, for tree fruit $y= x+10^{-1.0303}$, and for citrus $y= x+10^{-1.4747}$. No suitable data transformations were found for butterscotch, trueness to type or floral. For those attributes, permutation tests (N=10,000) were used to generate an empirical F distribution and determine the probability of the observed mean differences between classes and strains. Empirical p values were corrected for multiple testing using the Bonferonni method. Pearson's r rank correlation coefficients were calculated for all possible pairs of attributes.

Chemical Analysis

Chemical analyses were carried out to determine the concentration of the given chemicals in a sample of wine from each of the wine, oak and *S. paradoxus* strains listed in Table 2-1. A basic chemistry panel (free sulfur dioxide, molecular sulfur dioxide, total sulfur dioxide, titratable acidity, pH and volatile acidity), higher alcohol and fusel oil panel (acetaldehyde, ethyl acetate, methanol, 1-propanol, iso butanol, A-amyl alcohol and I-amyl alcohol), and sulfides panel (hydrogen sulfide, methyl mercaptan, ethyl mercaptan, dimethyl sulfide, dimethyl disulfide, diethyl sulfide, diethyl disulfide) was performed by ETS Laboratories (St. Helena, CA, USA). Individual ANOVAs were performed on each chemical to test for significant differences between classes.

Results

Human discrimination of wines fermented using wine yeast and non-wine yeast

A triangle discrimination test was used to determine the ability of humans to discriminate between wines fermented using different yeast strains (see Table 1 for a description of strains). In the discrimination test, participants were presented with three samples of wine, two of which were fermented using the same strain and one of which was fermented using a different strain. Participants were asked to identify (discriminate) the wine sample they thought was different. We hypothesized first that humans can discriminate between wines fermented using strains of the same class (i.e. wine or wild) significantly more often than random. Second, we hypothesized that humans can discriminate between wines fermented using wine strains and those fermented using wild strains significantly more often than when presented with wines fermented using two different strains of the same class (i.e. wine or wild).

To test these hypotheses we measured rates of discrimination between all pairwise combinations of four grape wine strains (CDB, PR, M33 and M8) and four wild strains (YPS163, YPS1000, YPS1009 and YJM454) using the triangle test. For each type of comparisons, the proportion of correct classifications was significantly higher than 33%, the proportion expected by chance, indicating that humans can distinguish between wines produced by different strains regardless of their class, and establishing human perception as a selectable yeast phenotype. The ability of participants to discriminate between wines produced by wild strains was the lowest at 40% (Binomial test, $p = 0.023$), followed by wine strains at 47% ($p < 0.001$), and was highest between wine and wild strains at 56% ($p < 0.001$) (Figure 2-1). No single comparison showed evidence of being an outlier based on the number of correct and incorrect decisions for each comparison (within wine, within wild, and between wine and wild). The magnitude of discrimination (47%) between wine strains was not significantly different from the magnitude of discrimination between oak strains (40%). However, discrimination between wine and wild strains

(56%) was significantly greater than that within either group (Chi-square test, $p = 0.040$ and $p = 0.001$ for comparisons between wine and wild strains to those within wine and within wild, respectively) (Figure 2-1).

A second discrimination experiment was performed to measure the ability of humans to discriminate wines fermented using two *S. paradoxus* strains (N17 and YPS138) with wines fermented using a grape wine strain (CDB) and a wild strain (YPS163) of *S. cerevisiae*. We measured the ability of participants to discriminate between wines fermented using the wine and wild *S. cerevisiae* strains as well as their ability to discriminate between wines fermented using each *S. cerevisiae* strain and each of the two different *S. paradoxus* strains. Strikingly, the wines fermented using wine and wild strains were as different from each other as either was to wines fermented using *S. paradoxus* (Table 2-2). Discrimination of wines fermented using wine and wild *S. cerevisiae* strains was not significantly different from the same pairwise comparison made in the previous experiment, and the ability of humans to discrimination between wines fermented using *S. cerevisiae* strains and *S. paradoxus* was not significantly different for either strain of *S. paradoxus*.

Although most strains of *S. cerevisiae* have been found in association with vineyards and oak trees, strains have also been found in association with other wine fermentations, including sake and palm wine. To test the hypothesis that human perceived differences between wines fermented using grape wine and wild strains is not simply a result of historical use for the production of alcoholic beverages, we used a third discrimination experiment to measure the ability of participants to discriminate between wines fermented using either a palm wine (PW5), sake (K12) grape wine (CDB or M8) or wild (YPS100 or YPS1009) strain of *S. cerevisiae*. Subjects were unable to distinguish between wine fermented using the palm wine strain and wild strains, but were able to distinguish wine fermented using the palm wine strain and grape wine strains. The ability of participants to distinguish between wine fermented using the palm wine

strain and the grape wine strains was similar to the degree of differentiation observed when subjects discriminated between wines fermented using grape wine and wild strains (Table 2-3). In contrast, the wine fermented using the sake strain was significantly different from that fermented using the wild strains, but not significantly different from the wines fermented using grape wine and palm wine strains.

Quantification of sensory attributes

The results of our discrimination tests demonstrate that *S. cerevisiae* strains produce wines that can be discriminated by human perception. However, discrimination testing does not allow us to quantify differentiation for specific attributes. To determine which sensory attributes contribute to the perceived sensory differences between wines fermented using different strains, the same wines used in our discriminatory analysis were used for quantitative descriptive analysis (see Table 2-1 for a description of strains.) As described in Materials and Methods, twelve attributes (cabbage, wet dog, oxidized, mushroom, astringency, acidity, hay/straw, butterscotch, tree fruit, trueness to type, citrus and floral) were chosen for analysis and a trained panel of experts evaluated each wine for those twelve attributes using a quantitative scale.

Principal component analysis (PCA) was used to evaluate differences in scores for the twelve attributes. The first two principal components together explained 35.4% of the variance. The mean and standard error of the first two principal components for each strain is shown in Figure 2-2. The first principal component axis, which explains 23.7% of the variation, was loaded most heavily by cabbage, wet dog, oxidized and mushroom attributes in the negative direction, and by butterscotch, tree fruit, trueness to type, citrus and floral attributes in the positive direction (

Table 2-4. Significant correlations between descriptive and chemical attributes.

comaprison type	attribute1	attribute2	correlation coefficient [1]	P value
descriptive	cabbage	citrus	-0.69	0.029
descriptive	cabbage	hay/straw	0.67	0.035
descriptive	cabbage	mushroom	0.78	0.007
descriptive	cabbage	oxidized	0.81	0.005
descriptive	cabbage	trueness	-0.65	0.044
descriptive	citrus	mushroom	-0.76	0.011
descriptive	citrus	trueness	0.64	0.045
descriptive	floral	trueness	0.74	0.014
descriptive	hay/straw	mushroom	0.79	0.006
descriptive	hay/straw	trueness	-0.62	0.056
descriptive	mushroom	trueness	-0.78	0.007
descriptive	oxidized	hay/straw	0.66	0.039
descriptive	oxidized	mushroom	0.7	0.025
descriptive	oxidized	trueness	-0.72	0.020
descriptive	tree fruit	astringency	-0.67	0.036
descriptive	wet dog	mushroom	0.65	0.041
chemical	Aamyl	acidity	0.73	0.017
chemical	acetaldehyde	phenyl.ethanol	0.72	0.020
chemical	butanol	ethyl octanoate	-0.86	0.004
chemical	butanol	phenyl ethanol	0.76	0.010
chemical	butanol	VA	-0.66	0.039
chemical	ethyl acetate	isoamyl acetate	0.7	0.025
chemical	ethyl acetate	phenyl ethanol	-0.82	0.004
chemical	ethyl acetate	VA	0.98	0.000
chemical	ethyl hexanoate	ethyl octanoate	0.74	0.014

chemical	ethyl isobutyrate	isoamyl acetate	0.69	0.027
chemical	ethyl isobutyrate	isobutyl acetate	0.68	0.030
chemical	ethyl isobutyrate	total so2	-0.66	0.040
chemical	ethyl propionate	isoamyl alcohol	-0.63	0.050
chemical	ethyl propionate	isobutanol	-0.88	0.001
chemical	ethyl-2-methylbutyrate	butanol	0.82	0.004
chemical	ethyl-2-methylbutyrate	dimethyl sulfide	-0.65	0.044
chemical	ethyl-2-methylbutyrate	ethyl octanoate	-0.68	0.030
chemical	ethyl-3-methylbutyrate	isoamyl alcohol	0.65	0.042
chemical	free so2	molecular so2	0.96	< 0.001
chemical	free so2	total so2	0.75	0.013
chemical	free so2	trueness	0.66	0.037
chemical	isobutanol	dimethyl sulfide	0.66	0.038
chemical	isobutanol	isoamyl alcohol	0.74	0.015
chemical	isobutanol	pH	-0.65	0.044
chemical	isobutyl acetate	dimethyl sulfide	0.8	0.005
chemical	isobutyl acetate	ethyl -2-methylbutyrate	-0.65	0.041
chemical	molecular so2	total so2	0.67	0.035
chemical	molecular so2	trueness	0.77	0.009
chemical	pH	butterscotch	-0.87	0.001
chemical	phenyl ethanol	VA	-0.85	0.002
chemical	propanol	pH	-0.64	0.049
chemical	propanol	TA	0.65	0.042
chemical	total so2	TA	-0.7	0.024
chemical	VA	citrus	0.64	0.046
chemical and descriptive	acetaldehyde	citrus	-0.63	0.050
chemical and descriptive	butanol	oxidized	0.67	0.036

chemical and descriptive	dimethyl sulfide	tree fruit	-0.67	0.033
chemical and descriptive	ethyl acetate	astringency	0.65	0.042
chemical and descriptive	ethyl hexanoate	floral	0.65	0.040
chemical and descriptive	ethyl isobutyrate	tree fruit	-0.76	0.011
chemical and descriptive	ethyl octanoate	floral	0.73	0.017
chemical and descriptive	ethyl-2-methylbutyrate	astringency	-0.74	0.015
chemical and descriptive	ethyl-2-methylbutyrate	citrus	-0.69	0.027
chemical and descriptive	isobutyl acetate	astringency	0.7	0.023
chemical and descriptive	isobutyl acetate	tree fruit	-0.86	0.001
chemical and descriptive	propanol	acidity	0.73	0.016
chemical and descriptive	propanol	wet dog	-0.7	0.023
chemical and descriptive	TA	acidity	0.93	< 0.001

[1] correlation coefficient is Pearson's r rank correlation coefficient

Table 2-5). The grape wine strains along with the lab strain W303, which is closely related (genetically) to grape wine strains (Winzeler et al. 2003; Rothstein 1977; Rothstein et al. 1977), are associated with positive values on the first principal component axis, while wild, palm wine, sake and *S. paradoxus* strains are associated with negative values on this axis. The second principal component axis, which explains 11.7% of the variation, was loaded most heavily by astringency, acidity, wet dog, floral and cabbage attributes in the negative direction and by butterscotch, hay/straw and mushroom attributes in the positive direction (Figure 2-2). Significant correlations were found between sensory attributes, supporting the relationships inferred through PCA (Table 2-4). Linear discriminant analysis (LDA) was performed to determine the predictive value of the twelve attributes. Overall, the linear descriptive analysis was able to correctly 51% of observations (67% for oak, 65% for grape wine, 36% for sake, 33% for paradoxus, 27% for lab, and 9% for palm wine strains). In agreement with PCA, the combination of variables that optimized the predictive value for grape wine (67%), oak (70%), and *S. paradoxus* (25%) included wet dog, citrus and floral.

To determine if there was a significant difference between classes of strains for principal components, we performed ANOVA on the principal components scores for each axis, as described in Materials and Methods. The class term, grape wine, wild, palm wine, sake, and *S. paradoxus* was significant for the first principal component ($p < 0.001$), but not for the second principal component ($p=0.124$). The strain term, which represents random strain effect within each class, was not significant for either of the first two principal components (p = 0. 816 and p = 0.591 respectively). A post-hoc Tukey's HSD test revealed that wines fermented using grape wine strains are significantly different from those fermented using wild, palm wine and *S. paradoxus* strains for the first principal component (Table 2-6) but not significantly different from sake or lab strains. Despite some levels of discrimination between sake, palm, wild and *S. paradoxus* strains (Table 2-2 and Table 2-3), these classes are not significantly different from one

another for wine attributes captured by the first principle component, which readily distinguishes grape wine strains from other strains of *S. cerevisiae* and *S. paradoxus*. Similarly, linear discriminate analysis is able to predict class member ship for each wine replicate 65% and 67% of the time for wine and wild strains, respectively, but only 27% of the time, on average, for the other classes.

In agreement with the PCA analysis, wine attributes that are significantly different between classes by ANOVA include the undesirable attributes cabbage, wet dog, oxidized and mushroom, and the desirable attributes citrus, and floral (Table 2-6). Differences in the mean class scores for these attributes are depicted in Figure 2-3. Wines fermented using wild, palm wine, sake and *S. paradoxus* strains scored higher for undesirable attributes whereas wines fermented using grape wine strains and the lab strain scored higher for desirable attributes. Post-hoc Tukey's HSD tests revealed that cabbage, wet dog, citrus and floral attributes differentiated between grape wine strains and other strains, but not between any classes of non-grape wine strains (Table 2-6). Mushroom aroma was variable between many classes, differentiating grape wine strains from wild *S. cerevisiae* and *S. paradoxus* strains, but also differentiating sake strains from wild *S. cerevisiae* and *S. paradoxus* strains (Table 2-6). Oxidized aroma did not differentiate wine strains from any other class of strains (Table 2-6). The only significant differences between strains within a class was between two grape wine strains, M33 and CDB ($p = 0.044$) for mushroom aroma. The results of this quantitative analysis support our results of the initial discrimination tests, showing that human perceived differences between wines produced by grape wine strains and other classes of strains are significantly greater than differences within each class.

Chemical Analysis

A final experiment was conducted to test if the flavor and aroma attributes that contribute to the ability of humans to discriminate between wines fermented using wine strains and those

fermented using wild strains and *S. paradoxus* strains are due to differences in chemical concentrations produced during fermentation. The chemical composition of the wines was evaluated for 14 commonly produced yeast metabolites associated with wine flavor. No significant difference between classes of strains was observed.

Discussion

Many *Saccharomyces* yeasts preferentially ferment sugar into alcohol in the presence of oxygen despite the higher energy yield of respiration (de Deken 1966). Yet, grape wine is predominantly produced using a genetically homogeneous subgroup of *S. cerevisiae* strains. We initiated this study to determine if an association could be drawn between the genetic bottleneck observed for traditional wine making strains of *S. cerevisiae* and the production of quality grape wine.

We have established human perception as a selectable yeast phenotype, and also demonstrated that divergence in wine aroma and flavor, a putative domestication phenotype, is coupled with the genetic divergence between these two groups. Furthermore, the magnitude of divergence between grape wine and wild strains of *S. cerevisiae* compared to *S. paradoxus* suggests rapid enological divergence of the wine strains from their wild ancestors. The enrichment of several desirable wine attributes and/or the elimination of several undesirable wine attributes imply that wine strains may have been intentionally or inadvertently domesticated for the production of quality grape wine.

Grape wine and non-grape wine strains are differentiated by both desirable and undesirable sensory attributes. We found that the sulfurous attributes cabbage and wet dog make a major contribution to differences between wines produced by grape wine strains and those produced by wild strains of both *S. cerevisiae* and *S. paradoxus*. Citrus and floral attributes make similar contributions to the difference between grape wine and wild *S. cerevisiae* strains. However, it is possible that these desirable sensory attributes were present in wines produced by wild strains at levels similar to those produced by wine strains, but were detected at a lower level by humans due to the masking effect of sulfurous attributes. The oxidized aroma, which is associated with acetaldehyde, also contributed to the difference between wine and wild strains, as did the mushroom aroma, but the latter was more heavily loaded on the second principle

component. Although the second principal component was not significantly different among classes of yeast strains, it tended to differentiate wild strains of *S. cerevisiae* and *S. paradoxus* (Figure 2-2). The attributes, astringency, acidity, hay/straw and butterscotch were also more heavily loaded on the second principle component but did not make significant contributions to differences between classes of yeast strains.

Loss of undesirable traits is common in domesticated species, for example loss of seed shattering in crop species (Doebley et al. 2006), and loss of aggression in dogs (Lindberg et al. 2007). Often, undesirable traits persist in domesticated species. For example, in rice (Londo & Schaal 2007) weedy traits include increased competitive ability and variable levels of seed shattering (Gealy 2005). Persistence of undesirable sulfur compounds during wine production could be caused in part by the presence of wild *S. cerevisiae* strains that have migrated into vineyards or vineyard strains with mixed backgrounds. Vineyard strains with mixed backgrounds have been identified in a number of studies (Gangl et al. 2009; González et al. 2006; Lopandic et al. 2007; Naumova et al. 2005; Groth et al. 1999; de Barros Lopes et al. 2002).

Loss of sulfur flavors is one of a growing number of differences between wine and wild strains of *S. cerevisiae*. The selective elimination of various undesirable wine attributes, particularly pungent sulfur-containing compounds, is consistent with current practice in the wine industry. The production of hydrogen sulfide, thiols (mercaptans) and related sulfur-containing compounds during fermentation is a major problem in wine production (Swiegers & Pretorius 2007). Commercial wine strains of *S. cerevisiae* (Swiegers & Pretorius 2007), as well as *S. bayanus* (Ugliano et al. 2009) differ in their production of sulfur compounds, which is often influenced by fermentation conditions and grape juice composition. Copper sulfate can be added to wines to reduce sulfur-related defects by chelating thiol (-SH) containing compounds (Swiegers & Pretorius 2007), but in the United States copper sulfate must remain below 0.5 parts copper per million parts wine . Other differences that have been documented for grape wine and wild strains

of *S. cerevisiae* include growth and fermentation parameters, (Spor et al. 2009) resistance to copper (Fay et al. 2004) and sulfite (Park & Bakalinsky 2000), two chemicals related to vineyards and wine production, and freezing (Will et al. 2010). Wine strains of *S. cerevisiae* also exhibit a substantial decrease in sporulation efficiency that is also likely to be associated with domestication (Gerke et al. 2006).

We found the largest differences in wine attributes between wines produced by grape wine and wild strains of *S. cerevisiae,* which was equal to the differences between wine strains of *S. cerevisiae* and *S. paradoxus*. This degree of phenotypic divergence within *S. cerevisiae* is quite high given that the genetic divergence between *S. cerevisiae* and *S. paradoxus* is 25 times higher than that between a wine and wild strain of *S. cerevisiae*, as measured by the synonymous substitution rate (Doniger et al. 2008). Enological divergence among grape wine strains was similar to that among wild strains, despite the latter showing 3.6 times more genetic diversity (Fay & Benavides 2005). However, this pattern is consistent with previous studies which revealed substantial variation in stress response (Kvitek et al. 2008) and growth and fermentation parameters (Spor et al. 2009) among grape wine strains compared to other *S. cerevisiae* strains. In addition, the increased phenotypic diversity combined with a reduction in genetic diversity is consistent with other domesticated organisms (e.g. varietal differences in crops (Doebley et al. 2006)). In the case of wine flavor and aroma characteristics, the large amount of phenotypic diversity within wine strains could be due to selection for different desirable attributes, against undesirable attributes, or for attributes specific to different grape varieties.

The smaller enological differences between the sake, palm wine and wild strains is not surprising. Relative to wild strains of *S. cerevisiae*, strains of *S. paradoxus* scored higher for undesirable characteristics, except for wet dog, and lower for desirable characteristics (Figure 2-3). The lack of divergence between wild *S. cerevisiae* and *S. paradoxus* strains could be a simple consequence of constraints placed upon them by their shared environment. Sake and

palm wine strains produce grape wine with attributes that are similar to wild strains. However, the low levels of differentiation among these groups could be due to the measurement of grape wine characteristics rather than sake or palm wine specific characteristics. Consistent with this possibility, sake strains exhibit a number of sake fermentation flavor characteristics that differ from those generated by a laboratory strain (Katou et al. 2009; Katou et al. 2008). Thus, differentiation between wine and wild strains does not appear to be simply correlated with use in alcohol production.

Measurement of chemical differences among wines revealed a number of quantitative differences, but none that significantly differentiated wine and wild strains. Some of the measured compounds could contribute to cabbage odor (methyl mercaptan, dimethyl sulfide, dimethyl disulfide) while others could contribute a fruity odor (acetaldehyde, ethyl acetate). However, hundreds of compounds are known to influence wine flavor and aroma (Swiegers & Pretorius 2005), many of which could contribute to attributes that distinguish wine and wild strains. Sulfides and mercaptans are often described as having an aroma of cooked cabbage (Swiegers & Pretorius 2005). Volatile sulfur compounds (Fan 2005) and branched or complex aldehydes have been associated with wet dog odor (Young et al. 2002). Both 'mushroom' and 'oxidized' have been used to describe metabolites in the Ehrlich pathway. The Erhlich pathway oxidizes and reduces fusel aldehydes into fusel alcohols and fusel acids, respectively (Hazelwood et al. 2008). Aldehydes can generate an oxidized flavor and the Ehrlich pathway can convert methionine to methional, which has a cabbage/cauliflower odor, and subsequently to 3-methylthiopropyl acetate, which has a mushroom/garlic odor (Swiegers & Pretorius 2005; Etschmann et al. 2008). The Ehrlich pathway is especially active when amino acids are the sole nitrogen source, as is the case during wine fermentation, and genes within the Ehrlich pathway are differentially regulated during wine fermentation (Backhus et al. 2001).

Previous studies have revealed that the use of different commercial (Molina et al. 2009; Swiegers et al. 2009; Carrau et al. 2008; Barbosa et al. 2009; Vilela-Moura et al. 2010; Mendes-Ferreira et al. 2009; Loscos et al. 2007; Mateos et al. 2006; Masneuf-Pomarède et al. 2006; Howell et al. 2004; Estevez et al. 2004; Herjavec et al. 2003) or indigenous (Callejon et al. 2010; Orlić et al. 2010; Romano et al. 2003; Nurgel et al. 2002; Wondra & Berovic 2001) strains of *S. cerevisiae* can affect the production of flavor and aroma compounds, and in some cases have also shown that humans can detect those differences (Molina et al. 2009; Swiegers et al. 2009; Carrau et al. 2008; Callejon et al. 2010; Wondra & Berovic 2001). Determining the genetic contribution of *S. cerevisiae* to wine flavor and aroma characteristics is challenging (Bisson & Karpel 2010). Not only do yeast metabolites interact to form certain flavors and aromas, but grape composition and fermentation conditions affect *S. cerevisiae* metabolite production (Bisson & Karpel 2010). Despite this difficulty, several examples of genes underlying wine flavor and aroma differences have been identified. Genes involved in the production of fusel oils (higher alcohols), volatile organic acids, esters, sulfur-containing volatiles, carbonyl compounds, volatile aglycones and cys-conjugates have been identified (reviewed in (Bisson & Karpel 2010)). Genetic variation at these loci between wine and wild strains of *S. cerevisiae* may account for some of the observed differences in wine flavor and aroma, but further work will be needed to dissect the genetic basis for the sensory differentiation we observed between wine and wild strains of *S. cerevisiae*.

While most differences in wine quality are attributable to grapes, which differ by variety, location and year, there is a growing body of evidence that wine quality is also influenced by the yeast (Swiegers & S. Pretorius 2005; Bisson & Karpel 2010), specifically in the production of undesirable sulfur aromas (Swiegers & Pretorius 2007; Bisson & Karpel 2010). Our results show that wild *S. cerevisiae* may contribute several undesirable wine characteristics, resulting in low quality wine. If wild populations of *S. cerevisiae* are present in vineyards during grape harvesting, they may contribute to problem fermentations. Selection against low quality wine may thus

provide an explanation for the strong degree of genetic differentiation between wine and wild yeast strains. By identifying the genetic determinants of undesirable attributes present in wild yeast populations, it may be possible to further improve existing commercial wine strains as well as better understand the origins and evolution of wine strains.

Table 2-1. Yeast strains used in this study.

Strain	Class	Origin
W303	laboratory	related to the laboratory strain S288c
Cotes des Blancs (CDB)	grape wine	Commercial wine strain originating from Germany
Pasteur Red (PR)	grape wine	Commercial wine strain originating from France
M33	grape wine	Vineyard, Italy
M8	grape wine	Vineyard, Italy, 1993
YPS163	wild	Oak exudate, Pennsylvania, United States, 1999
YPS1000	wild	Oak exudate, New Jersey, United States, 2000
YPS1009	wild	Oak exudate, New Jersey, United States, 2000
YJM454	wild	Clinical isolate (blood), United States, pre-1994
PW5	palm wine	Raphia Palm tree, Aba, Abia state, Nigeria, 2002
AKU-4011 (K12)	sake	Commercial Sake wine, Japan
N17	*S. paradoxus*	Oak exudate, Tartarstan, Russia
YPS138	*S. paradoxus*	Oak soil, Pennsylvania, United States, 1999

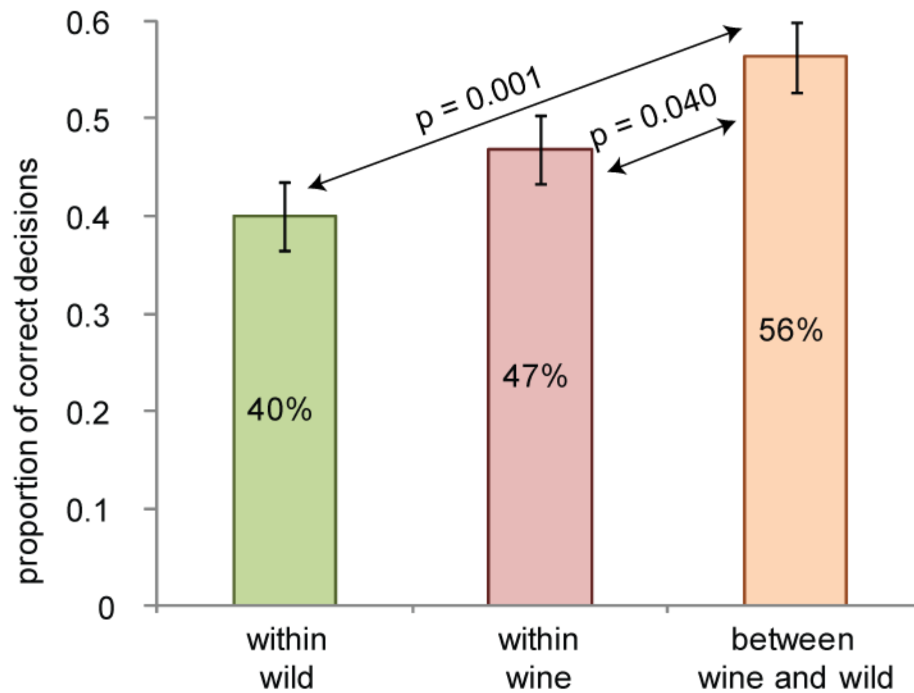


Figure 2-1. Humans can discriminate between wines fermented using different strains of *S. cerevisiae*.

The proportion of correct decisions for the triangle discrimination test is shown for grape wine (CDB, M33, M8 and PR) and wild (YPS163, YPS1009, YPS1000 and YJM454) strains of *S. cerevisiae.*

Table 2-2. Discrimination of wines produced by *S. cerevisiae* and *S. paradoxus* strains.

comparison	correct	N	p
wine v *S. paradoxus*	0.46	96	0.007
wild v *S. paradoxus*	0.45	96	0.004
wine v wild	0.43	96	0.021

Strains are CDB (grape wine), YPS163 (wild), N17 and YPS138 (*S. paradoxus).*

Table 2-3. Discrimination of wines produced by palm wine, sake, grape wine and wild strains of *S. cerevisiae*.

comparison	**correct**	**N**	**p**
palm wine - wild	0.36	39	0.301
palm wine – grape wine	0.52	48	0.002
sake - wild	0.50	48	0.006
sake – grape wine	0.42	48	0.086
palm wine - sake	0.46	24	0.068

Strains are PW5 (palm wine), K12 (sake), CDB and M8 (wine), YPS100 and YPS1009 (wild).

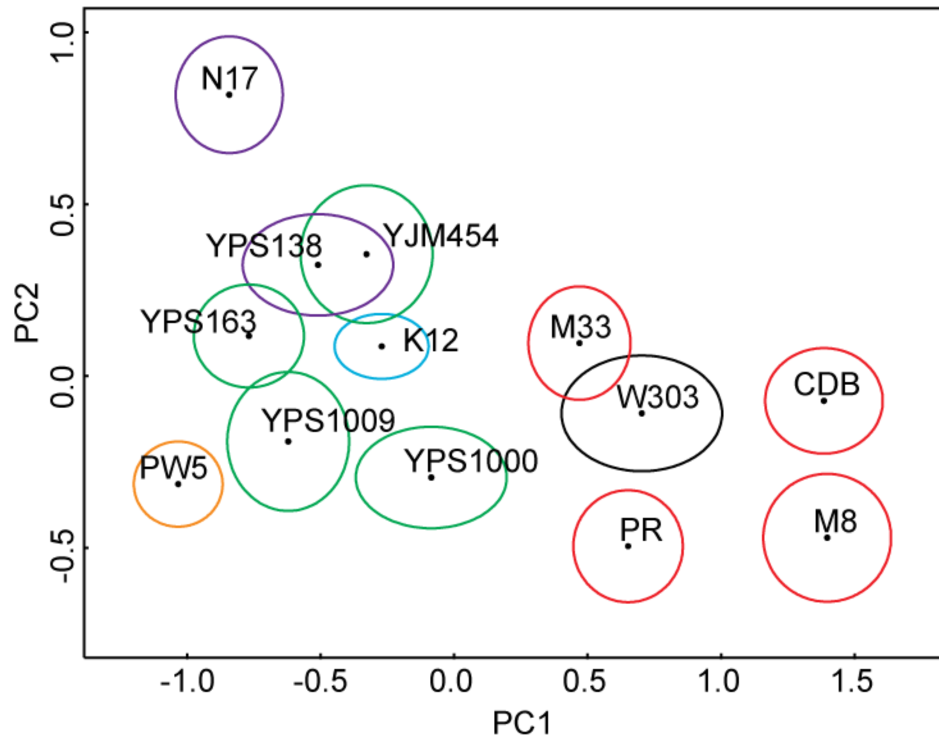


Figure 2-2. Grape wine and non-grape wine strains of *S. cerevisiae* are differentiated for wine flavor and aroma attributes.

Strain means (points) and standard error (ellipses) of the first two principal components for 12 wine attributes Grape wine strains are in red, wild in green, *S. paradoxus* in purple, sake in light blue, palm wine in orange, and laboratory in black.

Table 2-4. Significant correlations between descriptive and chemical attributes.

comaprison type	attribute1	attribute2	correlation coefficient [1]	P value
descriptive	cabbage	citrus	-0.69	0.029
descriptive	cabbage	hay/straw	0.67	0.035
descriptive	cabbage	mushroom	0.78	0.007
descriptive	cabbage	oxidized	0.81	0.005
descriptive	cabbage	trueness	-0.65	0.044
descriptive	citrus	mushroom	-0.76	0.011
descriptive	citrus	trueness	0.64	0.045
descriptive	floral	trueness	0.74	0.014
descriptive	hay/straw	mushroom	0.79	0.006
descriptive	hay/straw	trueness	-0.62	0.056
descriptive	mushroom	trueness	-0.78	0.007
descriptive	oxidized	hay/straw	0.66	0.039
descriptive	oxidized	mushroom	0.7	0.025
descriptive	oxidized	trueness	-0.72	0.020
descriptive	tree fruit	astringency	-0.67	0.036
descriptive	wet dog	mushroom	0.65	0.041
chemical	Aamyl	acidity	0.73	0.017
chemical	acetaldehyde	phenyl.ethanol	0.72	0.020
chemical	butanol	ethyl octanoate	-0.86	0.004
chemical	butanol	phenyl ethanol	0.76	0.010
chemical	butanol	VA	-0.66	0.039
chemical	ethyl acetate	isoamyl acetate	0.7	0.025
chemical	ethyl acetate	phenyl ethanol	-0.82	0.004
chemical	ethyl acetate	VA	0.98	0.000
chemical	ethyl hexanoate	ethyl octanoate	0.74	0.014

chemical	ethyl isobutyrate	isoamyl acetate	0.69	0.027
chemical	ethyl isobutyrate	isobutyl acetate	0.68	0.030
chemical	ethyl isobutyrate	total so2	-0.66	0.040
chemical	ethyl propionate	isoamyl alcohol	-0.63	0.050
chemical	ethyl propionate	isobutanol	-0.88	0.001
chemical	ethyl-2-methylbutyrate	butanol	0.82	0.004
chemical	ethyl-2-methylbutyrate	dimethyl sulfide	-0.65	0.044
chemical	ethyl-2-methylbutyrate	ethyl octanoate	-0.68	0.030
chemical	ethyl-3-methylbutyrate	isoamyl alcohol	0.65	0.042
chemical	free so2	molecular so2	0.96	< 0.001
chemical	free so2	total so2	0.75	0.013
chemical	free so2	trueness	0.66	0.037
chemical	isobutanol	dimethyl sulfide	0.66	0.038
chemical	isobutanol	isoamyl alcohol	0.74	0.015
chemical	isobutanol	pH	-0.65	0.044
chemical	isobutyl acetate	dimethyl sulfide	0.8	0.005
chemical	isobutyl acetate	ethyl -2-methylbutyrate	-0.65	0.041
chemical	molecular so2	total so2	0.67	0.035
chemical	molecular so2	trueness	0.77	0.009
chemical	pH	butterscotch	-0.87	0.001
chemical	phenyl ethanol	VA	-0.85	0.002
chemical	propanol	pH	-0.64	0.049
chemical	propanol	TA	0.65	0.042
chemical	total so2	TA	-0.7	0.024
chemical	VA	citrus	0.64	0.046
chemical and descriptive	acetaldehyde	citrus	-0.63	0.050
chemical and descriptive	butanol	oxidized	0.67	0.036

chemical and descriptive	dimethyl sulfide	tree fruit	-0.67	0.033
chemical and descriptive	ethyl acetate	astringency	0.65	0.042
chemical and descriptive	ethyl hexanoate	floral	0.65	0.040
chemical and descriptive	ethyl isobutyrate	tree fruit	-0.76	0.011
chemical and descriptive	ethyl octanoate	floral	0.73	0.017
chemical and descriptive	ethyl-2-methylbutyrate	astringency	-0.74	0.015
chemical and descriptive	ethyl-2-methylbutyrate	citrus	-0.69	0.027
chemical and descriptive	isobutyl acetate	astringency	0.7	0.023
chemical and descriptive	isobutyl acetate	tree fruit	-0.86	0.001
chemical and descriptive	propanol	acidity	0.73	0.016
chemical and descriptive	propanol	wet dog	-0.7	0.023
chemical and descriptive	TA	acidity	0.93	< 0.001

[1] correlation coefficient is Pearson's r rank correlation coefficient

Table 2-5. PCA loadings and ANOVA p values for individual attributes.

	PCA loadings		attribute p value[1]	
attribute	**PC1**	**PC2**	**class**	**strain**
cabbage	-0.435	-0.143	**< 0.000**	**0.013**
wet dog	-0.370	-0.256	**0.013**	0.599
oxidized	-0.311	0.000	**0.006**	0.118
mushroom	-0.209	0.446	**< 0.000**	**0.017**
astringency	0.000	-0.482	0.122	0.739
acidity	0.000	-0.274	0.110	**0.009**
hay/straw	0.000	0.444	0.112	0.086
butterscotch	0.130	0.419	0.186	0.476
tree fruit	0.229	0.000	0.148	0.328
trueness	0.355	0.000	0.213	0.515
citrus	0.372	0.000	**0.000**	0.156
floral	0.436	-0.152	**0.006**	**0.016**

[1] p values are from ANOVA except for butterscotch and trueness to type which are from the smallest Bonferroni corrected p value for any pairwise comparison obtained from a permutation test of mean differences between classes or strains

Table 2-6. Tukey's HSD (Honestly Significant Difference) p values.

Class1	Class2	PC1[1]	cabbage[1]	wet dog[1]	oxidized[1]	mush room[1]	citrus[1]	floral[2]
grape wine	lab	0.974	0.675	0.829	0.751	1.000	0.985	10.833
grape wine	wild	**0.000**	**0.003**	**0.008**	0.399	**0.000**	**0.001**	0.951
grape wine	palm	**0.001**	**0.000**	0.304	0.183	0.883	0.120	0.707
wine	paradoxus	**0.000**	**0.002**	0.840	0.084	**0.000**	**0.001**	**0.006**
wine	sake	0.223	**0.021**	1.000	0.701	1.000	**0.012**	3.414
sake	lab	0.844	0.759	0.947	0.243	1.000	0.304	7.686
sake	wild	0.990	0.971	0.264	1.000	**0.041**	0.975	14.604
sake	palm	0.731	0.935	0.613	0.980	0.921	0.986	8.373
sake	paradoxus	0.905	1.000	0.973	0.992	**0.015**	1.000	3.209
paradoxus	lab	0.156	0.732	1.000	0.028	**0.014**	0.190	0.618
paradoxus	wild	0.984	0.971	0.534	0.875	0.957	0.948	1.023
paradoxus	palm	0.991	0.812	0.897	1.000	0.256	0.981	8.759
palm	lab	0.110	0.207	0.983	**0.048**	0.894	0.705	3.083
palm	wild	0.858	0.367	1.000	0.869	0.522	1.000	6.509
wild	lab	0.283	0.939	0.907	0.123	**0.038**	0.425	6.306

[1] p value from Tukey’s HSD

[2] Bonferroni corrected p value from a permutation test of mean differences between classes

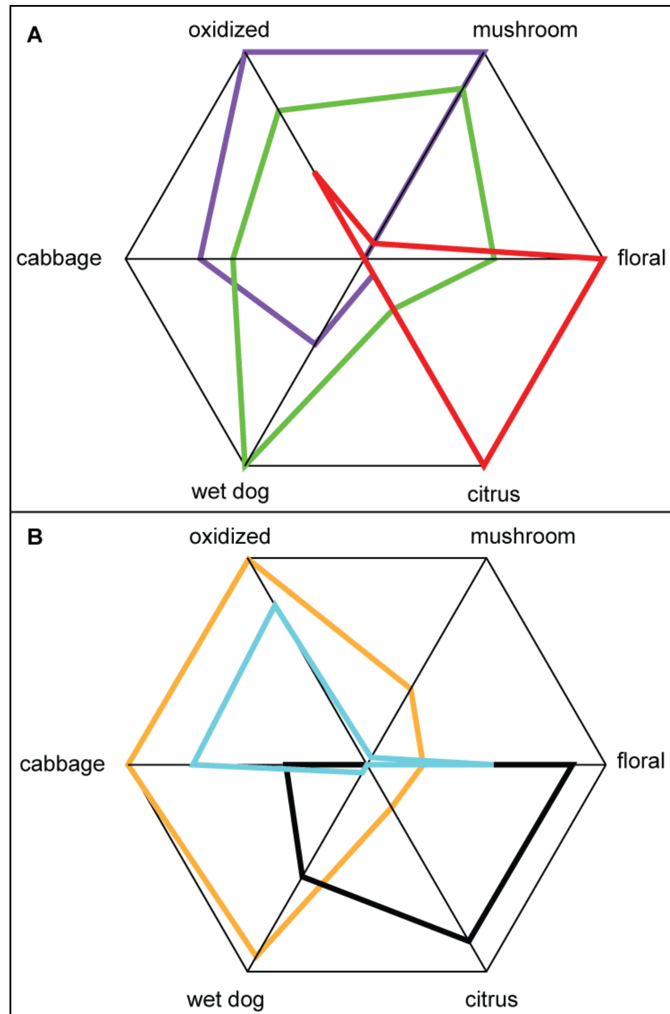


Figure 2-3. Grape wine strains produce desirable wine attributes and wild strains produced undesirable wine attributes.

(A) Class means for grape wine strains (red), wild strains (green), and *S. paradoxus* strains (purple) and (B) means for the palm wine (orange), sake (blue) and laboratory (black) strains are shown for each of the six wine attributes that distinguish grape wine strains from other non-grapewine strains. Means were scaled from 0 (center) to 1 (spokes), where 0 represents the lowest mean score and 1 represents the highest mean score for any class.

Chapter 3 : Isolation and Characterization of *Saccharomyces* species from Vineyard and Non-Vineyard locations

Introduction

The budding yeast *Saccharomyces cerevisiae* is one of the most widely studied fungal species and is primarily associated with the production of bread, beer, and wine (Liti et al 2009, Fay & Benavides 2005, Legras et al 2007, Aa et al. 2006). Given its close and historical association with humans, *S. cerevisiae* has been thought to represent a fully domesticated species (Mortimer 2000). In contrast, *S. cerevisiae* has also been isolated from more natural habitats, suggesting that *S. cerevisiae* has a role in ecological processes, and that wild populations not associated with humans persist in the environment. *Saccharomyces paradoxus*, the most closely related sister species to *S. cerevisiae*, is also found in association with the same natural habitats (e.g. tree bark, soil) as the 'wild' *S. cerevisiae* (Naumov et al. 1998; Sniegowski et al. 2002; Sampaio & Gonçalves 2008), but is not typically associated with human activity.

There have been many collections of populations of *S. cerevisiae* from fermentations and from vineyards (Goddard et al. 2010; Valero et al. 2007; Csoma et al. 2010; Orlić et al. 2010; Versavaud et al. 1995; Lopandic et al. 2007; Redzepovic et al. 2002; Pramateftaki et al. 2000; Mortimer & Polsinelli 1999; Blanco et al. 2010; Schuller & Casal 2006; Garijo et al. 2008; Török et al. 1996; Romano et al. 2003; Li et al. 2010) as well as collections of natural populations (Aa et al. 2006; Sampaio & Gonçalves 2008; Sniegowski et al. 2002; Naumov et al. 1998; Koufopanou et al. 2006). However, these studies have mainly focused on the isolation of *S. cerevisiae* from a single habitat type and the variation within that habitat, rather than from multiple habitats within the same geographic region. Even though *S. paradoxus* has been isolated from natural environments (oak trees), sampling of this species is similarly restricted. Thus the distribution and ecological habitat of both species is not fully resolved.

Using globally collected samples, researchers have begun to probe the genetic structure of *S. cerevisiae* through the analysis of molecular markers, leading to new hypotheses regarding the domestication of *S. cerevisiae*. Initially, results from an analysis of five genetic loci

demonstrated a signature of a strong genetic bottleneck associated with the *S. cerevisiae* strains utilized in the production of wine (Fay & Benavides 2005). This signature has since been confirmed with the analysis of complete genome sequences (Liti et al. 2009). While genetic evidence points to the domestication of wine strains (Liti et al. 2009; Fay & Benavides 2005), the *S. cerevisiae* isolates used to generate this hypothesis represent samples that span decades and continents, and were not originally collected to test specific predictions about domestication. It remains to be seen whether the differentiation between 'wine' and 'wild 'strains is ecologically relevant when comparing strains isolated contemporarily from the same geographical locations. In order to assess the ecological distribution of *S. cerevisiae* and *S. paradoxus* populations in human associated (vineyard) and natural (non-vineyard) locations, we describe the isolation and characterization of yeast populations from grapes and oak trees. This study was designed to isolate, compare, and characterize populations of *S. cerevisiae* and S. *paradoxus* from shared temporal and spatial habitats in order to test hypotheses regarding the genetic genetic structure and domestication of wine strains of *S. cerevisiae* in an ecological context.

Materials and Methods

Study Sites

Samples were collected from a total of eight study sites: two vineyards and two non-vineyard locations each in Missouri and Oregon, USA. In Missouri, vineyard sites were located in Ste. Genevieve county (Chaumette Vineyards) and St. Charles County (Augusta Winery) Non-vineyard sites were in St. Louis county (Tyson Research Center), and Washington County (L. Watrud, personal property). In Oregon, vineyard sites were located in Polk County (Whistling Dog Cellars) and Benton County (Tyee Wine Cellars). Non-vineyard sites were in Benton County (Chip Ross State Park and M. Bollman, personal property) (Figure 3-1).

Sampling and Enrichment

Samples were collected from two different environments at vineyard locations; from damaged grapes and from adjacent, vineyard-associated oak trees. At non-vineyard locations samples were collected from oak trees. Damaged grapes and oak trees were chosen for sampling based on previously published studies showing high recovery rates (Sampaio & Gonçalves 2008; Mortimer & Polsinelli 1999; Naumov et al. 1998; Sniegowski et al. 2002). Damaged grapes were removed from the vine using ethanol sterilized forceps and macerated using an ethanol sterilized metal rod. Oak tree samples were taken from bark, twig and surrounding soil found at the base of established trees > 8.9 cm (3.5 inches) in diameter. Oak bark samples were scraped from the tree using ethanol sterilized knives, twigs were cut using ethanol sterilized scissors, and soil was collected using ethanol sterilized spatulas. All types of samples were placed into sterile plastic 15 ml screw cap conical vials. Samples were collected in 2008 from all 8 locations during the harvest season for vineyards in Missouri (September), and Oregon (October). Additional samples were collected from the Missouri Chaumette Vineyard and Tyson sampling locations in 2009. See Table 3-1 and Table 3-2 for a list of the samples collected at each location. A description of the entire sampling and enrichment procedure is found in Figure

3-2. In addition to grape and oak samples, samples from dejuiced grape mash and from a spontaneously generated wine fermentation were collected at Chaumette Vineyards. Samples from dejuiced grape mash were enriched following the procedure used for grape and oak samples. Spontaneous fermentation samples were not enriched, but rather plated directly onto YPD agar medium for colony recovery.

Samples were enriched for *S. cerevisiae* and other yeasts species that favor similar growing conditions by adding 6 mL of sterile enrichment media to the sample, closing the tube and allowing it to ferment. Two different types of enrichment media were used, a high sugar medium (H), YPD containing 10% dextrose and 5% ethanol, adjusted to pH 5.3 (Mortimer & Polsinelli 1999) and a low sugar medium (L) containing 6.7 g/L yeast nitrogen base, 1% w/v glucose, and 8% v/v ethanol, an adaptation from Sampaio & Gonçalves (2008) to determine which enrichments increase the recovery of *S. cerevisiae*. After 7 days of fermentation, a 200 µl sample was transferred into a new 15 ml vial with 6 mL of fresh sterile enrichment media, and allowed to ferment for an additional 4 days. Following the second fermentation, 2 µl of enriched medium was plated onto YPD plates, and incubated at 30ºC for 2 days. One to six colonies from each plate were restreaked for purity, and frozen stock cultures of an overnight (YPD) culture were prepared in 15% glycerol at -80ºC. For samples collected in 2009, only the high sugar enrichment medium was used for both stages of enrichment, and only colonies that resembled *S. cerevisiae* were restreaked and frozen.

Isolate screening and species identification

Colonies that resembled bacteria were tested on YPD agar containing 10 mg/L chrloramphenicol and 100 mg/L ampicillin, bacterial-specific antibiotics. If colonies failed to survive antibiotic screening (indicating likely bacterial species) they were excluded from the study. Remaining “yeast-like” colonies were further screened with molecular methods to identify isolates belonging to the *Saccharomyces* sensu stricto group (Figure 3-2). DNA was purified from each

isolate by resuspending a colony grown on YPD in 100 µl of10 mg/ml lyticase with a small amount of glass beads in a 96 well PCR plate. Plates were sealed and incubated at 37 ^{0}C for 15 minutes, followed by a brief vortexing for 2-3 seconds and incubation at 95 ^{0}C for 10 minutes. The resulting DNA was then used as a template for a multiplex PCR assay (Nardi et al. 2006). The assay included two primer pairs, one specific to the *Saccharomyces* sensu stricto group, and the other which acts as a universal fungal primer (Table 3-3). Amplification of two PCR products indicated presence of *Saccharomyces* sensu stricto specific priming, and thus identification of *Saccharomyces* species. PCR reactions were carried out in a 25 µl reaction using 3 µl of DNA template, 0.5 µl of each primer at 10 µM concentration, 1 µl Taq polymerase, 1.2 mM DNTPs, and 4 mM MgCl2. PCR reactions were incubated at 94 ^{0}C for 2 minutes followed by 35 cycles of 94 ^{0}C for 30 seconds, 51 ^{0} C for 30 seconds and 72 ^{0} C for 2 minutes, followed by a final incubation at 72 ^{0} C for 7 minutes.

Isolates that were identified as *Saccharomyces* sensu stricto using this method were further classified using ribotyping; restriction digests of the intergenic transcribed spacer region (ITS) (McCullough et al. 1998). An initial digestion by the restriction enzyme *HaeIII* was first used to differentiate *S. cerevisiae* and *S. paradoxus* from *S. mikitae*, *S. bayanus*, and *S. kudriavzevii.* A second digestion by either *BfaI* or *MwoI* was used to further differentiate species within these two groups, respectively (Table 3-4).

Isolates from a spontaneous fermentation that were positive for antibiotic resistance but did not belong to the *Saccharomyces* sensu stricto group were identified using DNA sequencing of the ITS gene region. DNA was isolated as described above. The primers ITS-1 (5' – TCC GTA GGT GAA CCT GCG G – 3') and ITS-4 (5' – TCC TCC GCT TAT TGA TAT GC – 3') were used for PCR and sequencing as described previously (M. J. McCullough et al. 1998). PCR reactions were carried out in a 25 µl reaction using 2 µl of DNA template, 0.5 µl of each primer at 10 µM concentration, 1 µl Taq polymerase, 1.2 mM dNTPs, and 4 mM MgCl2. PCR reactions were

incubated at 94 ºC for 3 minutes followed by 30 cycles of 94 ºC for 1 minute, 60 º C for 30 seconds and 72 º C for 2.5 minutes, followed by a final incubation at 72 º C for 5 minutes. PCR reactions were cleaned up using exoSAP prior to Big Dye sequencing reactions. Sequencing reactions contained 2 µl of PCR product, 0.325 µl of each 10 µm primer, 0.5 µl of Big Dye RR mix 3.1 and 1.75 µl Big Dye 5x Buffer v 3.1 in a total volume of 10 µl. Sequencing reactions were incubated at 96ºC for 1 minute followed by 29 cycles of 96ºC for 10 seconds, 50ºC for 10 seconds. Sequencing reactions were submitted to the Genome Sequencing Center at Washington University for Sequencing. Resulting sequences were manually trimmed using Lasergene SeqMan software (DNASTAR, Inc. Madison, WI), and BLASTED against the SGD fungal genome database available at http://www.yeastgenome.org/cgi-bin/blast-fungal.pl.

Assimilation of carbon and nitrogen sources

Isolates were assessed for their ability to assimilate different carbon compounds. Isolates were tested using growth on different agar media containing yeast nitrogen base (YNB) with a final concentration of 2% for the carbon source. Carbon sources included glucose, raffinose, lactose, maltose, galactose, and sucrose. Ethanol assimilation was tested in the same manner, with a final concentration of 3% ethanol. Assimilation of nitrate was tested using growth on agar containing yeast carbon base (YCB) along with 2% nitrate. Yeast strains were grown overnight in Yeast Peptone Dextrose media (YPD), and then diluted 1:1000 into YNB for carbon assimilation, or YCB for nitrate assimilation tests. After an overnight incubation, 5 µl of each culture was dispensed onto agar test plates (as described above). Plates were incubated at room temperature, and scored after 7 days. Strains were scored as positive (growth) or negative (no growth) for assimilation of each carbon or nitrogen compound. A negative control (YNB or YCB with the carbon or nitrogen source) was also tested. Strains that were positive for growth on YNB or YCB without a carbon or nitrogen source were considered false positives, and not scored for the assimilation assays using that base. Assimilation of glucose served as a positive control.

Fermentation of carbon sources

The fermentative ability of yeast isolates was also evaluated for the carbon sources glucose, galactose and raffinose using a protocol adapted from (Heard & Fleet 1990). Yeast strains were grown overnight in Yeast Peptone Dextrose media (YPD), and then diluted 1:1000 into YNB for carbon assimilation, or YCB for nitrate assimilation tests. After an overnight incubation, 20 µl of each culture was added to 60 µl of YNB with 20% of the carbon source, plus 40 µl YNB with 0.17 g/L bromocresol green (a pH indicator), and overlaid with 80 µl of mineral in round bottom assay plates (Costar, Inc. Bethesda, MD). Strains that developed of yellow color, indicating the presence of acid (and thus fermentation) were considered positive for fermentation, and those without color development were considered negative. Positive and negative controls strains were included on each plate, and plates were scored when the positive control strains developed a yellow color, after 1-5 days.

Copper sulfate resistance

Copper sulfate resistance has been shown to be variable between in *S. cerevisiae* and is thought to represent an adaptation to vineyard life (Fay et al. 2004). Resistance to copper sulfate was tested using growth on YPD agar containing 2.5 or 7.5 mM copper sulfate (Cu_2SO_4). Yeast strains were grown overnight in YPD, and 5 µl of each culture was dispensed onto test plates. Plates were incubated at room temperature, and scored after 7 days. Strains were scored as positive (growth) or negative (no growth) for copper sulfate resistance.

Statistical Analysis

Statistical significance for enrichment and isolation, assimilation and fermentation and copper sulfate resistance was examined using Chi-square tests and Fisher's exact test. When the number of levels for the test was greater than 2, a Chi-square test was performed. If the Chi-square test was significant ($p < 0.05$), Fishers exact tests for each contrast was performed independently. Statistical analysis was carried out using R (R Development Core Team 2009).

Results

Evaluation of enrichment procedures for the recovery of Saccharomyces sensu stricto species

In 2008, four different enrichment procedures were evaluated for their effectiveness and specificity in isolating *Saccharomyces* species. These procedures used two different enrichment media in two stages (see Materials and Methods). The first used high sugar medium for both stages of enrichment (HH), the second used high sugar medium for the first stage and low sugar medium for the second stage (HL), the third used low sugar medium for the first stage and high sugar medium for the second stage (LH), and the fourth used low sugar medium for both stages (LL). A total of 1,084 samples were processed into 3,535 enrichment sub-samples. From these enrichments, a total of 3,109 isolates were streaked, purified, and examined for differences in carbon assimilation and fermentation and ability to tolerate copper sulfate.

All four enrichment methods were significantly different in the proportion of samples with successful isolations, and also in the proportion that yielded more than a single colony morphology (Table 3-5). The HH enrichment method yielded the greatest number of isolates (1118), and also had the highest proportion of enrichments with more than a single colony morphology (39%). In contrast, the LL enrichment method yielded the least number of isolates (466) and only16% yielded more than a single colony morphology (Table 3-5). Enrichments that used L medium for the first stage had fewer colonies to test, but more enrichments yielding a single colony morphology. These enrichments (LH, LL) also had an overall higher proportion of isolates that failed the bacterial-specific antibiotic test (24%) (Table 3-6). In contrast, the enrichments using H for the first stage (HH, HL), had a significantly smaller proportion of isolates that failed the antibiotic test (12-14%) (Table 3-6). Enrichments that used H medium for the first stage yielded the highest numbers of *S. cerevisiae* and *S. paradoxus* isolates, but the HH

enrichment resulted in significantly less *S. cerevisiae* and *S. paradoxus* than HL relative to the total number of isolates tested (Table 3-7).

Variation in isolation rates of Saccharomyces species

There was substantial variation in the rates of *S. cerevisiae* and *S. paradoxus* isolation from the different sample substrates (e.g. grape, bark, soil, twig). Isolation rates for both *S. cerevisiae* and *S. paradoxus* were very low (2%) for grape samples. Oak samples yielded higher proportions of *S. paradoxus* (16-43%) than *S. cerevisiae* (5-17%). For both species soil sub-samples had the highest isolation rates, followed by bark sub-samples, and then by twig sub-samples. The isolation rate from twigs was significantly lower than from bark or soil for both species, but the difference in isolation rate between bark and soil was only significant for *S. paradoxus* (Table 3-8).

Samples were collected from vineyard and non-vineyard sites in Missouri (USA), and Oregon (USA) (Figure 3-1). Despite equal sample collection sizes the proportion of *S. cerevisiae* and *S. paradoxus* isolated from these locations varies significantly (Figure 3-3 and Figure 3-4 A). Most notably, the proportion of *S. cerevisiae* (17%) relative to *S. paradoxus* (11%) was significantly higher in Missouri, while in Oregon it was significantly lower (2% v. 32%) (Figure 3-4 A). In fact, 17 out of 22 total *S. cerevisiae* isolates from Oregon were isolated from a single location (MB) (Figure 3-3). In Missouri, the isolation rates for both *S. cerevisiae* and *S. paradoxus* were significantly higher in non-vineyard locations relative to vineyard locations (Figure 3-4 B). For *S. cerevisiae* this difference was driven by a lower isolation rate from grape samples relative to oak tree samples (Figure 3-4 C), and was not significantly different between vineyard and non-vineyard oak samples (Figure 3-4 D). For *S. paradoxus*, however, isolation rates from non-vineyard oak samples were significantly higher than for vineyard oak samples (Figure 3-4 D).

Composition of a spontaneous fermentation

A total of 43 colonies were isolated from the sample collected from the spontaneous fermentation at Chaumette Vineyard in 2008. A single colony was identified as *S. cerevisiae* while 42 were not *Saccharomyces* sensu stricto species. Examination of these isolates using ITS PCR identified 28 of these 42 other isolates from the fermentation to the genus level. The majority of these isolates, 16 out of 28 (57%) where most similar in ITS sequence to *Hanseniaspora* spp., most likely *H. vineae*. We also found nine isolates with sequence similarity to *Pichia membranifaciens*, two to *Issatchenkia terricola*, and one to *Kluveromyces thermotolerans*.

Carbon and nitrogen utilization

To test for phenotypic differentiation between *Saccharomyces* sensu stricto species and other yeast species that we recovered from different habitats, we measured the assimilation and fermentation ability of several different carbon sources and nitrogen (assimilation of raffinose, lactose, maltose, galactose, sucrose, ethanol and nitrate, and fermentation of galactose and raffinose). The results presented are from samples collected from two vineyards and one non-vineyard location in Missouri. For all comparisons, p values less than 0.1 are reported. Across all habitats, *Saccharomyces* sensu stricto species varied from other yeast species in the assimilation of maltose, galactose and ethanol as well as fermentation of both galactose and raffinose (Figure 3-5). In all cases a lower proportion of non sensu-stricto isolates were able to utilize these compounds.

There was no variation in the assimilation of carbon or nitrogen within *S. cerevisiae* and *S. paradoxus;* strains were positive for raffinose, maltose, galactose, sucrose and ethanol assimilation and negative for lactose and nitrate assimilation. Fermentation of galactose was slightly lower in non-vineyard populations of both *S. cerevisiae* and *S. paradoxus,* but only significantly so for *S. paradoxus* ($p = 0.012$), in which 91% of non-vineyard isolates could ferment

galactose compared to 96% of vineyard isolates. Raffinose fermentation was not variable for either species.

Although *Saccharomyces* species were not highly variable for assimilation and fermentation phenotypes, there was clear differentiation between non *Saccharomyces* sensu-stricto species for maltose, galactose and ethanol assimilation as well as galactose and raffinose fermentation (Figure 3-6). In all cases, a smaller proportion of vineyard grape yeasts were able to utilize these resources. Most of the variation was attributable to these differences. However, a significantly lower proportion of yeast isolated from vineyard oak trees was able to assimilate galactose and ethanol than their non-vineyard counterparts (Figure 3-6). This comparison remained significant for both vineyard locations considered independently (Figure 3-7). In addition to differences between vineyard and non-vineyard communities, we also found variation in carbon use between non *Saccharomyces* sensu stricto yeast strains isolated from different sub-samples on oak trees. The proportion of isolates from soil that could assimilate maltose was marginally less than that from bark or twig samples, but significantly greater for galactose fermentation and significantly lower for raffinose fermentation (Figure 3-8).

Copper sulfate resistance in yeast communities

To further assess differentiation of yeast communities, we measured the ability of isolates to tolerate and grow in the presence of copper sulfate. Results presented are from strains isolated from two vineyards and one non-vineyard location in Missouri. For all comparisons, p values less than 0.1 are reported. For non sensu stricto yeast isolates, there was no significant difference in copper sulfate resistance between isolates from vineyard or non-vineyard habitats (Figure 3-9). The proportion of *Saccharomyces* sensu stricto isolates able to grow at 2.5 mM or 7.5 mM copper sulfate was significantly lower than non sensu stricto yeast isolates (Figure 3-9). At the 2.5 mM concentration, nearly all *S. paradoxus* isolates were capable of growth on copper sulfate, compared to around 70% of *S. cerevisiae* isolates ($p = 0.046$). At 7.5 mM, however, a greater

proportion of *S. cerevisiae* isolates were able to grow on copper sulfate, although the difference is not significant (Figure 3-9).

Within *S. cerevisiae,* there is significant variation in copper sulfate resistance between vineyard and non-vineyard habitats. Specifically, the proportion of isolates resistant to 7.5 mM copper sulfate is significantly greater in vineyards than outside of vineyards (Figure 3-10). The difference is not driven by differentiation between isolates from grapes and oak trees within the vineyard, but rather by the differences between isolations from vineyard and non-vineyard habitats. This is evident based on the significant difference between vineyard and non-vineyard oak isolates, but lack of differentiation between vineyard oak and grape isolates (Figure 3-10). Notably, the proportion of resistant *S. cerevisiae* isolates is significantly lower for soil samples compared to both bark and twig samples (Figure 3-10). When considered independently, copper sulfate resistance between the two vineyards locations was significantly different. At 7.5 mM the proportion of resistant isolates from Chaumette (30%) was significantly higher than the proportion from Mount Pleasant (0%) (Fisher's exact test $p < 0.001$). Unlike for *S. cerevisiae,* resistance to copper sulfate in *S. paradoxus* was not significantly different between the vineyard and non-vineyard habitats.

Discussion

We have shown that sampling and enrichment strategies have a significant impact on the rate of *S. cerevisiae* and *S. paradoxus* isolation. High sugar enrichments yielded a lower proportion of *Saccharomyces* isolates, but due to the high number of colonies resulting from these enrichments, they also yielded the highest absolute number of *Saccharomyces* isolates. This enrichment medium was also characterized by fewer bacterial isolates than the low sugar medium. Therefore, for samples where *Saccharomyces* is expected to be in low abundance on a given sample substrate, it may be beneficial to use HH or HL enrichment protocols. If *Saccharomyces* is expected to be in high abundance, using the LH or LL enrichment protocols may significantly reduce the number of non *Saccharomyces* species, while still allowing for sufficient recovery of *Saccharomyces* isolates.

We isolated 43 colonies from a spontaneous fermentation at Chaumette Vineyards in 2008. Of these isolates, only one was *S. cerevisiae.* We identified a number of the other isolates to the genus level, and found that they were yeasts that have been previously associated with the early stages of fermentation (Querol et al. 1990; Fleet 1993). In spontaneous fermentations, *S. cerevisiae* is reported to dominate other species in middle to late stages of fermentation (Fleet 1993), thus it is possible that had we sampled the fermentation at a later stage, *S. cerevisiae* would have been in higher frequency.

The isolation rate of *S. cerevisiae* from grape samples (2%) was very low compared to isolations rates from vineyards in Italy (20%), (Mortimer & Polsinelli 1999), despite using the same enrichment protocol. However, previous studies have demonstrated that the abundance of *S. cerevisiae* on grapes in vineyards is highly associated with the ripening of grapes (Valero et al. 2007). Differences between isolation rates could be due to this harvest-dependent presence of *S. cerevisiae* or could be due to differences in other untested environmental parameters. Isolation rates of *S. cerevisiae* and *S. paradoxus* from oak bark (14-28%), were similar to previous studies

(Sampaio & Gonçalves 2008), potentially indicating more comparable yeast communities between distant natural environments than between distant vineyards locations.

We also found significant differences in the isolation rates of *S. cerevisiae* and *S. paradoxus* based on geographical location. Despite utilizing the same enrichment protocol, isolates of *S. cerevisiae* were more prevalent from Missouri, whereas in Oregon *S. cerevisiae* was nearly absent and *S. paradoxus* was more common. It is possible that different environmental selective pressures between Missouri and Oregon habitats have resulted in different habitat use between the species. This includes both differences in biotic (e.g. different trees or grapes species), or abiotic (e.g. temperature, humidity) pressures. It is also possible that we artificially created these differences due to an undetected difference in the enrichment procedure, or to differences involved in the transport and handling of samples prior to enrichments.

Within Missouri samples, *S. paradoxus* isolates represented a significantly higher proportion of the community on non-vineyard oak trees than they did on vineyard oak trees, suggesting that the oak tree habitat within and outside of the vineyard may have different biotic or abiotic selective pressures. As the proportion of *S. cerevisiae* did not differ between these two oak tree habitats, this result could indicate increased competition for the specific niche occupied by *S. paradoxus* on oak trees within vineyards. In support of this observation, significant variation in assimilation and fermentation phenotypes within non-*Saccharomyces* sensu stricto isolates was also observed, also suggesting that there are differences in community structure between the habitats. However, it is possible that the observed fermentation and assimilation differences between oak tree communities (vineyard vs. non-vineyard) could also be due in part to migration of species between the grape habitat and oak habitat within the vineyard and may not specifically imply that the oak tree habitat itself is significantly different in vineyard and non-vineyard locations.

Results from community and species specific differences in copper sulfate resistance demonstrated clear differences between vineyard and non-vineyard habitats. Strains of *S. cerevisiae* isolated from vineyard habitats are more likely to be resistant at high levels of copper sulfate than non-vineyard isolates. However, the difference was not correlated with the grape or oak tree habitat, suggesting migration of resistant *S. cerevisiae* between grapes and oak trees. While it is possible that application of copper sulfate could drift to oak trees and result in selection pressure for resistance, it is typically applied directly to grape vines. There was a significant difference in copper sulfate resistance between oak tree microhabitats. Significantly fewer soil isolates show resistance to copper sulfate than bark or twig samples. This observation also potentially supports the observation that copper sulfate stress may be restricted within the vineyard through application techniques.

Non-*Saccharomyces* sensu stricto isolates showed no variation in resistance to copper sulfate between habitats and were typically resistant. Since non-vineyard *S. cerevisiae* are typically not resistant to copper sulfate, but other yeasts (non-sensu strict) are resistant, copper sulfate resistance, which is attributed to increased copy number variation at the *CUP1* locus (Fogel et al. 1983), appears to be an acquired trait associated with *S. cerevisiae* in vineyards. This result is in agreement with previous studies that have documented variation in copper sulfate resistance within *S. cerevisiae* (Fay et al. 2004). We have demonstrated that *S. cerevisiae* and *S. paradoxus* can be isolated with varying success through several different enrichment procedures at locations within the US. We also provide evidence for differentiation of assimilation and fermentation phenotypes at the community level between habitat types (e.g. vineyard, non-vineyard, grape, oak). Copper sulfate resistance is specifically associated with vineyard isolates of *S. cerevisiae,* congruent with the hypothesis that resistance to copper sulfate represents an adaptation to vineyard life (Fay et al. 2004; Mortimer 2000). An increase in copper sulfate resistance for *S. paradoxus* or other species within vineyards is not apparent, even though the resistance phenotype exists in those species. This observation suggests that either copper

sulfate resistance is not an adaptation to vineyard life, that the strength of selection is variable between species, or that selective pressures in these vineyards have changed.

In the future, this and other collections will be essential for appropriately testing hypotheses regarding the demography, genetic structure and community ecology of *Saccharomyces* in the context of local adaptation and domestication.

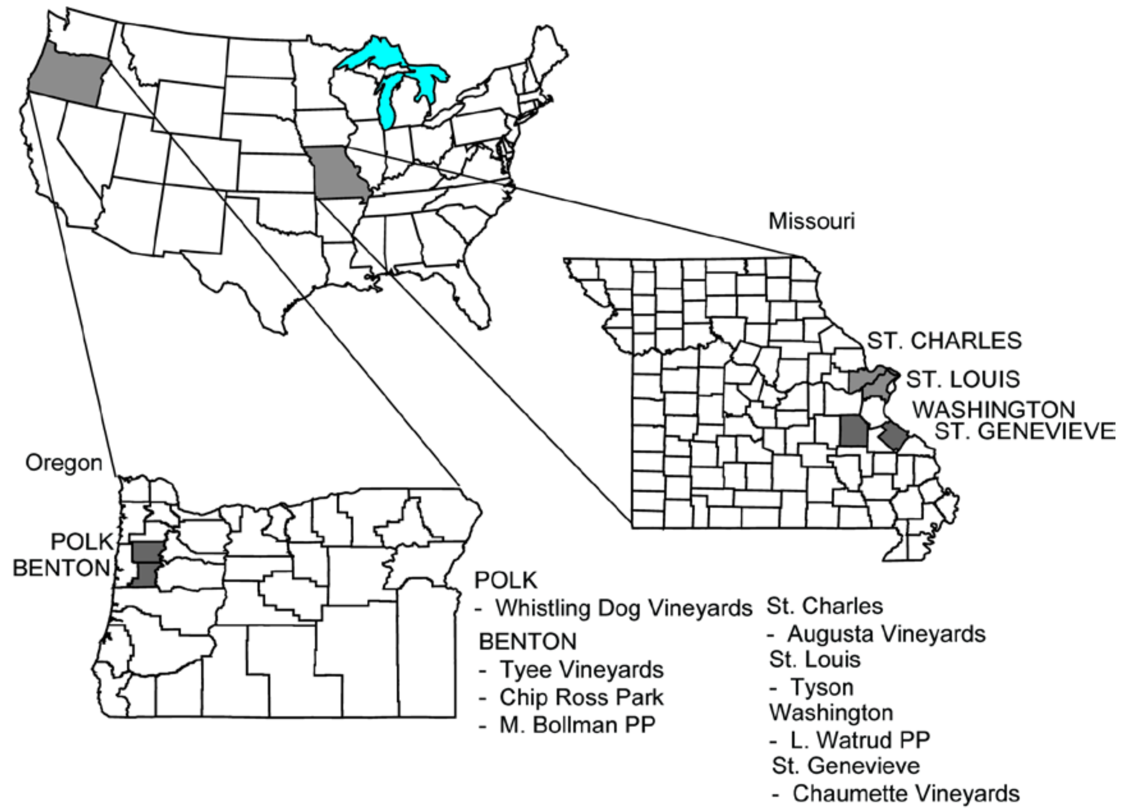


Figure 3-1. Sampling locations for this study.

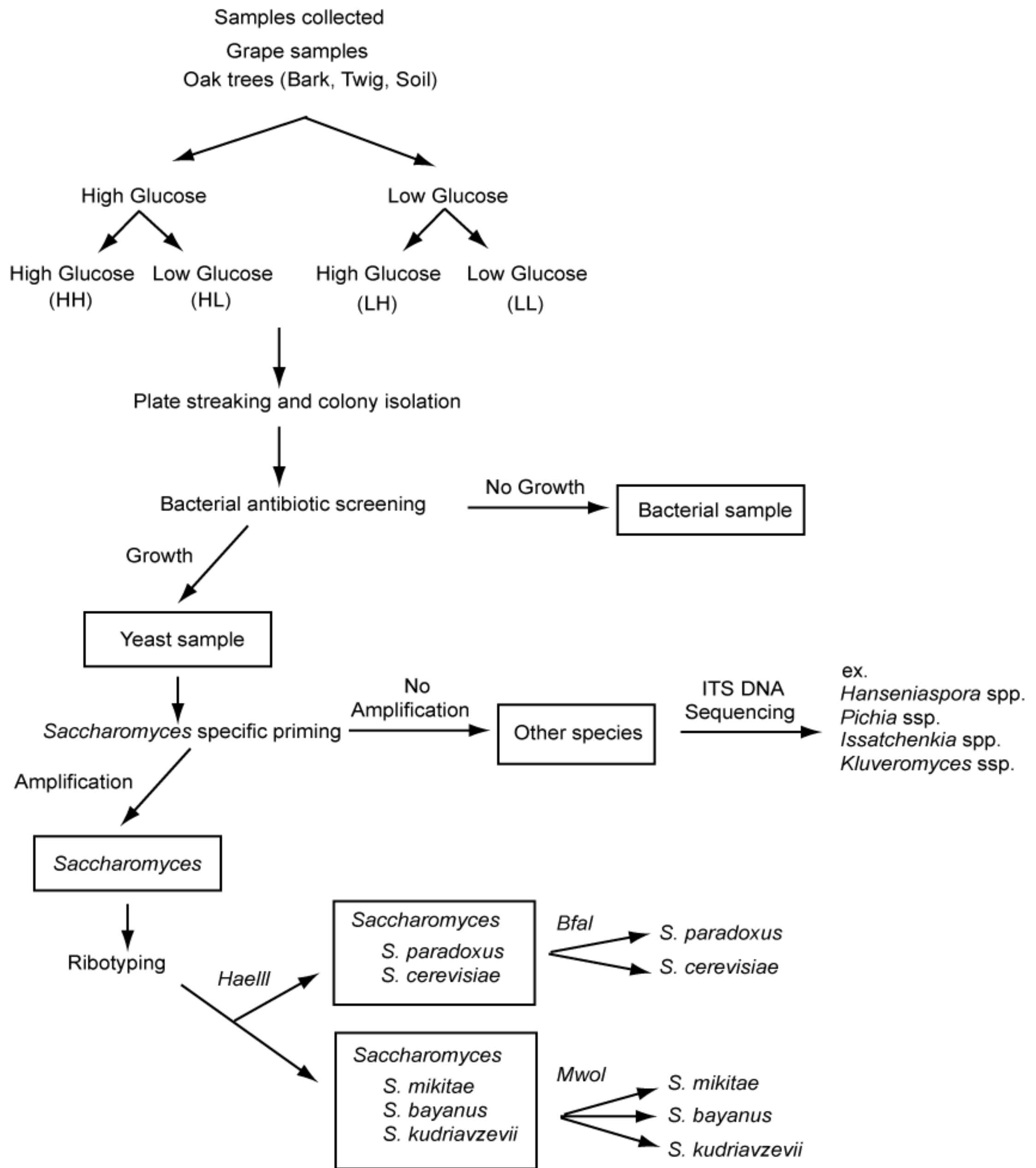


Figure 3-2. Sampling and isolation procedure for *Saccharomyces* spp.

Table 3-1. Samples collected in 2008.

State	Site	Location	Grape	Bark	Soil	Twig	Total
MO	non-vineyard	L.W. pp	-	53	32	14	99
		Tyson	-	42	30	23	95
	vineyard	Chaumette	100	52	17	19	188
		Mount Pleasant	100	19	13	19	151
OR	non-vineyard	Chip Ross	-	51	21	27	99
		M.B. pp	-	39	25	26	90
	vineyard	Tyee	100	32	22	24	178
		Whistling Dog	100	40	29	15	184
Total			400	328	189	167	1084

Table 3-2. Samples collected in 2009.

State	Site	Location	Grape	Grape mash	Bark	Soil	Total
MO	non-vineyard	Tyson	-	-	66	33	99
	vineyard	Chaumette	271	5	80	40	396

Table 3-3. Multiplex PCR assay for *Saccharomyces* sensu stricto.

Forward Primer	**Reverse Primer**	**Product Size (bp)**	**Specificity**
SAC18F	SAC18R	900	Fungi
SAC26F	SAC26R	471	*Saccharomyces* sensu stricto

From (Nardi et al. 2006)

Table 3-4. Size of expected fragments for *Saccharomyces* sensu stricto ribotyping.

	HaeIII[1]	***BfaI (MaeI)***[1]	***MwoI***[2]
S. cerevisiae[1]	311, 231, 172, 127	607, 154, 80	340, 207, 168, 126
S. paradoxus[1]	312, 229, 172, 128	760, 81	339, 294, 208
S. mikitae[2]	484, 228, 126	606, 151, 81	335, 295, 208
S. bayanus[1]	481, 229, 128	604, 153, 81	344, 336, 129, 29
S. kudriavzevii[2]	484, 229, 123, 5	607, 155, 79	339, 208, 139, 126, 29

Expected fragment sizes are in base pairs.

[1] From (McCullough et al. 1998)

[2] From this study

Table 3-5. The performance of different enrichment media on overall isolate recovery in 2008.

First enrichment medium	Second enrichment medium	Total Number of enrichments	Percent of enrichments yielding colonies to test	Percent of enrichments with colonies to test yielding more than one colony morphology
High Sugar	High Sugar	885	^{A} 85%	^{A} 39%
	Low Sugar	884	^{B} 75%	^{B} 26%
Low Sugar	High Sugar	883	^{C} 56%	^{B} 31%
	Low Sugar	883	^{D} 45%	^{C} 16%
Total		3535	65%	19%

Contrasts that are not significantly different ($p > 0.05$, Fisher’s exact test) are indicated by a shared letter.

Table 3-6. The proportion of bacteria recovered by different enrichment media in 2008.

First enrichment medium	Second enrichment medium	Isolates tested	Percent of isolates that failed the antibiotic test
High Sugar	High Sugar	1118	[A] 14%
	Low Sugar	857	[A] 12%
Low Sugar	High Sugar	668	[B] 24%
	Low Sugar	466	[B] 24%
Total		3109	17%

The antibiotics used in the test (chloramphenicol and ampicillin) inhibit bacterial growth, but do not affect the growth of fungi.

Contrasts that are not significantly different ($p > 0.05$, Fisher's exact test) are indicated by a shared letter.

Table 3-7. The proportion of *Saccharomyces* species recovered by different enrichment media in 2008.

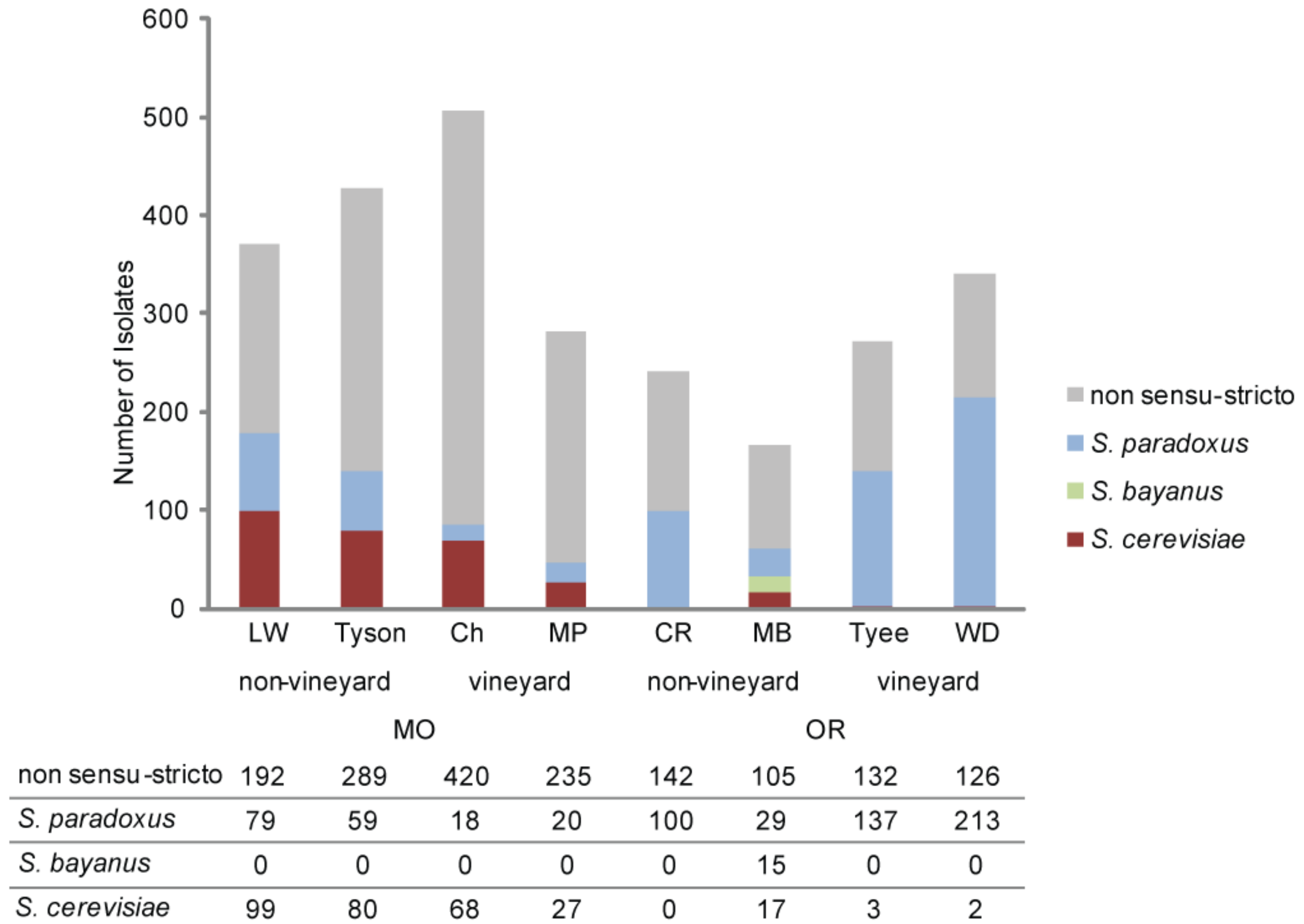
First enrichment medium	Second enrichment medium	isolates tested	*S. cerevisiae*	*S. paradoxus*	*S. bayanus*	Non sensu-stricto
High Sugar	High Sugar	960	^{A} 89 (9%)	^{A} 207 (22%)	10 (1%)	^{A} 654 (68%)
	Low Sugar	744	^{B} 90 (12%)	^{B} 201 (27%)	4 (1%)	^{B} 449 (60%)
Low Sugar	High Sugar	501	^{B} 60 (12%)	^{B} 139 (28%)	1 (0%)	^{B} 301 (60%)
	Low Sugar	359	^{B} 52 (14%)	^{B} 109 (30%)	0 (0%)	^{B} 198 (55%)
Total		2564	291 (11%)	656 (26%)	15 (1%)	1602 (62%)

Contrasts that are not significantly different ($p > 0.05$, fisher's exact test) are indicated by a shared letter.

Table 3-8. The proportion of *Saccharomyces* species recovered from different substrates in 2008.

Sample Substrate	**Isolates tested**	***S. cerevisiae***	***S. paradoxus***	***S. bayanus***	**Not sensu stricto**
Bark	1058	[A] 149 (14%)	[A] 294 (28%)	5 (0%)	[A] 610 (58%)
Berry	410	[B,E] 10 (2%)	[B] 8 (2%)	0 (0%)	[B] 392 (96%)
Soil	654	[A] 112 (17%)	[C] 284 (43%)	7 (1%)	[C] 251 (38%)
Twig	443	[C] 24 (5%)	[D] 69 (16%)	3 (1%)	[D] 347 (78%)
Spontaneous Fermentation	42	[C,E] 1 (2%)	[B] 0 (0%)	0 (0%)	[B] 41 (98%)
Total	2607	296 (11%)	655 (25%)	15 (1%)	1641 (63%)

Contrasts that are not significantly different ($p > 0.05$, fisher's exact test) are indicated by a shared letter.



	LW	Tyson	Ch	MP	CR	MB	Tyee	WD
non sensu-stricto	192	289	420	235	142	105	132	126
S. paradoxus	79	59	18	20	100	29	137	213
S. bayanus	0	0	0	0	0	15	0	0
S. cerevisiae	99	80	68	27	0	17	3	2

Figure 3-3. The number of *Saccharomyces* sensu stricto isolates recovered by location in 2008.

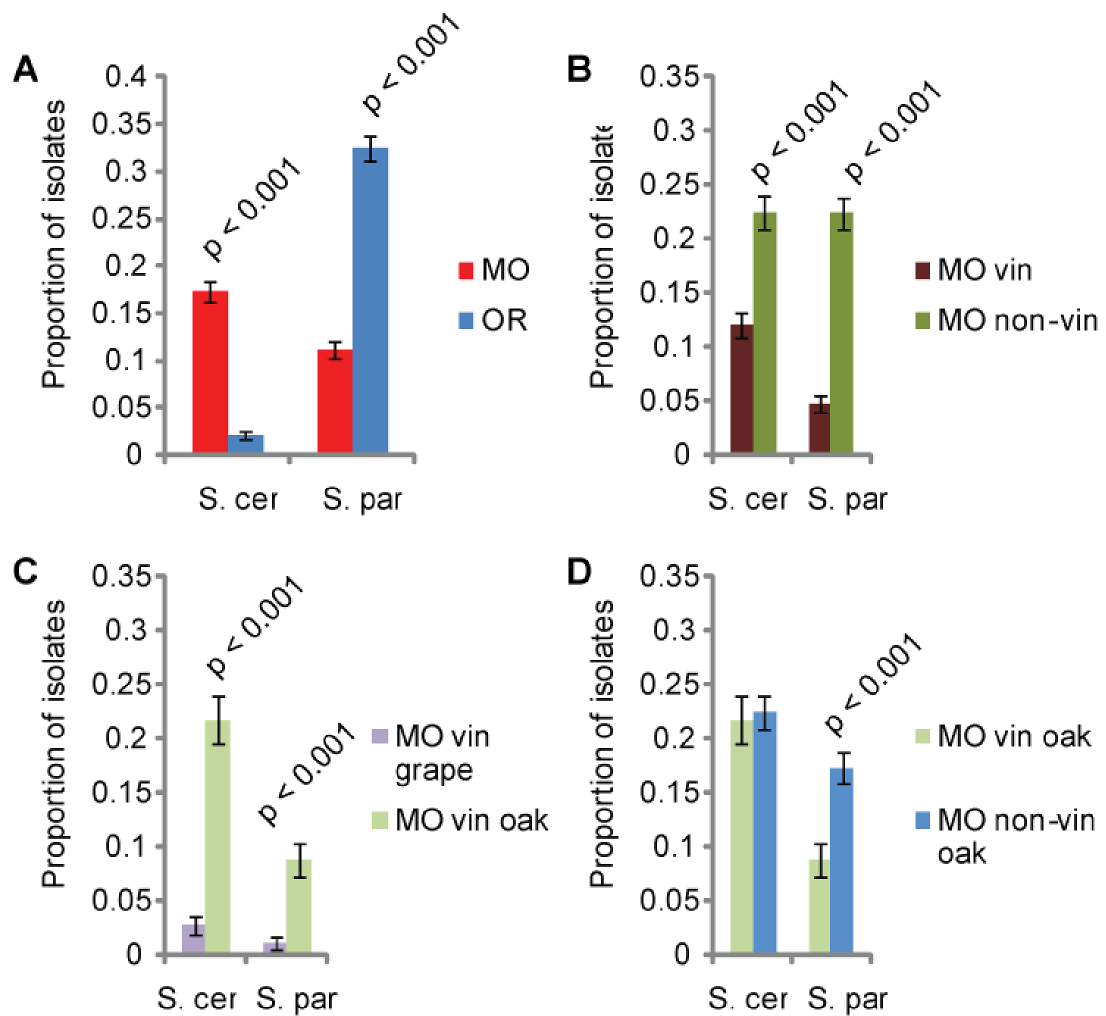


Figure 3-4. The proportion of *S. cerevisiae* and *S. paradoxus* isolates from Missouri and Oregon in 2008.

The proportion of *S. cerevisiae* (S. cer) and *S. paradoxus* (S. par) isolates out of the total number of yeast isolates tested for (A) Missouri and Oregon, (B) Vineyards and non-vineyard locations in MO, (C), grapes and oaks in MO vineyards, and (D) oak from vineyard and non-vineyard locations in MO. P values less than or equal to 0.1 are shown.

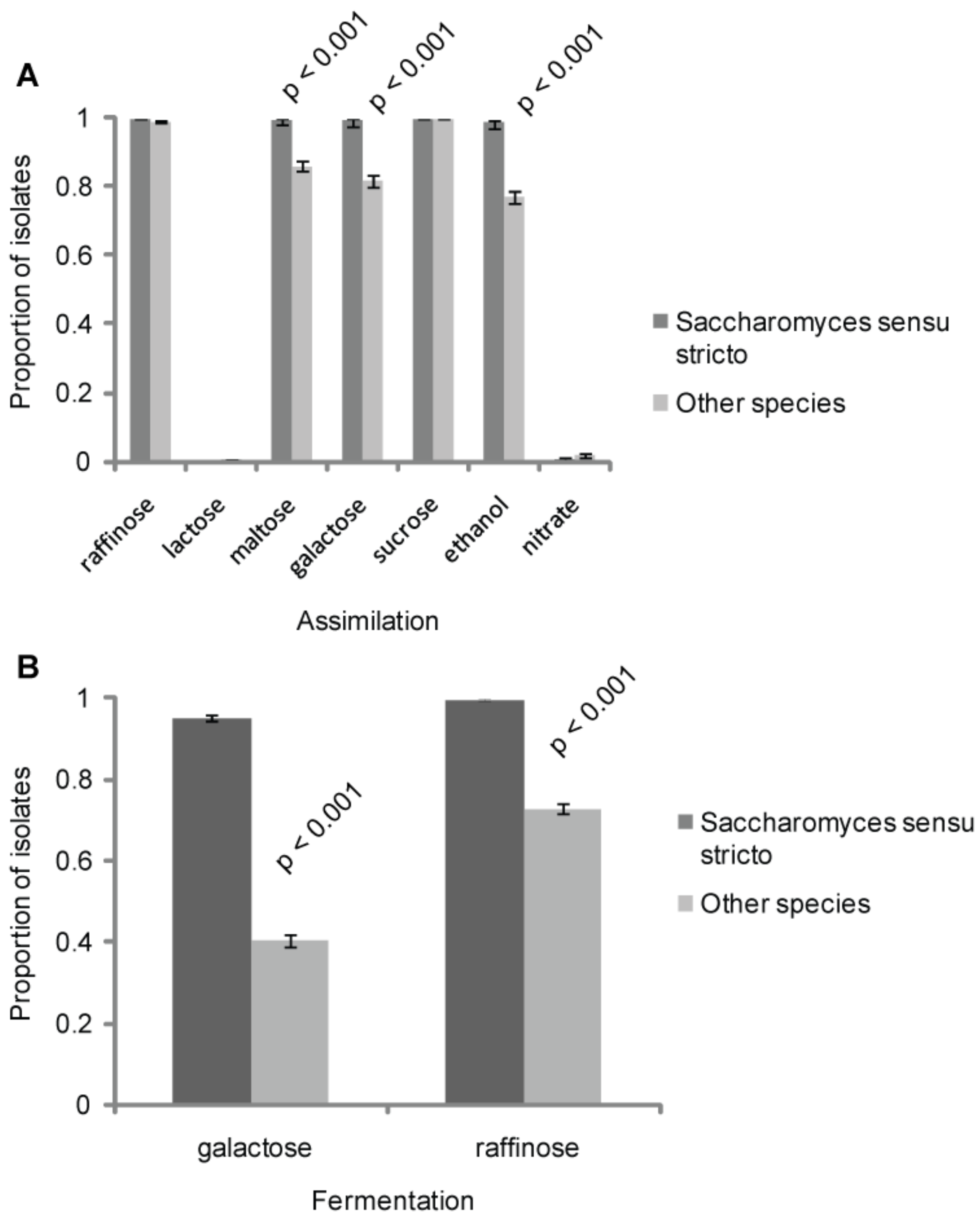


Figure 3-5. Variation in carbon and nitrogen use between *Saccharomyces* and non-*Saccharomyces* species from Missouri.

The proportion of all yeast isolates able to (A) assimilate or (B) ferment of various carbon and nitrogen sources. P values less than or equal to 0.1 are shown.

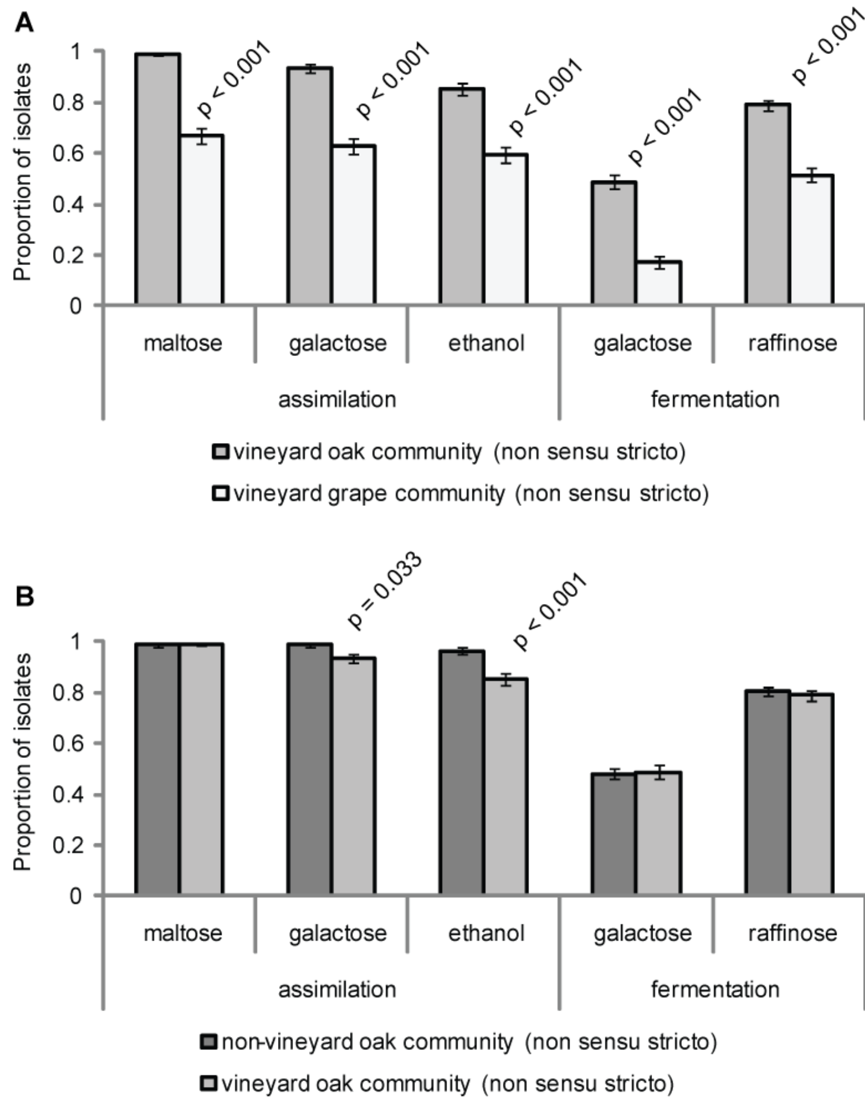


Figure 3-6. Variation in carbon and nitrogen use between oak and grape yeast communities in Missouri vineyards.

The proportion of non *Saccharomyces* sensu stricto isolates (A) from Missouri vineyard grapes and vineyard oaks, and (B) from Missouri vineyard oaks and non-vineyard oaks able to assimilate or ferment various carbon and nitrogen sources. P values less than or equal to 0.1 are shown.

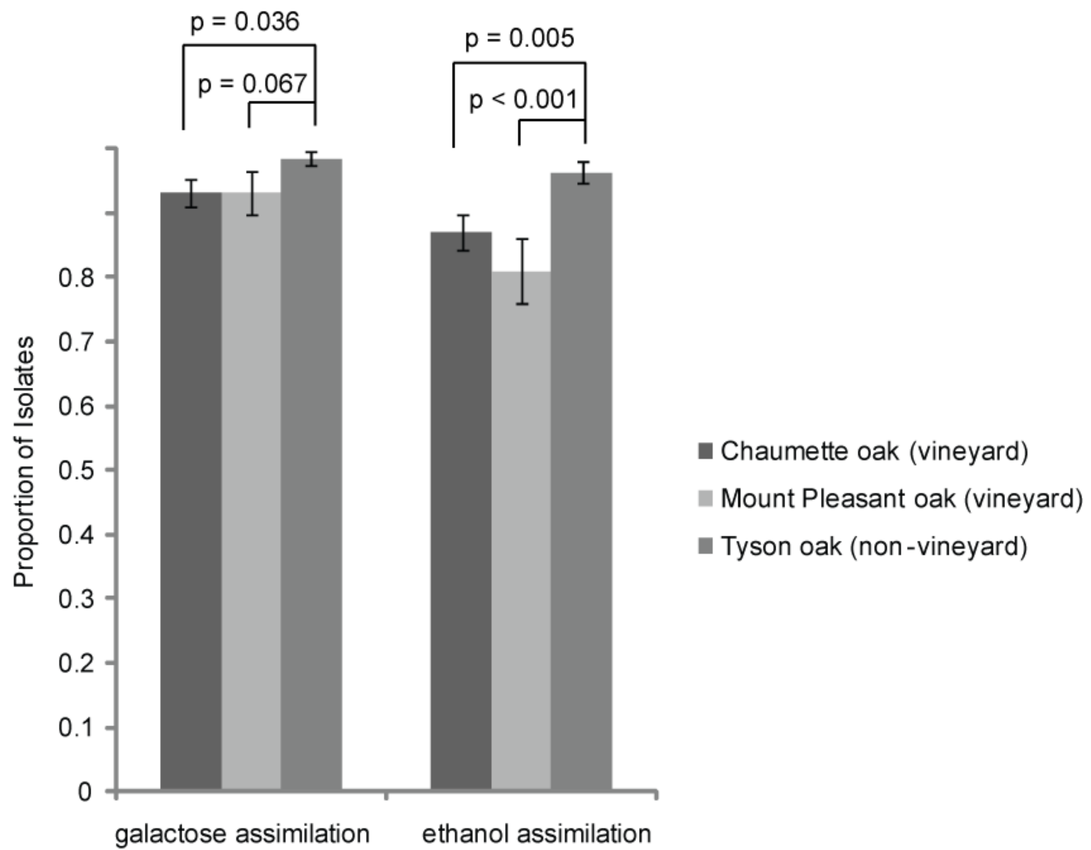


Figure 3-7. Variation in carbon and nitrogen use by non *Saccharomyces* sensu-stricto yeast species living on oak trees at three different locations.

The proportion of non *Saccharomyces* sensu stricto isolates from oak trees at different locations in Missouri able to assimilate or ferment various carbon and nitrogen sources. P values less than or equal to 0.1 are shown.

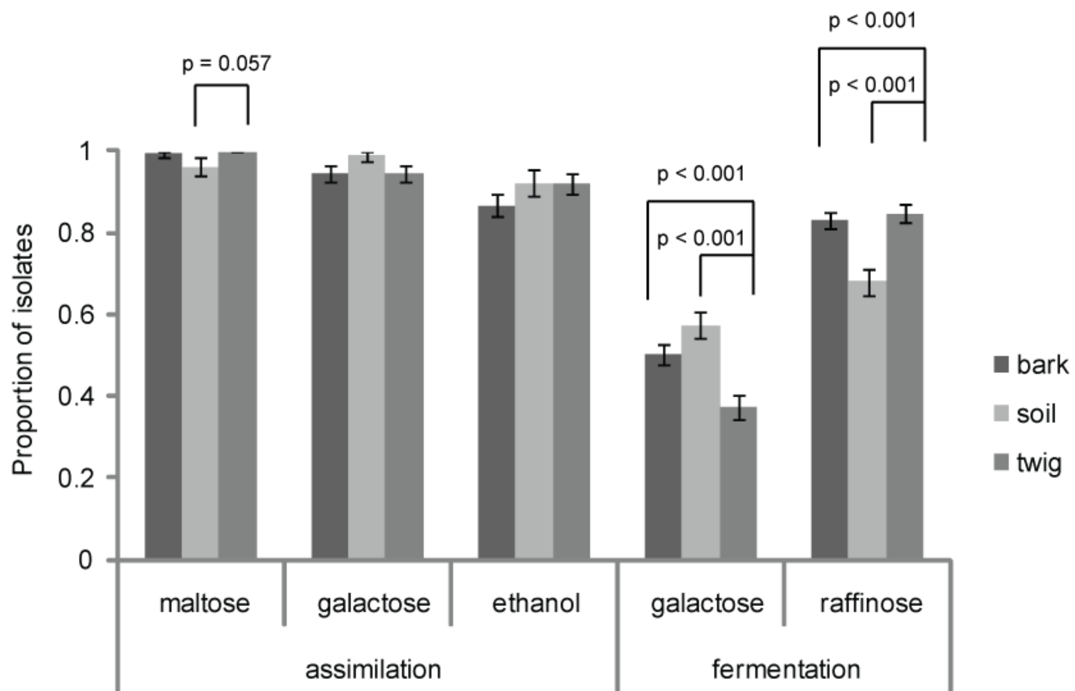


Figure 3-8. Variation in carbon and nitrogen use between bark, twig and soil non *Saccharomyces* sensu stricto yeast isolates.

Variation in assimilation and fermentation of various carbon and nitrogen sources. P values less than or equal to 0.1 are shown.

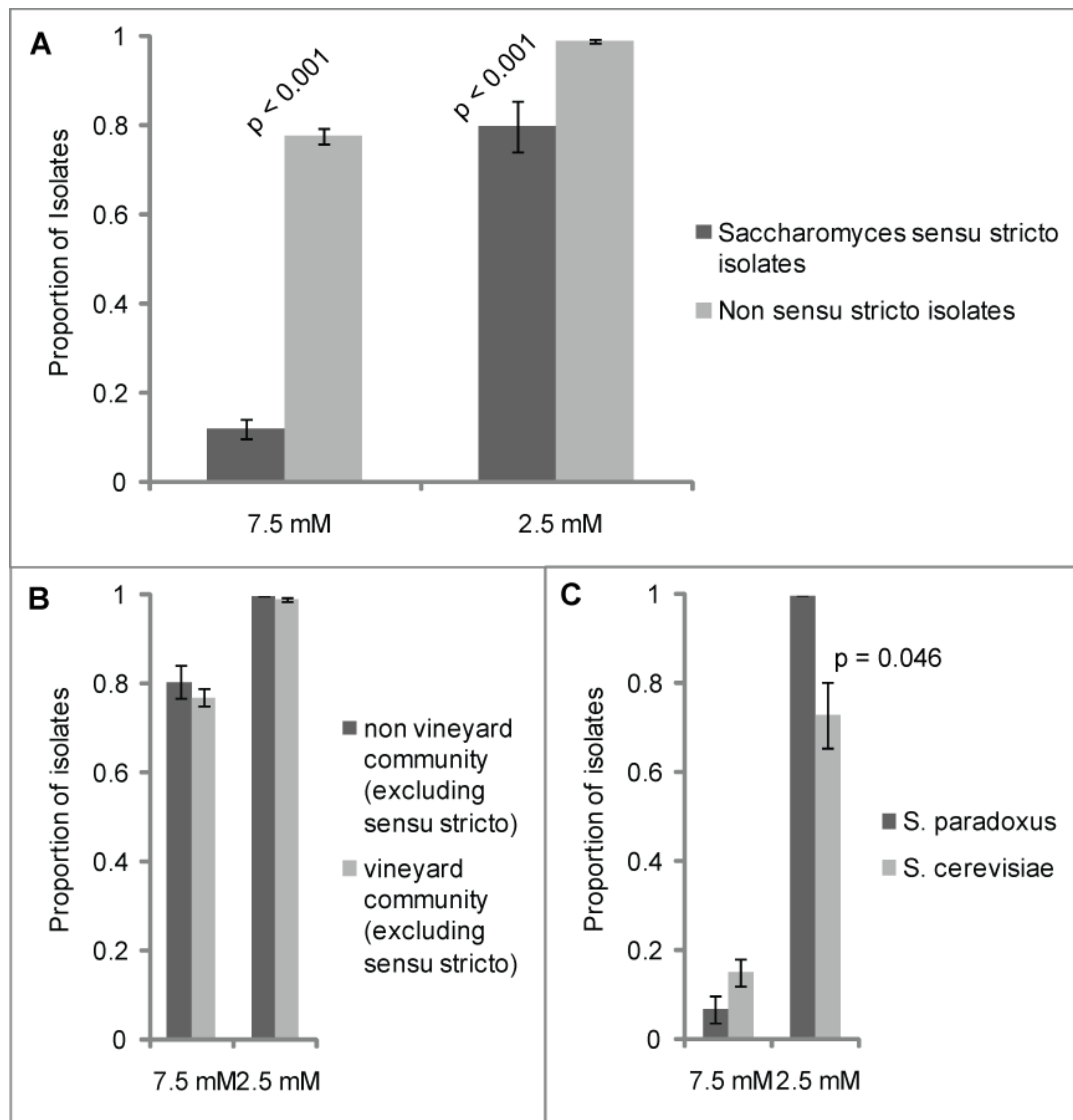


Figure 3-9. Copper sulfate resistance in vineyard and non-vineyard yeast communities.

The proportion of isolates able to grow at 7.5 mM and 2.5 mM copper sulfate (Cu_2So_4) for (A) *Saccharomyces* sensu strict and non- *Saccharomyces* sensu strict isolates, (B) non sensu strict isolates from vineyards and non-vineyards, and (C) *S. cerevisiae* and *S. paradoxus*. P values less than or equal to 0.1 are shown.

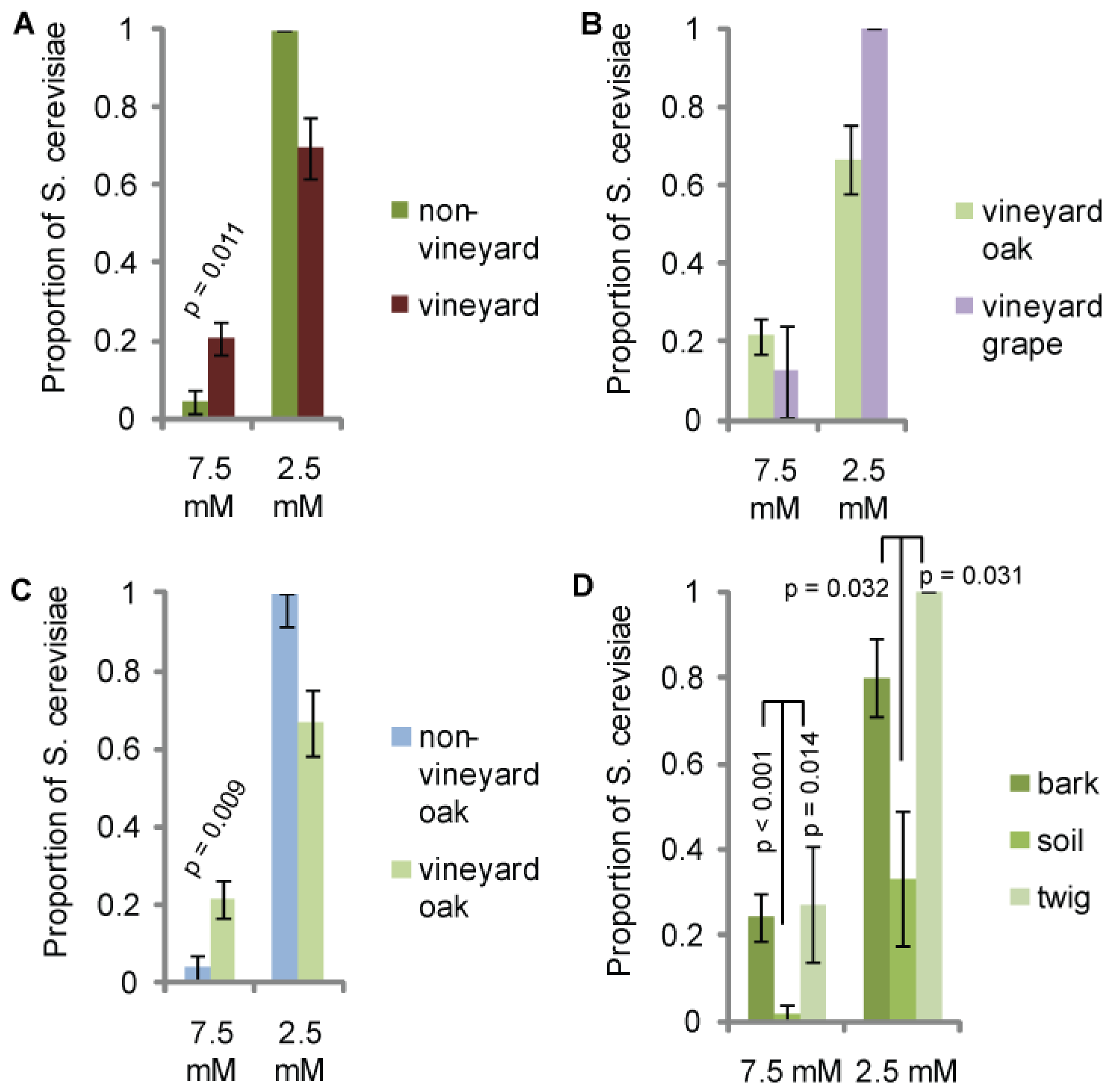


Figure 3-10. Copper sulfate resistance in vineyard and non-vineyard strains of *S. cerevisiae.*

The proportion of *S. cerevisiae* isolates able to grow at 7.5 mM and 2.5 mM copper sulfate (Cu_2So_4) for (A) vineyard and non-vineyard isolates, (B) vineyard grape and vineyard oak isolates, (C) vineyard and non-vineyard oak isolates, and (D) different types of oak samples. P values less than or equal to 0.1 are shown.

Chapter 4 : Population Genetics of Vineyard and Non-Vineyard Populations of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*

Introduction

Although the budding yeast *Saccharomyces cerevisiae* has been utilized as a model organism for decades and was the first organism for which the entire nuclear genome was sequenced (Goffeau et al. 1996), the population genetics of the species has not been investigated until recently (Winzeler et al. 2003; Fay & Benavides 2005; Aa et al. 2006; Legras et al. 2007; Liti et al. 2009; Schacherer et al. 2009; Diezmann & Dietrich 2009). Previous studies have demonstrated that the population genetic structure of *S. cerevisiae* is driven by ecological differentiation rather than geographical distance (Legras et al. 2007; Schacherer et al. 2009; Fay & Benavides 2005; Liti et al. 2009; Diezmann & Dietrich 2009) whereas divergence within *S. paradoxus,* the closest wild relative of *S. cerevisiae,* seems to be driven by geographical distance (Koufopanou et al. 2006; Liti et al. 2009). The primary difference between these two species is that *S. cerevisiae* is intimately associated with humans; we use this species for the production of wine, beer and other alcoholic beverages, baking, and biofuel production. It has been postulated that the differences in population genetic structure can be attributed to the association of *S. cerevisiae* with humans, either indirectly through changes in dispersal and migration patterns (Legras et al. 2007), or through artificial selection in the form of domestication (Fay & Benavides 2005).

Hypotheses differ regarding the level of association between *S. cerevisiae* and humans. Because of its close association with humans and long history as a laboratory and genetic model system, one hypothesis is that the entire species has been domesticated, implying that strains isolated from 'wild' habitats represent escaped isolates (Mortimer 2000). In contrast, another hypothesis is that one subgroup of *S. cerevisiae*, strains associated with winemaking, have been domesticated from 'wild' strains occurring in natural habitats (Fay & Benavides 2005). The evidence for domestication lies in the genetic signature of wine strains; they carry only a fraction of the genetic diversity of the species as a whole, likely as a result of a domestication-associated

genetic bottleneck (Fay & Benavides 2005; Liti et al. 2009). Furthermore, genetic evidence suggests that the wine strains have been derived from 'wild' strains, rather than the opposite (Fay & Benavides 2005).

Genetic bottlenecks associated with the domestication of crops are common, as only a select number of individuals are repeatedly propagated (Doebley et al. 2006). The extent of genetic reduction in diversity depends on the size of the population during domestication and the length of time of domestication (Eyre-Walker et al. 1998). This process has also been associated with the relaxation of selective constraints, resulting in an excess of nonsynonymous substitutions in the domesticated lineage (Lu et al. 2006). Although putatively domesticated strains of *S. cerevisiae* exhibit both a population bottleneck (Fay & Benavides 2005; Liti et al. 2009) as well as a slight increase in the ratio of nonsynonymous to synonymous mutations (Doniger et al. 2008), it is unknown whether this is common for domesticated fungal species.

Although humans use many species of fungi in food production (Hesseltine 1965), there are very few studies of genetic differentiation within these species. The only example to date to examine the genetics of domestication in a fungal species comes from *Aspergillus oryzae,* a fungus used to prepare soy sauce, sake and miso that is thought to have been domesticated from wild populations of *Aspergillus flavus* (Geiser et al. 1998). These two species are phenotypically nearly indistinguishable aside from the production of the secondary metabolite aflatoxin (Geiser et al. 1998). Additionally, although the domestication is believed to have occurred approximately 2,000 years ago, there seems to be no genetic signature associated with the domestication (Rokas 2009). Although there are few studies of fungal domestication, it is seems likely that the phenotypic and genotypic indicators of domestication in fungi may not parallel those patterns observed in plant and animal species. For example, it may be that bottlenecks in fungi are less severe because entire populations of cells rather than a few individuals are selected on every

generation, or that persistent asexual reproduction decreases the efficacy of selection (Rokas 2009).

Additional challenges exist for testing the hypothesis of domestication in *S. cerevisiae.* While previous studies have provided genetic evidence that points to the domestication of wine strains, the *S. cerevisiae* isolates that have been used to infer domestication were collected over a span of several decades, and across many continents (Fay & Benavides 2005; Liti et al. 2009). It remains to be seen whether the differentiation between 'wine' and 'wild 'strains is ecologically relevant when comparing strains isolated contemporarily at a single geographical location. The goals of the study are to examine the genome-wide population genetic structure of *S. cerevisiae* within the vineyard, where both the 'domesticated' and 'wild' habitats occur together, and also beyond the vineyard to determine if the global population structure observed for *S. cerevisiae* can be recapitulated, and if so, whether there is evidence for barriers to gene flow that contribute to sympatric coexistence.

Materials and Methods

Strains

S. cerevisiae and *S. paradoxus* strains were collected from a total of eight study sites: two different vineyard and two different non-vineyard locations were sampled from the states of Missouri and Oregon, USA. In Missouri, vineyard sites were located in Ste. Genevieve County (Chaumette Vineyards) and St. Charles County (Mount Pleasant Winery), and non-vineyard sites were in St. Louis County (Tyson Research Center), and Washington County (L. Watrud, personal property) In Oregon, vineyard sites were located in Polk County (Whistling Dog Cellars) and Benton County (Tyee Wine Cellars), and non-vineyard sites were in Benton County (Chip Ross State Park and M. Bollman, personal property). See Chapter 3 for details on sample collection and Table 4-1 for a description of the strains used in this study. Seven additional strains were also genotyped: four strains isolated in Wisconsin were provided by Audrey Gasch (Table 4-3), and three strains isolated in Ecuador were provided by Javier Carvajal (Table 4-3).

Genotyping

Restriction-site associated DNA tags (RAD tags) were sequenced using a protocol based on Baird et al. (2008). Genomic DNA was isolated using ArchivePure DNA Yeast & Gram -+ Kits (5 Prime, Inc.), quantified using the Quant-it™ dsDNA HS Assay (Invitrogen Corporation), adjusted to a standard concentration, and digested for 60 minutes at 37ºC in a 50 µl reaction with 5 units (U) each of *MfeI* and *MboI* (New England Biolabs, Inc.), followed by heat inactivation for 20 minutes at 65ºC. Digested genomic DNA was ligated to P1 adaptor, a modified Solexa© adaptor (2006 Illumina, Inc., all rights reserved; top: 5' –ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT xxxx – 3' [x = barcode], bottom: 5'- Phos – AATT xxxx AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT - 3', and P2 adaptor, a modified Solexa© adaptor (2006 Illumina, Inc., all rights reserved; top: 5' - Phos – GAT CCT CAG GCA TCA CTC GAT TCC TCC GAG AAC AA – 3' : bottom: 5' - CAA GCA GAA GAC GGC ATA CGA CGG AGG AAT CGA GTG ATG

CCT GAG – 3' with 1000 U concentrated T4 DNA ligase (New England Biolabs, Inc.) at room temperature for 20 minutes, followed by heat inactivation at 65ºC for 20 minutes. Ligated and digested DNA was pooled and purified using a QIAquick PCR Purification Kit (Qiagen, Inc.). Fragments from 150-500 bp were isolated using a QIAquick Gel Extraction kit (Qiagen, Inc.). Fragments were then PCR amplified using 5-10 ng DNA, 25 µl Phusion High-Fidelity PCR Master Mix (New England Biolabs,Inc.), 0.5 µM of each modified Solexa© pcr primer: (solexa pcr forward P1 5' - AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CT - 3' and solexa pcr reverse P2 5' - CAA GCA GAA GAC GGC ATA CGA - 3'), and water to a final volume of 50 µl. Cycling conditions were 98ºC for 1 minute followed by 14-18 cycles of 98ºC for 10 seconds, 60ºC for 30 seconds, 72ºC for 30 seconds, and a final extension at 72ºC for 4 minutes. The resulting PCR product was purified using a QIAquick PCR Purification Kit (Qiagen, Inc.) and adjusted to 10 nm. Illumina Solexa protocols were followed for sequencing.

Sequence Analysis

Raw sequence reads were processed to reduce sequencing artifacts within the data using Perl scripts (K.E.H.). First, reads were separated by barcodes, which were examined for quality and removed prior to mapping. Reads with a Phred-scaled sequence quality score of less than 20 for any bp within the barcode, as well as reads with an unknown barcode sequence, were removed. For *S. cerevisiae,* reads were aligned to the SGRP reference genome (available at http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html) (Liti et al. 2009) using the short read alignment program Bowtie (Langmead et al. 2009). Reads that aligned to more than a single region in the reference genome were discarded from the analysis. Alignment statistics for *S. cerevisiae* can be found in Table 4-7. Alignment rates to the SGRP *S. paradoxus* reference genome were low (< 50%) for most *S. paradoxus* strains, including the control strain YPS138, likely due to the large amount of sequence divergence between North American isolates and the European isolates used to generate the reference genome (Liti et al. 2009). A new assembly was created using the North American *S. paradoxus* strains UFRJ50791, UFRJ50816,

A12, A4, YPS138 and DBVPG6304, that resulted in 5-6x coverage and was used for alignment (J. Fay, personal communication). Alignment statistics for *S. paradoxus* to this new assembly are found in Table 4-8.

In some cases, restriction sites (cycles 5-9 of Solexa sequencing) were removed during image analysis, prior to generating raw reads. In those cases, reads that did not align adjacent to a *MfeI* restriction site (AATG), allowing for one bp difference in the restriction site, were filtered from the data set. When restriction sites were not removed prior to the generation of raw reads, reads that lacked an *MfeI* restriction site at the beginning of the read were filtered from the data set.

After alignment, the first four and last four base pairs of reads were discarded. Any position in an aligned read with a Phred-scaled sequence quality score of less than 15 was masked by converting that position to an ‘n,’ changing its quality score to 0, and removing it from the calculation of sequence coverage at that position. Consensus pileups for each strain were generated using Samtools (Li et al. 2009). Sequenced positions with a consensus quality score of less than 40 or with less than 3x coverage were filtered out of the data set. Single nucleotide polymorphisms (SNPs) in the data set were retained if the SNP quality score was greater than or equal to 20, and there were no more than 2 SNPs in a 10 bp window. The results of filtering, along with sequence coverage and quality, are found in Table 4-7 for *S. cerevisiae* and Table 4-8 for *S. paradoxus*. The number of sequenced positions, SNPs and heterozygous positions for each strain are found in Table 4-9 for *S. cerevisiae* and Table 4-10 for *S. paradoxus.*

During each run we included two control strains with independent complete genome sequence data in order to estimate the false positive rate for SNPs resulting from Solexa sequencing. The expected number of false positives was calculated for each control strain as: $(FP * T)$, where the false positive rate is $\mathrm{FP} = \mathrm{X} / \mathrm{C}$, C is the number of positions where there is non-ambiguous sequence information for both Solexa strains and both reference sequences for

which the sequence of the two references are identical, X is the number of SNPs found only in the Solexa sequence for the control strain but not in either reference, and T is the total number of Solexa sequenced positions for the strain. False discovery rate estimates are found in Table 4-11.

Due to the properties of RAD tagging and Solexa sequencing, certain regions of the genome may not be sequenced in every isolate and lead to incorrect population genetic inferences. To adjust for this possibility, the sequence data set was compiled for RAD genotyped *S. cerevisiae* strains, and any position that was sequenced for at least 53 of the 54 strains was retained for population genetic analysis. After filtering, the data set included 210,566 base pairs, representing about about 1.75% of the *S. cerevisiae* genome. Additional genome sequences for a diverse set of *S. cerevisiae* strains have been described previously (Liti et al. 2009a). Genotypes for these strains were extracted from the alignments available at http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html. See Table 4-4 for a list of strains. Sequenced positions for the SGRP alignments with a Phred score of less than 20 were converted to "N"s. We also included a set of newly sequenced *S. cerevisiae* strains (Table 4-6), available at http://www.genetics.wustl.edu/jflab/data4.html. Genotype information for these strains was obtained using BLAST (Altschul et al. 1990). The genotype of the reference sequence for each aligned read was blasted (blastn) against assembled contigs for each newly sequenced genome.

The sequence data were compiled separately for RAD genotyped *S. paradoxus* strains, and any position that was sequenced for at least 24 of the 40 strains was retained for population genetic analysis, similar to the filtering employed for *S. cerevisiae*. The filtered data set included 281,944 base pairs, representing approximately 2.4% of the *S. paradoxus* genome. Additional genome sequences for a diverse set of *S. paradoxus* strains have been described previously (Liti et al. 2009). Genotypes for these strains were extracted using BLAST (Altschul et al. 1990) in the

same manner as for unpublished genome sequences. See Table 4-4 for a list of strains. The assemblies of these strains are available at http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html.

Statistical Analysis

Positions in noncoding regions, coding regions, two-fold , four-fold and non- degenerate sites were extracted from the SGRP reference genome annotation (available at http://www.sanger.ac.uk/research/projects/genomeinformatics/sgrp.html) (Liti et al. 2009) for *S. cerevisiae*. Sequence diversity was estimated as the average number of base differences per site (π) using MEGA4 (Tamura et al. 2007). All positions containing alignment gaps, missing or ambiguous data were eliminated only in pairwise sequence comparisons. The number of private alleles and minor allele frequencies were calculated using perl scripts (K.E.H.). For minor allele frequencies in *S. cerevisiae*, only one representative of each clonal group was included in the analysis. The neutral expectation for minor allele frequencies was calculated from Watterson's θ (Watterson 1975) following (Lu et al. 2006). Linkage disequilibrium (r^2), and the distance at which linkage disequilibrium decays by ½ (LD ½) was calculated using perl scripts (K.E.H.).

Phylogenetic Analysis and Population Structure

Phylogenetic trees were inferred with MEGA4 (Tamura et al. 2007) using the Neighbor-Joining method with 1,000 bootstrap replicates. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons.

Population structure was examined using the model-based program STRUCTURE (Pritchard et al. 2000). Model-based population assignment programs can often be influenced by linkage disequilibrium between informative sites. For *S. cerevisiae* we measured linkage disequilibrium as r^2 using perl scripts (K.E.H). For *S. cerevisiae,* population structure was inferred from 3,323 parsimony informative sites assuming uncorrelated allele frequencies with no linkage.

Three initial simulations at K =1 were used to infer lambda, which was set at 0.4929 for subsequent simulations. Ten replicate simulations were performed for each inferred number of populations (K), for K = 2 through K = 10 with a burn-in period of 10,000, followed by 10,000 additional Markov Chain Monte Carlo replications. We used the admixture model with independent allele frequencies. CLUMPP (Jakobsson & Rosenberg 2007) was used to assess the similarity between replicate STRUCTURE results (G') in order to determine the relative likelihood of multimodality of the inferred population structure. We used the Fullsearch algorithm to compare permutations for K=2 and K=3. For K=4-10, the GREEDY algorithm was used to reduce computational time necessary to run the permutations. Using the GREEDY algorithm, we specified the total number of permutations to test (10,000 for K=4 and K=5, 50,000 for K =6 and K=7, 10,000 for K=8 and 100 for K=9 and K=10). DISTRUCT (Rosenberg 2003) was used to visualize the results.

Population structure in *S. paradoxus* was examined as in *S. cerevisiae,* for K=2 through K=5 with lambda equal to 1. Similarity was assessed using the Fullsearch algorithm to compare 10 permutations for each inferred number of populations.

Genetic Admixture and Introgression

For *S. cerevisiae* we first used STRUCTURE to assign genotypes to structured populations without using any prior information regarding origin, using the method described above. After this unsupervised run, previously sequenced individuals from wine and wild populations were used to define 'learning' populations for introgression and admixture analysis. Admixture analysis was performed with the POPFLAG info module of STRUCTURE (Pritchard et al. 2000) so that the 'learned' populations were used to infer introgression and admixture of the North American isolates sampled in this study, along with isolates from Wisconsin.

Results

Distribution of Genetic Markers and Heterozygosity in S. cerevisiae

Our data set included 3,233 SNPs, which are distributed across the genome. We examined the physical distribution of SNPs across the genome, and found that less than 0.3% of SNPs are within 10 kb of each other (Figure 4-1). The majority of sampled strains were found to be homozygous at identified SNPs with less than 0.01% of sites having more than one allele. However, a few strains demonstrated relatively high levels of heterozygosity, including the strains isolated from cherry samples (DCM6 and DCM21), a wine strain isolated from a vineyard grape (KEH00415), a wine strain isolated from a spontaneous wine fermentation (KEH02575), and the strains isolated from ancient fermentation vessels from Ecuador (Table 4-9). The cherry strains have previously demonstrated heterozygous wine and oak haplotypes for at least one genetic locus (A. Gasch, personal communication), and are likely to be recent hybrids between wine and oak strains. The slightly increased levels of heterozygosity in some strains may indicate recent or ongoing gene flow, or may also have arisen due to sequencing and alignment errors. Although we removed DNA sequence reads that aligned to more than one region in the reference genome, heterozygous positions may represent alignment errors due to the presence of multi-copy genes or redundant sequences in the genotyped strains that are represented by a single copy in the reference genome.

Phylogenetic Analysis of S. cerevisiae

A bootstrap consensus phylogeny tree was constructed using the neighbor-joining method, based on pairwise genetic distances (nucleotide p-value) between 117 OTUs at 210,566 nucleotide sites (Figure 4-2). Results of this analysis demonstrate that all of the *S. cerevisiae* strains collected in North America belong to either the previously described European/wine/vineyard or North American oak populations (Fay & Benavides 2005; Liti et al. 2009; Legras et al. 2007; Schacherer et al. 2009; Aa et al. 2006), with the exception of two strains

isolated from cherries in Wisconsin: DCM6 and DCM21 (Figure 4-2). The inclusion of strains within either the wine or wild North American groups is highly supported by bootstrap values of 98-100.

The Distribution of S. cerevisiae Diversity

Both 'wine' and 'North American oak' genotypes were found at vineyard locations (Figure 4-3 and Figure 4-4), whereas only 'North American oak' genotypes were found at non-vineyard locations (Figure 4-4). Although both of these genotypes are present within vineyards, their distribution is not correlated with sample substrate, i.e. wine genotypes are found on both grapes and oak trees and North American oak genotypes are also found on both grapes and oak trees. Due to the incongruence between genotype and sample substrate, I will refer to the strains related to the previously described Vineyard/European lineages as 'wine' strains and those related to the North American 'oak' lineage as 'wild' strains, regardless of the substrate from which they were isolated. Although wine genotypes were present in vineyards, both on grapes and on oak trees, at non-vineyard locations we uncovered only wild genotypes.

Wine genotype strains of *S. cerevisiae* isolated from Chaumette Vineyard were found both in close physical proximity to the winery, and also dispersed throughout the vineyard. Similarly, wild strains were dispersed throughout the vineyard (Figure 4-5). Four *S. cerevisiae* samples were isolated from the Mount Pleasant Winery, two from an unknown Oak species (*Quercus sp*) and two from grape samples. Each of these four strains was resolved as an oak genotype. These two vineyard locations differ in that the Chaumette Vineyard has a functioning winery adjacent to vineyard, whereas Mount Pleasant winery ferments and produces wine at a separate location found off the premises.

Three wine genotype strains of *S. cerevisiae* sampled from Oregon locations were isolated from soil sub-samples taken from the base of the Oregon White Oak (*Quercus garryana*).

Two of these strains were from the Tyee vineyard, while a single strain was from the Whistling Dog vineyard. Whistling Dog, like Mount Pleasant, has an off-premise winery.

Wild strains exhibit a clonal population structure, where many of the strains isolated have nearly identical genotypes, with no apparent geographical structure. Within wild strains, two clonal subpopulations (defined as a single clade in which the pairwise nucleotide p-value between any two strains within the group is less than 0.0002) contain 24 of the 27 strains in the group (Figure 4-4). The dominant clone (KEH00729, 20 isolates) is widespread, found in both vineyards and non-vineyard locations in Missouri and Oregon. The oak strains from Wisconsin (DY8 and DY9), as well as another US oak tree strain (T7) are also very closely related to KEH00729 (nucleotide p-distances are 0.00091, 0.00048 and 0.00119 respectively) and are also likely to be related to the dominant clone. The second subpopulation (KEH00411, 4 isolates) was found at both a vineyard and a non-vineyard location in Missouri. Wine strains, in contrast to the wild strains, have a less clonal structure with the exception of one clonal subpopulation (KEH02580) (Figure 4-3). These strains were all collected at Chaumette Vineyards in 2009. Six of the eight strains were isolated from grapes, and two from oak trees. Other clonal groups include NCYC110 and DBVPG6044 (W. African), the three strains from Ecuador, YPS606 and YPS1009-jf (Wild N. American), YJM975, YJM981 and YJM978 (wine), UWOPS05-227-2, UWOPS03-461-4 and UWOPS05-217-3 (Malaysian). Both the W. African and Malaysian clonal groups include all of the strains that make up the previously described ‘clean’ lineages (Liti et al. 2009) for those groups.

Despite low variation within clonal subgroups, the overall nucleotide diversity (π) within the wild strains (0.134) is higher than the diversity within the wine strains (0.064). In fact, the amount of nucleotide diversity within the wine lineage is only about 16% of that found in the entire sampled population and only 48% of that found in the wild strains. The reduction in diversity within wine strains is mostly due to synonymous sites, as evident by the ratio π_N/π_S (0.357), which is nearly twice as high as for the entire *S. cerevisiae* population (0.189, Chi-square $p < 0.001$),

and nearly three times as high as wild strains (0.131, Chi-square $p < 0.001$) (Table 4-12). The amount of diversity within wine strains at nonsynonymous sites (0.041) is nearly equal to that of wild strains (0.049). When comparing wine and wild strains, at synonymous sites 37% percent of the diversity is due to variation within wine or wild strains and 63% of the diversity is due to differentiation between the groups, whereas at nonsynonymous sites 41% of the diversity is within populations, and 59% is between (Chi-square $p < 0.001$). The ratio π_N/π_S between wine and wild strains (0.131) is similar to π_N/π_S within oak strains (0.151), but still significantly higher (Chi-square $p = 0.002$) (Table 4-12).

Allele frequencies in S. cerevisiae

We also measured the proportion of private alleles, both monomorphic and polymorphic within wine and wild strains. For this analysis we included wine and wild strains along with populations that do not show any evidence for admixture (e.g. Malaysian, West African, and Sake strains) (Liti et al. 2009). When comparing the total number of private alleles and their type (polymorphic or monomorphic) we noted a significant difference between the number of alleles at 4-fold degenerate and nondegenerate sites for both wine and wild strains (Chi-square test, $p < 0.001$ for wine strains and $p = 0.023$ for wild strains) (Figure 4-6). In wine strains, there is a significant excess in the number of private alleles at nondegenerate sites, whereas in wild strains there is a significant excess in the number of private alleles at 4-fold degenerate sites (Table 4-13). In wine strains, 35% of nonsynonymous private alleles are monomorphic, significantly higher than the 22% at 4-fold degenerate sites ($p = 0.017$) (Table 4-13). Private alleles in wild strains were almost always polymorphic, for both types of substitutions (Table 4-13).

We measured the minor allele frequency for wine and wild strains at both 4-fold degenerate sites and nondegenerate sites, and compared them to the neutral expectation (calculated based on Watterson's θ). In each case, the observed distribution was significantly different from the neutral expectation (Chi-square test, $p < 0.001$) (Figure 4-8).We observed an

excess of low frequency SNPs in wine strains for both 4-fold degenerate sites and nondegenerate sites, and an excess of higher frequency SNPs in wild strains at both 4-fold degenerate and nondegenerate sites. Because the frequency distribution of 4-fold degenerate sites did not match the neutral expectation, we compared 4-fold degenerate sites to nondegenerate sites, and found no significant differences.

Linkage Disequilibrium in S. cerevisiae

Linkage disequilibrium is known to vary between *S. cerevisiae* populations, and was previously reported to be high in lab strains, but low in wine strains (Schacherer et al. 2009). We used r^2 to measure linkage disequilibrium (which is a measure of LD that corrects for differences in allele frequencies) for the entire *S. cerevisiae* population as well as for wine and wild strains. We found LD decays to half of the maximum value (LD ½) within 2 kb in *S. cerevisiae* when all strains are considered, with a low level of LD at physically unlinked loci (r^2 = 0.075) (Figure 4-9). When wine strains and wild strains are considered independently, we observe different patterns. Wine strains show slightly more linkage when considered independently, with an LD ½ of 2.5 kb, and a slightly elevated level of linkage disequilibrium at physically unlinked loci (r^2 = 0.100). Although the wild strains have a LD ½ value of 2 kb, physically unlinked loci show a much higher level of linkage (r^2 = 0.196) (Figure 4-10). To account for the clonal nature of some isolates, LD was also measured within wine and wild strains with clonal isolates removed. LD ½ did not change for either population when clonal isolates were removed, but LD at physically unlinked loci decreased slightly in both cases, to 0.095 in wine strains and 0.182 in wild strains.

Population Structure within S. cerevisiae

An analysis of population structure was performed using the program STRUCTURE (Pritchard et al. 2000b) for 3,233 parsimony informative sites within *S. cerevisiae* strains. Ten replicate simulations were performed for each inferred number of populations (K), for K = 2 through K = 10. The replicate output with the highest ln likelihood for K=2 through K= 9, along

with the similarity index (G’), measured using CLUMPP (Jakobsson & Rosenberg 2007) is shown in Figure 4-12. The average likelihood continued to increase from values at K=2 with each additional population to K=10. However, the variance values in ln likelihood increased dramatically from less than 20% of the average likelihood at K2-9 to 500% of the average likelihood at K=10 (Figure 4-11), clearly indicating K=10 is an unlikely population structure.

For all simulations, both the wine/European lineage and the wild North American lineage were clearly differentiated. Replicate simulations at a given K value became less consistent above K=3, as indicated with decreased values for G’ (Figure 4-12), even though likelihood increased. Although values of similarity (G’) between replicate runs at K=8 and K=9 are roughly comparable, our results are congruent with a previous inference of population structure (Liti et al. 2009a) at K=7, providing resolution of sake, Malaysian, and West African lineages in addition to the wine/European and wild North American lineages. As such, our micro-scale sampling of yeast samples from vineyard and non-vineyard locations resolved previously identified subpopulations of *S. cerevisiae.*

Genetic Admixture and Introgression in S. cerevisiae

Individuals with genome sequence data from wine and wild populations, as defined by the population structure analysis described above, were used to define the wine and wild populations for the purpose of assessing genetic admixture of newly genotyped strains from North America. Potential introgression or admixture between the ‘wine’ and ‘wild’ genotypes was detected for a few of the US isolated strains and fell into two categories, those with predominately wine backgrounds, and those with predominately wild backgrounds (Figure 4-13).

Strains of *S. cerevisiae* sampled from cherries in Wisconsin are inferred to have a large proportion of admixture in their genetic background (64% wine background and 35% wild background). Similarly, four strains isolated from Chaumette Vineyard in Missouri are inferred to have a proportion of admixture from the wild genetic background. These strains include a single

isolate from a spontaneous wine fermentation, a single isolate from grape mash and a two isolates from vineyard grapes. The signature of introgression/admixture is preserved with the addition of out-groups using the unsupervised global population inference (Table 4-14).

Four strains with wild backgrounds (DY8, DY9, KEH00088 and KEH01135) were inferred to have between 5 and 15% wine ancestry (Table 4-14). For DY8 and DY9, strains isolated from oak trees in Wisconsin, this pattern is consistent with the results uncovered by the full unsupervised analysis, with only a small amount of potential admixture from a third population detected for DY8 (Table 4-14). The two other strains with oak background show moderate levels of introgression from the wine genotype (15%). However, this pattern is not upheld in the context of the global collection, as the inferred ancestry of these strains is preferentially assigned to a different population (other than wine or wild) with the addition of outgroups (Table 4-14).

Phylogenetic Analysis and Distribution of Genetic Diversity in S. paradoxus

A bootstrap consensus phylogeny tree was constructed using the neighbor-joining method, based on pairwise genetic distances (nucleotide p-value) between 66 OTUs at 281,944 nucleotide sites (Figure 4-14). Missing and ambiguous data were removed only for pairwise comparisons. Our analysis resolved the same four populations that complete genome sequencing uncovered: American, European, Far Eastern, and Hawaiian (represented by a single strain) (Liti et al. 2009a). Most of the strains isolated from Missouri and Oregon belong to the American clade. However, four of the strains isolated in this study as well as the control strain N17 seemed to form their own group that was somewhat related to both the European and Far Eastern clades (Liti et al. 2009). These isolates had poor alignment rates, resulting in a large proportion of missing data. We repeated the phylogenetic analysis with all missing data removed (Figure 4-15). Removing missing data resolved these strains as belonging to the European clade.

Isolates belonging to the American clade show further geographic structure. Strains from Missouri and strains from Oregon form two distinct clades, which are highly supported by

bootstrap analysis (Figure 4-16). Additionally, Oregon strains exhibit more extensive substructure compared to Missouri strains. Isolates from both vineyards are distinct from isolates from the non-vineyard locations (Figure 4-16). Four of the strains isolated in this study, three identical strains from Oregon and one from Missouri, were resolved as European genotypes (Figure 4-17).

Most of the diversity within *S. paradoxus* (93%) can be attributed to differentiation between groups rather than within groups (Table 4-15). Minor allele frequencies in *S. paradoxus* show a significant shift towards higher frequency alleles (Figure 4-18), likely due to population structure. When the American lineage is considered independently, we observe a slight but significant shift towards higher frequency alleles (Figure 4-18), which could, again, be influenced by population structure. Within the European lineage, there is a significant shift towards lower frequency alleles (Figure 4-18).

Population Structure in S. paradoxus

An analysis of population structure was performed using the program STRUCTURE (Pritchard et al. 2000) for 7,063 parsimony informative sites within *S. paradoxus* strains. Ten replicate simulations were performed for each inferred number of populations (K), for K = 2 through K = 15. The replicate output with the highest ln likelihood for K=2 through K= 4, along with the similarity index (G'), measured using CLUMPP (Jakobsson & Rosenberg 2007) is shown in Figure 4-19. The average likelihood increased from K=2 to K=4 with a similar and low amount of variance. However, when K = 5, the program failed to assign a fifth population, clearly indicating that 5 populations are highly unlikely. When K = 3, there is good resolution of the American, European, and Far Eastern populations. Increasing K to 4 increases the likelihood slightly. The major difference between 3 and 4 populations is the substructure between Missouri and Oregon strains from the American clade (Figure 4-19). The Hawaiian strain, in all cases appears to show a signal of genetic admixture, although it is likely that this is an artifact of sampling.

Discussion

In this study we investigate the hypothesis that strains of *S. cerevisiae* have been domesticated in association with the production of wine by examining the population genetics of *S. cerevisiae* isolated from different ecological niches within vineyard and non-vineyard locations in North America. First we show that distinct wine and wild populations of *S. cerevisiae,* which correspond to previously described European 'wine' genotypes and North American 'wild' genotypes (Liti et al. 2009) occur sympatrically within vineyards. However, wine stains are not established or do not persist in non-vineyard habitats. These two populations show major differences in population genetic parameters, indicating separate and distinct demographic histories. We provide evidence of gene flow between wine and wild yeast populations within vineyards, indicating a lack of physical or temporal barriers to gene exchange. In addition to *S. cerevisiae,* we also isolated populations of *S. paradoxus* within vineyard and non-vineyard habitats. Despite many similarities between *S. cerevisiae* and *S. paradoxus* we observe marked differences in their population genetic structure.

Genetic Differentiation between Saccharomyces cerevisiae Populations

The dominant genetic pattern observed within the areas that we sampled is the presence of two very distinct populations of *S. cerevisiae*; one that includes European 'wine' genotypes, and one that includes North American 'wild' genotypes. While there are several potential scenarios that could contribute to the population structure we observed, two likely mechanisms include the recent introduction or migration of allopatrically diverged isolates, or barriers to gene flow between locally adapted genotypes (Templeton 2006). In regard to the former scenario, a potential explanation for the persistence of distinct 'wine' and 'wild' populations in the United States is that a relatively recent dispersal of European winemaking strains has resulted in observed subdivision between 'wine' and 'wild' strains, and that given time this pattern will erode. This mechanism may be currently contributing to the observed population structure as the history

of US winemaking is relatively recent; commercial vineyards have been established within the last 300 years, and the wineries sampled in this study were established between 150 (Mount Pleasant Winery) and 20 (Chaumette Vineyards) years ago.

The second mechanism, genetic incompatibility or lack of gene flow between 'wine' and 'wild' strains, could be another explanation for the persistence of these distinct populations. Within the vineyard, as both wine and wild populations of *S. cerevisiae* were found on grapes and oak trees, there is no evidence for physical or temporal barriers to gene flow. In fact, we observed a signature of potential gene flow (5-10%) for several strains from Chaumette (Table 4-14), suggesting that gene flow between the European wine strains population and the wild North American population is possible. Moreover, we find that two strains isolated from cherries in Wisconsin appear to be recent hybrids. However, estimates of density, generation time and out-crossing rates within these habitats are highly variable (Murphy & Zeyl 2010), and preclude the ability to estimate the expected degree of admixture in these populations. *Saccharomyces* yeasts have mechanisms for both asexual and sexual reproduction that complicate the assessment of the frequency of gene flow from our data. For example, following sexual reproduction between wine and wild genotypes (gene flow), a resulting hybrid genotype would theoretically have a genetic signature of relatively equal proportions of admixture (~50:50). If this hybrid next propagated asexually through mitotic cell division, the hybrid genotype could persist for future generations at this level of admixture. However, if the hybrid undergoes sporulation, recombination between parental chromosomes during meiosis would decrease the signature of genetic admixture as well as heterozygosity in each resulting spore as a function of the recombination rate during meiosis. In this study we find support for both of these scenarios. The *S. cerevisiae* isolates from cherries, which have an admixture signature of roughly 40:60 may be indicative of recent hybridization or hybridization followed by asexual reproduction. In contrast, the vineyard strains with low levels of admixture, roughly 5-10%, could indicate hybridization

followed by sporulation. What is clear from our study is that there is no evidence for physical, temporal, or prezygotic genetic barriers between wine and wild genotypes in vineyards.

The apparent lack of barriers to gene flow between populations in the US raises interesting questions about the population genetic structure of *S. cerevisiae* within Europe, the source of most commercial strains (Johnston 1990). It is possible that the classically described wine strains (Fay & Benavides 2005; Liti et al. 2009) represent a specific 'wild' European genotype, and that the genetic bottleneck associated with those wine strains may have resulted from a founder event concurrent with the development and marketing of commercial starter strains, or through an artifact of sampling. However, the actual population structure of wild European strains remains an unanswered and open question, as wild European strains have not been broadly collected or extensively described. If wine strains and wild strains exist as distinct populations in wine producing regions of Europe that have been established for longer periods of time than the relatively young North American vineyards, it may be possible that the sympatric persistence of 'wine' and 'wild strains indicate fitness differences between strains or location adaptation to different environments.

Both wine and wild genotypes were sampled from vineyard grapes and vineyard oaks, yet only wild genotypes were isolated from non-vineyard locations. The presence of both genotypes in vineyards is most likely due to the migration of wine genotypes out of winery facilities and onto grapes and adjacent oak trees. The lack of wine genotypes isolated from non-vineyard locations may indicate that *S. cerevisiae* lacks sufficient dispersal ability to reach oak trees outside of vineyards. Very little is known about the dispersal range and mechanism for the movement of *S. cerevisiae* strains under normal conditions, although it has been postulated they are primarily transported by insects (Goddard et al. 2010; Mortimer & Polsinelli 1999), and there is evidence that they can be transported in oak barrels (Goddard et al. 2010). Another explanation for the restricted range of wine strains is that they are introduced seasonally, rather

than persist in the vineyard year round. Other studies have revealed that *S. cerevisiae* exists on grapes in high frequency only in the few weeks surrounding the grape harvest season (Valero et al. 2007), and that commercial wine making strains disseminate into the vineyard on a seasonal basis (Valero et al. 2005) which may limit the ability of wine genotypes to migrate to non-vineyard oaks. However, another study reported that commercial strains persist in the vineyard on a perennial basis (Schuller et al. 2005). There is evidence that *S. cerevisiae* can colonize wine cellars (Versavaud et al. 1995), but due to intense sanitation procedures employed in modern day wineries, the persistence of *S. cerevisiae* is unlikely (R. K. Mortimer 2000). The lack of wine genotypes isolated from non-vineyard locations could also suggest that wine genotypes are less fit than wild genotypes and unable to colonize the wild habitat. Although we did not evaluate fitness, previous studies have shown differentiation between wine and wild strains in freeze-thaw tolerance (Will et al. 2010), and suggest that other fitness differences may exist.

While the lack of migration ability is a reasonable explanation for the restriction of wine genotypes within vineyard locations, the complete lack of geographic structure within wild strains along with the recovery of the same clonal isolate from Missouri and Oregon locations suggests that dispersal may not be a limiting factor. However, it is possible that along with the distinct phenotypic (Chapters 1, 2 and 3) and genetic differences observed between wine and oak strains, migration ability has similarly diverged between these populations and contributes to the differences in their distribution.

A recent study conducted in New Zealand vineyards has also examined the population genetic structure of *S. cerevisiae* from spontaneous fermentations and a variety of substrates (e.g. spontaneous fermentation, vine bark, honeycomb, oak barrels) within vineyards (Goddard et al. 2010). Their results suggest that New Zealand strains are unique relative to previously described populations of *S. cerevisiae*, and provide evidence of population structure within vineyards. However, it is difficult to draw comparisons between this study and ours in relation to

the domestication of *S. cerevisiae* given that the methods used (microsatellite analysis of nine loci) resolved all New Zealand samples as discrete from previously differentiated *S. cerevisiae* populations (Liti et al. 2009).

Genetic Variation and Nucleotide Diversity within Saccharomyces cerevisiae Populations

We examined the level of genetic diversity, allelic frequencies and linkage disequilibrium for the global collection *S. cerevisiae,* and for both wine and wild populations independently. The results of this study build on previous reports of low diversity in wine strains compared to other *S. cerevisiae* populations (Fay & Benavides 2005; Liti et al. 2009), along with an excess in nonsynonymous polymorphism (Doniger et al. 2008) consistent with a historical population bottleneck. The excess in low frequency SNPs as well as relatively rapid decay in linkage disequilibrium support this scenario as well. The minor allele frequency is equally skewed at both neutral and non-neutral sites ($p = 0.200$), which provides no evidence for genome wide selection. The 22% of private nondegenerate alleles that are monomorphic within wine strains were likely fixed in the ancestral population, or through the combination of relaxed purifying selection and drift that typically accompany population bottlenecks (Templeton 2006).

Our results support previous studies that have reported a high amount of genetic diversity in wild strains relative to wine strains (Fay & Benavides 2005; Liti et al. 2009). However, we also find that the population structure of wild *S. cerevisiae* strains is dominated by several clones that exhibit no geographical structure. The skew towards higher frequency alleles and increase in linkage disequilibrium within wild strains is indicative of a recent population bottleneck with little subsequent increase in size. Whereas a population bottleneck is expected for the putative domestication of wine strains, the genetic signature of a recent bottleneck in wild North American strains is curious. This pattern could reflect the recent introduction of wild strains into North America, but also raises the possibility that strains of *S. cerevisiae* may not be well suited to the

oak tree habitat and that relatively few wild genotypes are capable of persisting at any given time. Specific wild genotypes that are fit in this environment could have experienced a range expansion, resulting in the apparent genetic signature of clonal sweeps. This pattern is frequently observed in populations of fungal pathogens (McDonald & Linde 2002) as well as clonal bacterial populations (Spratt & Maiden 1999), and has recently been show to occur in populations of *E. coli* even under a constant environment (Maharjan et al. 2006). It is also possible, however, that this pattern has been generated through some other form of cryptic population structure, or other neutral demographic processes. Each case raises interesting questions about the persistence and relative stability of 'wild' strains in natural habitats. Future studies of the temporal dynamics of wild populations could provide data pertinent to the hypothesis of *S. cerevisiae* domestication.

Differences in Population Genetic Structure between Saccharomyces cerevisiae and Saccharomyces paradoxus

Saccaromyces cerevisiae and *S. paradoxus* are nearly phenotypically indistinguishable and share complete chromosomal synteny (Dujon 2010). As demonstrated here and in previous studies, they show very different patterns of genetic diversity and population structure (Johnson et al. 2004; Koufopanou et al. 2006; Liti et al. 2009; G. I. Naumov et al. 1997). Of particular note is the correlation of genetic diversity with geographic distance observed in *S. paradoxus,* and the presence of genetic barriers between allopatrically diverged populations (Sniegowski et al. 2002). Similar to previous studies (Liti et al. 2009) the level of genetic diversity we observed within *S. paradoxus* was approximately 5 times greater than for *S. cerevisiae*. Whereas 93% of the overall variation was found between populations for *S. paradoxus*, only 63% of variation was found between wine and wild populations of *S. cerevisiae*. The pattern of genetic diversity observed in *S. paradoxus* is congruent with isolation between continents (i.e. North America, Europe, Asia), as previously reported (Liti et al. 2009; Johnson et al. 2004), but this study provides additional evidence demonstrating genetic differentiation in *S. paradoxus* associated with geographical

distance within a continent, specifically North America. *Saccharomyces paradoxus* isolates from Missouri and Oregon formed well supported groups within North American isolates and there is some support for geographic substructure within Oregon as well.

We also observe a major difference between *S. cerevisiae* and *S. paradoxus* regarding the movement of genes between populations. Four of the *S. paradoxus* strains isolated from Missouri and Oregon were found to cluster with European *S. paradoxus,* suggesting migration of European isolates into the US. While we observe admixture between the European (wine) and North American (wild) genotypes of *S. cerevisiae*, we find no evidence for genetic exchange between European and American *S. paradoxus* genotypes, although sample sizes are small for *S. paradoxus*. The migration of European *S. paradoxus* isolates and their genetic isolation from N. American strains has been observed previously in the North and Eastern US and Canada (Kuehne et al. 2007) and may be indicative of allopatric divergence leading to speciation. Indeed, hybrids between *S. paradoxus* strains from different geographical origins show a significant decrease in spore viability, indicating partial reproductive isolation (Sniegowski et al. 2002).

Differences observed between the two species extend beyond genetic structure; they also show evidence of ecological differentiation. *Saccharomyces paradoxus* is found in association with oak trees, soil and decaying leaf material, the same habitats in which *S. cerevisiae* can be found (Johnson et al. 2004), and is thought to represent the ancestral state for *S. cerevisiae.* However, *S. cerevisiae* can only be isolated from a portion of the geographical range of *S. paradoxus*. In this study both species were found in abundance from samples collected in Missouri, but *S. cerevisiae* was nearly absent from most Oregon locations (Chapter 3). Geographical restriction of *S. cerevisiae* has also been observed in Europe. For example, both *S. cerevisiae* and *S. paradoxus* were isolated from tree bark in Portugal, while only *S. paradoxus* could be recovered from tree bark in Germany (Sampaio & Gonçalves 2008).

A potential explanation for differences in habitat use between the two species is that *S. paradoxus* may be able to tolerate a wider range of environmental stresses associated with oak tree habitats than *S. cerevisiae*. If *S. paradoxus* is indeed more generalized in its ability to tolerate environmental stresses, we would expect it to be able to inhabit a wider geographic distribution, and as a result carry more genetic diversity than *S. cerevisiae*. In fact, this is the pattern we observed for *S. paradoxus*. Previous studies have also demonstrated the relatively higher levels of genetic diversity for *S. paradoxus* as compared to *S. cerevisiae* (Liti et al. 2009; Johnson et al. 2004). In order to discern the relationship between environmental amplitude and genetic structure, future studies are needed to quantify the range of environments in which *S. cerevisiae* and *S. paradoxus* are capable of persisting.

Complicated by unknown rates of sexual and asexual reproduction, dispersal vectors, and temporal variation in persistence, the expected genetic structure of fungal populations and specifically of domesticated fungi are unknown. Despite clear differences in population structure within *S. cerevisiae* and between *S. cerevisiae* and *S. paradoxus*, there are few studies to which we can draw parallels. The only other study of fungal domestication (*Aspergillus oryzae)* found relatively little genetic distinction between wild and domesticated populations. However, we may be able to draw some inferences from studies of plants and animals. Signatures of genetic bottlenecks are common in domesticated plants and animals and the observation of increases in nonsynonymous polymorphism has been documented for other species including rice (Lu et al. 2006) and dogs (Cruz et al. 2008). Relevant comparisons may also come from studies of population structure differences between *Arabidopsis thaliana* and *Arabidopsis lyrata*. Although not classically domesticated, *A. thaliana* is an important of a model system, comparable to *S. cerevisiae*. Like *S. cerevisiae*, the global distribution and genetic structure of *A. thaliana* has also been heavily influenced by human migration and land use change (Beck et al. 2007). The two species also exhibit major differences in population genetic structure, analogous to the relationship between *S. paradoxus* and *S. cerevisiae.* Similar to *S. cerevisiae, A. thaliana* has

much lower genetic diversity compared to its sister species (Nordborg et al. 2005), and exhibits much higher levels of linkage disequilibrium (S. Kim et al. 2007), both of which can be explained by differences in mating system (Ross-Ibarra et al. 2008). While *A. thaliana* reproduces mostly through self-fertilization and thus has many similarities to clonally propagating species (Abbott & Gomes 1989), *A. lyrata* is mostly self-incompatible and reproduces sexually (Mable & Adam 2007). Although differences in mating system can cause the differences in levels of diversity and linkage disequilibrium, they cannot explain the fact that *A. thaliana* shows a significant excess in low frequency alleles (Nordborg et al. 2005), whereas *A. lyrata* shows a skew towards higher frequency alleles (Ross-Ibarra et al. 2008). While the pattern in *A. thaliana* is consistent with population bottlenecks followed by continuing population expansion, demographic modeling has been used to infer that the skew towards higher frequency alleles in *A. lyrata* is indicative of population bottlenecks with little or no recovery (Ross-Ibarra et al. 2008). These studies demonstrate non equilibrium demographic processes that have shaped population level patterns of diversity within species, causing population genetic patterns that in some ways resemble those of *S. cerevisiae* and *S. paradoxus.* In the future, global population level sampling in *Saccharomyces* may be used in a similar way to infer demographic history.

Conclusions

This study represents one of the first examinations of genome wide population level differentiation within *Saccharomyces* species in a single ecological context. Distinct wine and wild populations of *S. cerevisiae* are observed within vineyards and each population has unique differences in genetic variation and nucleotide diversity. Results suggest a population bottleneck followed by population growth of wine strains, supporting the hypothesis of domestication. Although the genetic diversity within wild strains is much higher than in wine strains, the allelic frequencies within wild strains are consistent with a recent population bottleneck. This signature combined with the clonal nature of wild strains suggests possible environmental restriction or selection against certain genotypes in the wild, although this pattern could also be generated

through neutral processes. We find evidence for genetic exchange between the populations which may suggest that local adaptation is not the primary driving force of genetic differentiation between the populations. However, wine genotypes are restricted to vineyard locations, which may be indicative of fitness differences, and it remains to be seen whether gene flow between the populations results in individuals that are less fit. It is clear that *S. paradoxus* and *S. cerevisiae,* despite their similarities, have dramatically different population structure even in the same environment. Future studies of *S. cerevisiae* including increased global sampling, especially of European populations will be critical to assess the degree to which local adaptation or domestication is responsible for the presence of distinct populations of *S. cerevisiae.*

Table 4-1. RAD genotyped *S. cerevisiae* strains collected in this study.

Name	Collection Site	Location	Year	Source
KEH00012	Chaumette Vineyards	Missouri, USA	2008	*Q. stellata* (twig)
KEH00088	Chaumette Vineyards	Missouri, USA	2008	*Q. stellata* (soil)
KEH00221	Chaumette Vineyards	Missouri, USA	2008	*Q. stellata* (bark)
KEH00290	Chaumette Vineyards	Missouri, USA	2008	Chardonel grape
KEH00400	Chaumette Vineyards	Missouri, USA	2008	Chardonel grape
KEH00411	Chaumette Vineyards	Missouri, USA	2008	Chardonel grape
KEH00415	Chaumette Vineyards	Missouri, USA	2008	Chardonel grape
KEH00463	Mount Pleasant Winery	Missouri, USA	2008	*Quercus* spp. (soil)
KEH00497	Mount Pleasant Winery	Missouri, USA	2008	*Quercus* spp. (bark)
KEH00673	Mount Pleasant Winery	Missouri, USA	2008	Vidal Blanc grape
KEH00729	Mount Pleasant Winery	Missouri, USA	2008	Vidal Blanc grape
KEH01027	Tyson Research Center	Missouri, USA	2008	*Q. alba* (bark)
KEH01135	Tyson Research Center	Missouri, USA	2008	*Q. rubra* (soil)
KEH01146	Tyson Research Center	Missouri, USA	2008	*Q. rubra* (twig)
KEH01172	Washington County	Missouri, USA	2008	*Q. alba* (bark)
KEH01205	Washington County	Missouri, USA	2008	*Q. velutina* (soil)
KEH01267	Washington County	Missouri, USA	2008	*Q. alba* (bark)
KEH01422	Washington County	Missouri, USA	2008	*Q. rubra* (soil)
KEH01639	Whistling Dog Cellars	Oregon, USA	2008	*Q. garryana* (soil)
KEH01876	Tyee Wine Cellars	Oregon, USA	2008	*Q. garryana* (soil)
KEH01958	Tyee Wine Cellars	Oregon, USA	2008	*Q. garryana* (soil)
KEH02439	Benton County	Oregon, USA	2008	*Q. garryana* (soil)
KEH02441	Benton County	Oregon, USA	2008	*Q. garryana* (bark)
KEH02503	Benton County	Oregon, USA	2008	*Q. garryana* (bark)
KEH02509	Benton County	Oregon, USA	2008	*Q. garryana* (twig)
KEH02518	Benton County	Oregon, USA	2008	*Q. garryana* (twig)
KEH02575	Chaumette Vineyards	Missouri, USA	2008	Spontaneous fermentation
KEH02580	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02583	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02587	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape

Name	Collection Site	Location	Year	Source
KEH02588	Chaumette Vineyards	Missouri, USA	2009	Norton grape
KEH02595	Chaumette Vineyards	Missouri, USA	2009	Traminette grape
KEH02635	Chaumette Vineyards	Missouri, USA	2009	*Q. stellata* (bark)
KEH02707	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02714	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02724	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02773	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02809	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02884	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape mash
KEH02887	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape mash
KEH02926	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH02978	Chaumette Vineyards	Missouri, USA	2009	*Q. stellata* (soil)
KEH03027	Tyson Research Center	Missouri, USA	2009	*Q. rubra* (soil)
KEH03066	Tyson Research Center	Missouri, USA	2009	*Q. alba* (bark)
KEH01091	Tyson Research Center	Missouri, USA	2008	*Q. marilandica* (soil)

Table 4-2. RAD genotyped *S. paradoxus* strains collected in this study.

Name	Collection Site	Location	Year	Source
KEH00137	Chaumette Vineyards	Missouri, USA	2008	*Q. stellata* (bark)
KEH00160	Chaumette Vineyards	Missouri, USA	2008	*Q. alba* (soil)
KEH00197	Chaumette Vineyards	Missouri, USA	2008	*Q. acutissima* (bark)
KEH00458	Mount Pleasant Winery	Missouri, USA	2008	*Quercus* spp. (bark)
KEH00489	Mount Pleasant Winery	Missouri, USA	2008	*Quercus* spp. (soil)
KEH00537	Mount Pleasant Winery	Missouri, USA	2008	*Quercus* spp. (soil)
KEH00703	Mount Pleasant Winery	Missouri, USA	2008	Vidal Blanc grape
KEH00831	Tyson Research Center	Missouri, USA	2008	*Q. alba* (twig)
KEH00973	Tyson Research Center	Missouri, USA	2008	*Q. rubra* (bark)
KEH01169	Washington County	Missouri, USA	2008	*Q. alba* (bark)
KEH01258	Washington County	Missouri, USA	2008	*Q. alba* (soil)
KEH01348	Washington County	Missouri, USA	2008	*Q. velutina* (bark)
KEH01447	Washington County	Missouri, USA	2008	*Q. rubra* (soil)
KEH01508	Washington County	Missouri, USA	2008	*Q. velutina* (bark)
KEH01547	Whistling Dog Cellars	Oregon, USA	2008	*Q. garryana* (bark)
KEH01619	Whistling Dog Cellars	Oregon, USA	2008	*Q. garryana* (soil)
KEH01684	Whistling Dog Cellars	Oregon, USA	2008	*Q. garryana* (twig)
KEH01764	Whistling Dog Cellars	Oregon, USA	2008	*Q. garryana* (bark)
KEH01830	Whistling Dog Cellars	Oregon, USA	2008	*Q. garryana* (soil)
KEH01860	Tyee Wine Cellars	Oregon, USA	2008	*Q. garryana* (bark)
KEH01903	Tyee Wine Cellars	Oregon, USA	2008	*Q. garryana* (soil)
KEH01967	Tyee Wine Cellars	Oregon, USA	2008	*Q. garryana* (twig)
KEH02054	Tyee Wine Cellars	Oregon, USA	2008	*Q. garryana* (bark)
KEH02126	Tyee Wine Cellars	Oregon, USA	2008	Pinot Noir grape
KEH02128	Chip Ross Park	Oregon, USA	2008	*Q. garryana* (bark)
KEH02166	Chip Ross Park	Oregon, USA	2008	*Q. garryana* (soil)
KEH02219	Chip Ross Park	Oregon, USA	2008	*Q. garryana* (twig)
KEH02271	Chip Ross Park	Oregon, USA	2008	*Q. garryana* (bark)

Name	Collection Site	Location	Year	Source
KEH02367	Chip Ross Park	Oregon, USA	2008	*Q. garryana* (soil)
KEH02391	Benton County	Oregon, USA	2008	*Q. garryana* (twig)
KEH02446	Benton County	Oregon, USA	2008	*Q. garryana* (soil)
KEH02492	Benton County	Oregon, USA	2008	*Q. garryana* (twig)
KEH02499	Benton County	Oregon, USA	2008	*Q. garryana* (bark)
KEH02530	Benton County	Oregon, USA	2008	*Q. garryana* (soil)
KEH02647	Chaumette Vineyards	Missouri, USA	2009	*Q. stellata* (bark)
KEH02801	Chaumette Vineyards	Missouri, USA	2009	Chardonel grape
KEH03015	Tyson Research Center	Missouri, USA	2009	*Q. alba* (bark)
KEH03086	Tyson Research Center	Missouri, USA	2009	*Q. velutina* (soil)

Table 4-3. Additional RAD genotyped strains.

Name	Location	Year	Source
M22	Italy	NA	Vineyard
YPS163	Pennsylvania, USA	1999	Oak exudate
DCM21	Wisconsin, USA	2009	Cherry
DCM6	Wisconsin, USA	2009	Cherry
DY8	Wisconsin, USA	2009	Oak
DY9	Wisconsin, USA	2009	Oak
CLQCA_10_084	Ecuador		Ancient wine fermentation vessel
CLQCA_10_097	Ecuador		Ancient wine fermentation vessel
CLQCA_10_100	Ecuador		Ancient wine fermentation vessel

Table 4-4. Published *S. cerevisiae* genomes (SGRP) used in this study.

Name	Geographical Origin	Source
273614N	Royal Victoria Infirmary, Newcastle UK	Clinical isolate (Fecal)
322134S	Royal Victoria Infirmary, Newcastle UK	Clinical isolate (Throat-sputum
378604X	Royal Victoria Infirmary, Newcastle UK	Clinical isolate (sputum)
BC187	Napa Valley, CA, USA	Barrel fermentation
DBVPG1106	Australia, 1947	Grapes
DBVPG1373	Netherlands, 1952	Soil
DBVPG1788	Turku, Finland, 1957	Soil
DBVPG1853	Ethiopia, 1959	White Teff
DBVPG6040	Netherlands, 1947	Fermenting fruit juice
DBVPG6044	West Africa, 1925	Bili wine, from *Osbeckia grandiflora*
DBVPG6765	Unknown	Unknown
K11	Japan, 1981	Shochu sake strain
L_1374	Cauquenes, Chile, 1999	Fermentation from must Pais
L_1528	Cauquenes, Chile, 1999	Fermentation from must Cabernet
NCYC110	West Africa, pre-1914	Ginger beer from *Z. officinale*
NCYC361	Ireland, 1952	Beer spoilage strain from wort
RM11_1A		
S288c	Merced, California, USA, 1938	Rotting fig (laboratory strain)
SK1	USA, pre-1974	Soil (laboratory strain)
UWOPS03_461_4	Telok Senangin, Malaysia, 2003	Nectar, Bertram palm
UWOPS05_217_3	Telok Senangin, Malaysia, 2005	Nectar, Bertram palm
UWOPS05_227_2	Telok Senangin, Malaysia, 2005	*Trigona* spp (Stingless bee) collected near Bertam palm flower
UWOPS83_787_3	Great Inagua Island, Bahamas, 1983	Fruit, *Opuntia stricta*
UWOPS87_2421	Puhelu Road, Maui, Hawaii, 1987	Cladode, *Opuntia stricta*
W303	Lab strain from multiple crosses	Laboratory strain
Y9	Indonesia, pre-1962	Ragi (similar to sake wine)
Y12	Ivory Coast, pre-1981	Palm wine strain
Y55	France, between 1930-1960	Grape

Name	Geographical Origin	Source
YIIc17_E5	Sauternes, France	Wine
YJM789		
YJM975	Ospediali Riuniti di Bergamo, Italy, 1994-6	Isolated from vagina of patient suffering from vaginitis
YJM978	Ospediali Riuniti di Bergamo, Italy, 1994-6	Isolated from vagina of patient suffering from vaginitis
YJM981	Ospediali Riuniti di Bergamo, Italy, 1994-6	Isolated from vagina of patient suffering from vaginitis
YPS128	Pennsylvania, USA, 1999	Bark of *Q. rubra*
YPS606	Pennsylvania, USA, 1999	Soil beneath *Q. alba*
YS2	Australia	Baker strain
YS4	Netherlands, 1975	Baker strain
YS9	Singapore	Baker strain

For a complete description of these strains, see (Liti et al. 2009) and references within.

Table 4-5. Published *S. paradoxus* genomes (SGRP) used in this study.

Name	Geographical Origin	Source
A4*	Mont St-Hilaire, Quebec, Canada, 2003	Bark of *Q. rubra*
A12*	Mont St-Hilaire, Quebec, Canada, 2003	Soil beneath *Q. rubra*
CBS432	Moscow area, Russia, pre-1931	Bark of *Quercus* spp
CBS5829	Denmark, pre-1967	Mor soil, pH 3.6
DBVPG4650	Marche, Italy, pre-1992	Fossilized guano in a cavern
DBVPG6304*	Yosemite, California, USA	*Drosophila pseudoobscura*
IFO1804	Japan	Bark of *Quercus* spp
KPN3828	Novosibirsk, Siberia, Russia, 2003	Bark of *Q. robur*
KON3829	Novosibirsk, Siberia, Russia, 2003	Bark of *Q. robur*
N.17	Tartasan, Russia	Exudate of *Q. robur*
N.43	Vladivostok, Russia	Exudate of *Q. mongolica*
N.44	Ternei, Russia, 1987	Exudate of *Q. mongolica*
N.45	Ternei, Russia, 1987	Exudate of *Q. mongolica*
Q31.4	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q32.3	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q59.1	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q62.5	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q69.8	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q74.4	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q89.8	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q74.4	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q89.8	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
Q95.3	Windsor Great Park, UK, 1998	Bark of *Quercus* spp
S36.7	Silwood Park, UK, 1997	Bark of *Quercus* spp
T21.4	Silwood Park, UK, 1998	Bark of *Quercus* spp
UFRJ50791*	Catalao point, Rio de Janeiro, Brazil, pre-1992	*Drosophila* spp
UFRJ50816*	Tijuca Forest, Rio de Janeiro, Brazil, pre-1992	*Drosophila* spp
UWOPS91.917.1	Saddle Road, Island of Hawaii, 1991	Flux of *Myoporum sandwichense*
W7	Silwood Park, UK, 1996	Bark of *Quercus* spp

Name	Geographical Origin	Source
Y6.5	Silwood Park, UK, 2003	Bark of *Quercus* spp
Y7	Silwood Park, UK, 2003	Bark of *Quercus* spp
Y8.1	Silwood Park, UK, 2003	Bark of *Quercus* spp
Y8.5	Silwood Park, UK, 2003	Bark of *Quercus* spp
Y9.6	Silwood Park, UK, 2003	Bark of *Quercus* spp
YPS138*	Pennsylvania, USA	Soil beneath *Q. velutina*
Z1	Silwood Park, UK, 2003	Bark of *Quercus* spp
Z1.1	Silwood Park, UK, 2003	Bark of *Quercus* spp

For a complete description of these strains, see (Liti et al. 2009) and references within.

Strains used to construct a reference assembly for alignments in this study are indicated with an asterisk (*).

Table 4-6. Unpublished genome sequences used in this study.

Strain	Geographical Origin	Source
CBS7960	South Africa	Sugar cane
CLIB215	New Zealand	Baker strain
CLIB324	Vietnam	Baker strain
CLIB382	Japan	Beer
FL100	NA	Lab strain
I14	Italy	Vineyard
IL-01	US	Soil
NC-02	US	Forest
PW5	Nigeria	Palm wine
T7	US	Oak tree
T73	Spain	Wine
UC5	Japan	Sake
Y9	Indonesia	Ragi
WE372	South Africa	Wine
Y10	Philippines	Coconut
Y12	Africa	Palm wine
YJM269	NA	Apple juice fermentation
YJM320	USA	Clinical
YJM326	USA	Clinical
YJM428	USA	Clinical
YJM451	Europe	Clinical
YJM653	NA	Clinical
YPS1009	USA	Oak tree
M22	Italy	Vineyard
YPS163	Pennsylvania, USA	Oak exudate

Table 4-7. Alignment and sequencing statistics for *S. cerevisiae.*

ID	Reads	% Aligned	Sequenced Positions (bp)	% Filtered	Remaining positions (bp)	Average sequence coverage	Average quality
CLQCA_10_084	744803	86%	434486	18%	354473	117	218
CLQCA_10_097	447495	84%	410173	16%	344113	71	190
CLQCA_10_100	899652	85%	451235	20%	359163	139	222
DCM21	540007	88%	436489	15%	372117	83	192
DCM6	364717	88%	427882	15%	365311	57	171
DY8	817371	86%	468409	19%	379327	121	209
DY9	51220	81%	352333	32%	240913	10	58
KEH00012	218520	81%	426100	27%	312927	36	116
KEH00088	790866	81%	464738	18%	381173	106	163
KEH00221	480630	80%	487901	26%	360347	68	142
KEH00290	69438	83%	375359	40%	225809	15	71
KEH00400	784811	81%	474681	20%	382004	104	163
KEH00411	787637	81%	519526	23%	398496	101	158
KEH00415	827893	81%	474187	18%	386679	111	165
KEH00463	287276	82%	435430	26%	322862	45	127
KEH00497	395530	81%	449232	24%	342104	59	136
KEH00673	478797	81%	501518	25%	376146	65	140
KEH00729	624242	81%	505696	23%	387251	82	150
KEH01027	117853	83%	400262	33%	268953	22	92
KEH01091	1758937	88%	467530	17%	389691	258	225

ID	Reads	% Aligned	Sequenced Positions (bp)	% Filtered	Remaining positions (bp)	Average sequence coverage	Average quality
KEH01135	754143	80%	514269	23%	397378	96	157
KEH01146	485921	82%	457729	21%	362246	70	144
KEH01172	953849	80%	474988	19%	385929	125	171
KEH01205	848016	81%	465036	17%	386765	113	166
KEH01267	362434	81%	471716	26%	349874	53	132
KEH01422	819127	81%	514755	23%	397716	107	163
KEH01639	691219	82%	505189	23%	389790	94	157
KEH01876	586346	82%	478696	22%	374874	82	151
KEH01958	208456	81%	435217	28%	311912	34	113
KEH02439	200575	83%	432207	30%	303199	34	116
KEH02441	357667	82%	446406	24%	339958	54	133
KEH02503	663637	81%	477837	21%	377213	90	155
KEH02509	848441	80%	516327	22%	400978	108	161
KEH02518	829938	81%	511592	21%	402904	106	161
KEH02575	472797	81%	491026	25%	367162	66	140
KEH02580	240211	83%	449336	29%	319693	39	121
KEH02583	571871	82%	486543	22%	378315	79	148
KEH02587	392378	83%	437609	20%	348534	60	138
KEH02588	495003	83%	472574	22%	367179	71	143
KEH02595	847311	83%	482327	19%	389603	114	165
KEH02635	214653	83%	435586	28%	312462	36	117
KEH02707	444002	82%	471884	23%	361918	64	139

ID	Reads	% Aligned	Sequenced Positions (bp)	% Filtered	Remaining positions (bp)	Average sequence coverage	Average quality
KEH02714	646280	84%	452718	19%	368235	93	157
KEH02724	278386	83%	443132	25%	331544	44	125
KEH02773	807256	81%	467908	18%	381987	109	165
KEH02809	1073380	81%	500052	21%	393450	141	174
KEH02884	727963	81%	476035	19%	384223	98	159
KEH02887	414134	82%	432076	19%	348184	63	140
KEH02926	395743	82%	444360	22%	348581	58	137
KEH02978	632433	82%	446086	18%	364581	90	156
KEH03027	417207	80%	477990	25%	357823	59	136
KEH03066	422643	81%	476704	24%	361293	60	137
M22 (control)	281575	82%	421124	24%	319812	46	126
YPS163 (control)	752443	82%	477289	20%	381375	103	161
Average	567132	82%	461250	23%	357714	80	150

Table 4-8. Alignment and sequencing statistics for *S. paradoxus.*

ID	Reads	% Aligned	Sequenced positions (bp)	Percent filtered	Remaining positions (bp)	Average sequence coverage	Average quality
KEH00137	534560	50%	286995	35%	185356	82	189
KEH00160	434	88%	21771	100%	59	5	42
KEH00197	351871	90%	397762	20%	318545	57	171
KEH00458	2527	90%	100325	99%	527	6	46
KEH00489	7809	91%	203845	95%	9948	6	46
KEH00537	205709	89%	343379	14%	295788	35	131
KEH00703	146575	89%	336233	14%	288676	26	103
KEH00831	1121603	88%	381733	15%	325807	175	223
KEH00973	531780	89%	364780	12%	320365	85	197
KEH01169	513146	90%	399517	19%	325538	82	197
KEH01258	182075	0%	2596	93%	194	38	128
KEH01348	289043	90%	345631	14%	298167	50	164
KEH01447	103972	90%	328549	16%	274606	19	84
KEH01508	786707	89%	382742	14%	327313	122	212
KEH01547	269518	88%	363994	17%	303096	45	152
KEH01619	1054261	87%	386902	15%	330185	161	222
KEH01684	410560	88%	362777	14%	312462	66	180
KEH01764	771663	88%	370812	12%	325183	121	212
KEH01830	4848	88%	155937	99%	1429	8	52
KEH01860	556	84%	25901	100%	59	6	45

ID	Reads	% Aligned	Sequenced positions (bp)	Percent filtered	Remaining positions (bp)	Average sequence coverage	Average quality
KEH01903	18	36%	649	100%	0	0	0
KEH01967	77102	88%	324437	20%	259504	14	70
KEH02054	520698	89%	368105	15%	312518	86	202
KEH02126	1017338	89%	385649	14%	333356	155	219
KEH02128	257	45%	6785	100%	0	0	0
KEH02166	480901	90%	365269	12%	319671	77	191
KEH02219	339320	50%	264616	34%	173447	55	170
KEH02271	277736	50%	253168	33%	170479	46	153
KEH02367	760617	87%	389041	17%	323874	117	213
KEH02391	5312	90%	168032	98%	3044	6	45
KEH02446	888494	1%	11977	99%	173	2305	194
KEH02492	950234	87%	410613	19%	333338	142	217
KEH02499	412063	87%	366860	14%	315383	65	181
KEH02530	487278	49%	278693	35%	181693	75	185
KEH02647	1221430	89%	387927	15%	331446	189	226
KEH02801	558219	90%	368444	15%	313620	92	204
KEH03015	1286221	89%	410712	18%	337337	196	224
KEH03086	951729	89%	382266	15%	326283	150	223
N17 (control)	467420	50%	276623	35%	180916	73	186
YPS138 (control)	1041209	88%	391307	15%	333202	159	218
Average	475820	77%	284334	38%	219815	130	153

Table 4-9. SNPS and heterozygosity for *S. cerevisiae.*

ID	Number of loci (total bp)1	Variable loci (SNPs)2	Heterozygous loci	% Heterozygous loci
CLQCA_10_084	434486	1738	137	0.03%
CLQCA_10_097	410173	1645	113	0.03%
CLQCA_10_100	451235	1793	148	0.03%
DCM21	436489	1812	859	0.20%
DCM6	427882	1801	950	0.22%
DY8	468409	1655	24	0.01%
DY9	352333	1054	4	0.00%
KEH00012	426100	1506	6	0.00%
KEH00088	464738	2022	26	0.01%
KEH00221	487901	1221	9	0.00%
KEH00290	375359	732	0	0.00%
KEH00400	474681	1946	21	0.00%
KEH00411	519526	2008	39	0.01%
KEH00415	474187	1446	177	0.04%
KEH00463	435430	1592	5	0.00%
KEH00497	449232	1674	7	0.00%
KEH00673	501518	1888	13	0.00%
KEH00729	505696	1958	20	0.00%
KEH01027	400262	1275	5	0.00%
KEH01091	467530	1966	36	0.01%
KEH01135	514269	1996	27	0.01%
KEH01146	457729	1818	18	0.00%
KEH01172	474988	1953	22	0.00%
KEH01205	465036	1952	24	0.01%
KEH01267	471716	1712	8	0.00%
KEH01422	514755	1976	31	0.01%
KEH01639	505189	1300	20	0.00%
KEH01876	478696	1223	10	0.00%
KEH01958	435217	1003	2	0.00%
KEH02439	432207	1466	6	0.00%

ID	Number of loci (total bp)1	Variable loci (SNPs)2	Heterozygous loci	% Heterozygous loci
KEH02441	446406	1672	8	0.00%
KEH02503	477837	1897	20	0.00%
KEH02509	516327	2033	37	0.01%
KEH02518	511592	2037	28	0.01%
KEH02575	491026	1366	307	0.06%
KEH02580	449336	1069	7	0.00%
KEH02583	486543	1234	17	0.00%
KEH02587	437609	1166	5	0.00%
KEH02588	472574	1257	15	0.00%
KEH02595	482327	1971	26	0.01%
KEH02635	435586	1014	4	0.00%
KEH02707	471884	1793	17	0.00%
KEH02714	452718	1263	9	0.00%
KEH02724	443132	1102	5	0.00%
KEH02773	467908	1912	18	0.00%
KEH02809	500052	1373	19	0.00%
KEH02884	476035	1389	9	0.00%
KEH02887	432076	1169	6	0.00%
KEH02926	444360	1726	8	0.00%
KEH02978	446086	1243	5	0.00%
KEH03027	477990	1753	9	0.00%
KEH03066	476704	1775	20	0.00%
M22	421124	1026	4	0.00%
YPS163	477289	1923	23	0.00%
Average	461250	1580	63	0.01%

[1] Total sites after quality filtering

[2] SNPs are defined in relation to the *S. cerevisiae* reference genome from (Liti et al. 2009).

Table 4-10. SNPs and heterozygosity for *S. paradoxus.*

ID	Number of loci (total bp)[1]	Variable loci (SNPS)[2]	Heterozygous SNPS	% Heterozygous loci
KEH00137	185356	3308	5	0.003%
KEH00160*	59	1	0	0.000%
KEH00197	318545	544	2	0.001%
KEH00458*	527	3	0	0.000%
KEH00489*	9948	20	2	0.020%
KEH00537	295788	442	4	0.001%
KEH00703	288676	485	4	0.001%
KEH00831	325807	559	4	0.001%
KEH00973	320365	525	3	0.001%
KEH01169	325538	536	5	0.002%
KEH01258*	194	5	0	0.000%
KEH01348	298167	459	3	0.001%
KEH01447	274606	454	9	0.003%
KEH01508	327313	551	6	0.002%
KEH01547	303096	547	8	0.003%
KEH01619	330185	614	9	0.003%
KEH01684	312462	571	7	0.002%
KEH01764	325183	625	11	0.003%
KEH01830*	1429	5	0	0.000%
KEH01860*	59	0	0	0.000%
KEH01903*	0	0	0	0.000%
KEH01967	259504	483	3	0.001%
KEH02054	312518	585	5	0.002%
KEH02126	333356	653	7	0.002%
KEH02128*	0	0	0	0.000%
KEH02166	319671	568	6	0.002%
KEH02219	173447	3010	3	0.002%
KEH02271	170479	2943	2	0.001%
KEH02367	323874	583	5	0.002%
KEH02391*	3044	5	0	0.000%

ID	Number of loci (total bp)[1]	Variable loci (SNPS)[2]	Heterozygous SNPS	% Heterozygous loci
KEH02446*	173	9	0	0.000%
KEH02492	333338	583	4	0.001%
KEH02499	315383	596	6	0.002%
KEH02530	181693	3194	2	0.001%
KEH02647	331446	581	4	0.001%
KEH02801	313620	514	4	0.001%
KEH03015	337337	562	5	0.001%
KEH03086	326283	546	8	0.002%
N17	180916	3163	6	0.003%
YPS138	333202	487	8	0.002%
Average	185356	733	4	0.002%

[1] Total sites after quality filtering

[2] SNPs are defined in relation to the reference genome used for assembly (see Materials and Methods)

* Isolate was excluded from further analysis due to low sequence coverage

Table 4-11. False discovery rate estimates for solexa RAD genotyping.

Strain	**Species**	**False positive rate**	**Number of sequenced sites**	**Estimated number of false positives**
M22	*S. cerevisiae*	0.0000200	319812	6
YPS163	*S. cerevisiae*	0.0000120	381375	5
YPS138	*S. paradoxus*	0.0004059	333203	135
N17	*S. paradoxus*	0.0001594	180906	29

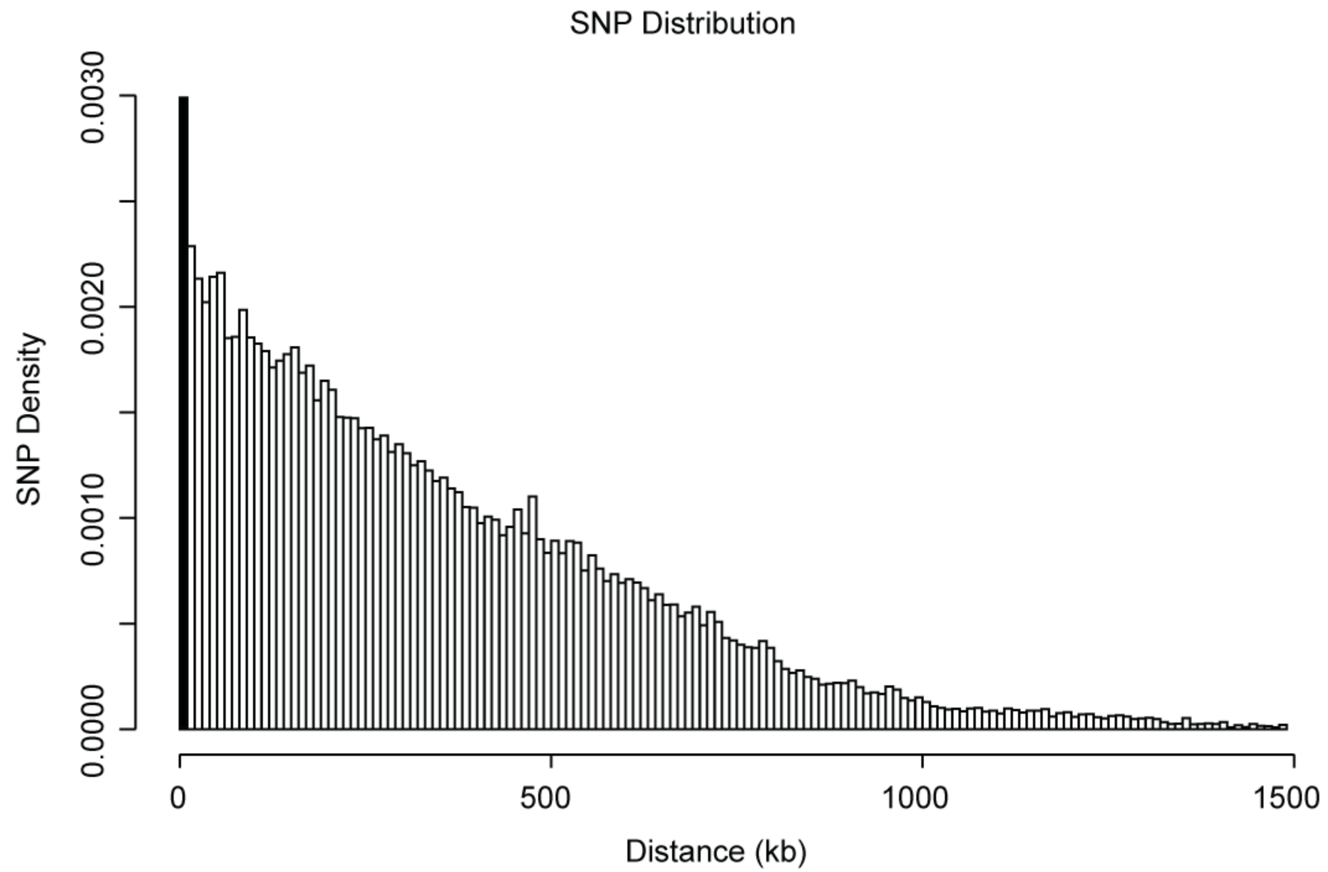


Figure 4-1. SNP density in *S. cerevisiae.*

SNP density (the proportion of pairwise comparisons among 3,323 parsimony informative SNPs) is shown as a function of distance in kilobases. Pairwise distance of 10kb or less make up less than 0.3% of the data (black bar).

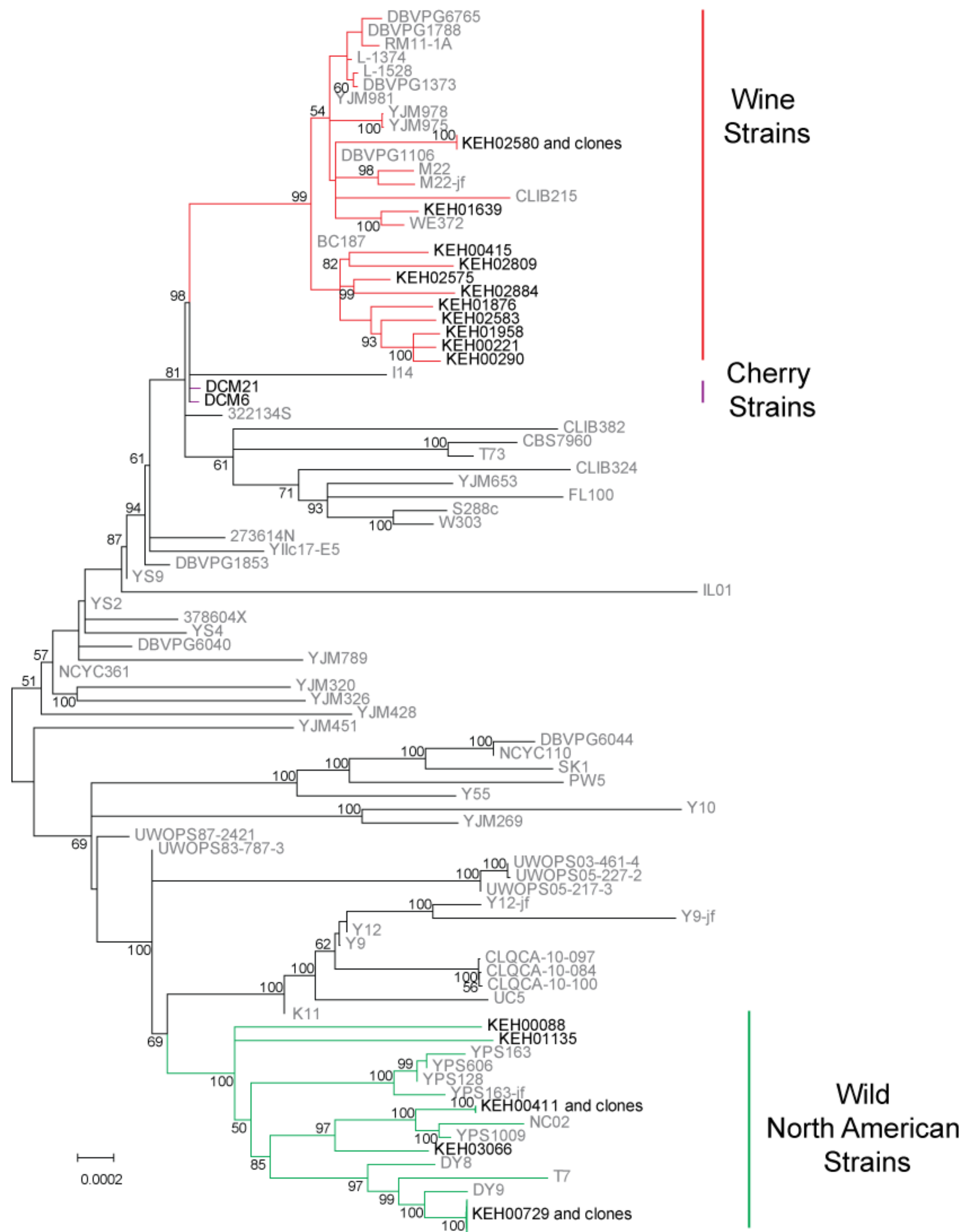


Figure 4-2. Neighbor-Joining phylogeny of *S. cerevisiae.*

The bootstrap consensus tree (1,000 replicates) is based on pairwise genetic distances (nucleotide p-value) at 220,996 positions. Bootstrap values less than 50 are not shown. The tree is drawn to scale.

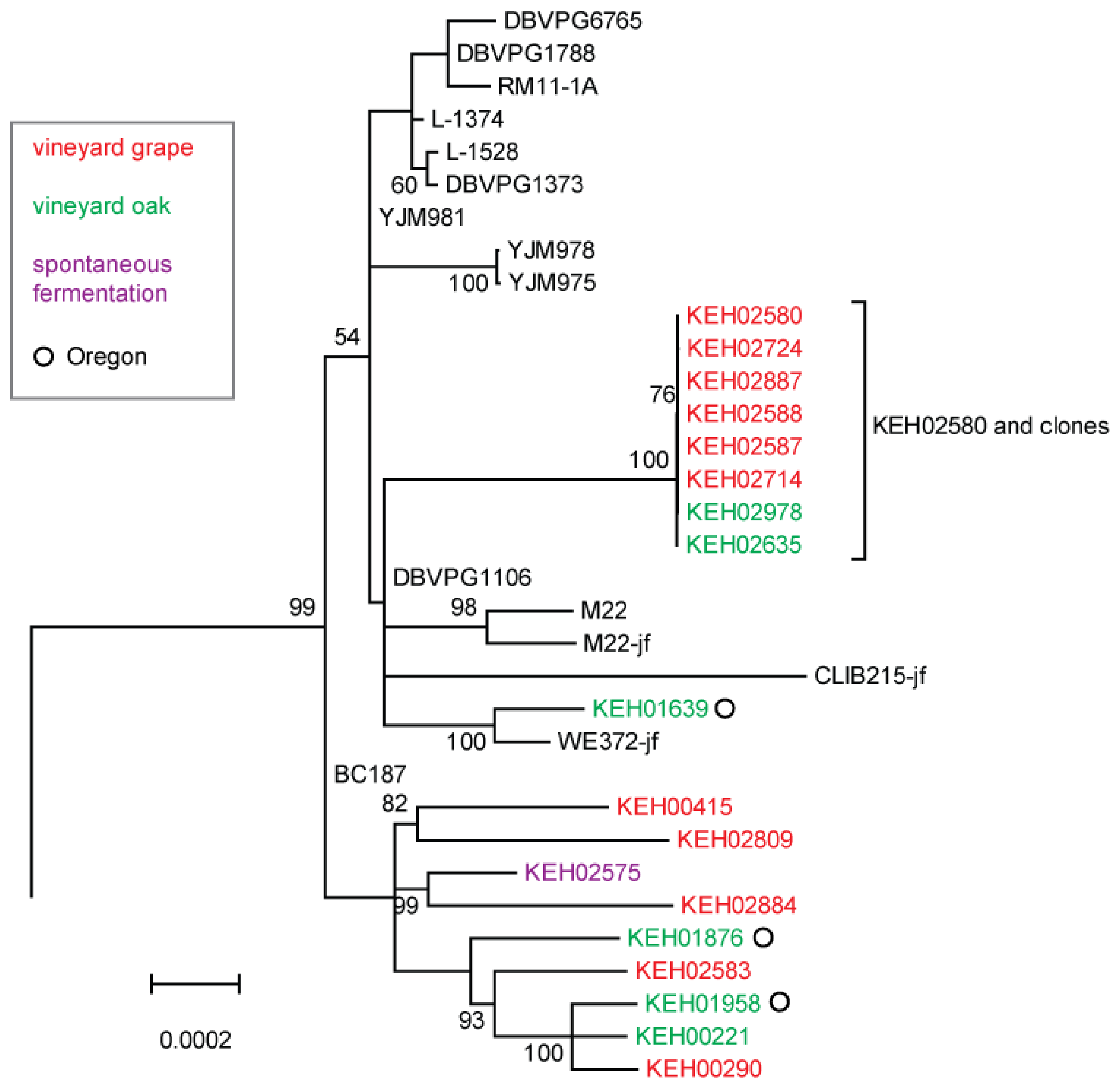


Figure 4-3. Neighbor Joining phylogeny of *S. cerevisiae* wine strains.

The wine lineage from the bootstrap consensus tree (1,000 replicates) of 115 taxa based on pairwise genetic distances (nucleotide p-value) at 220,996 positions. Bootstrap values less than 50 are not shown. The tree is drawn to scale.

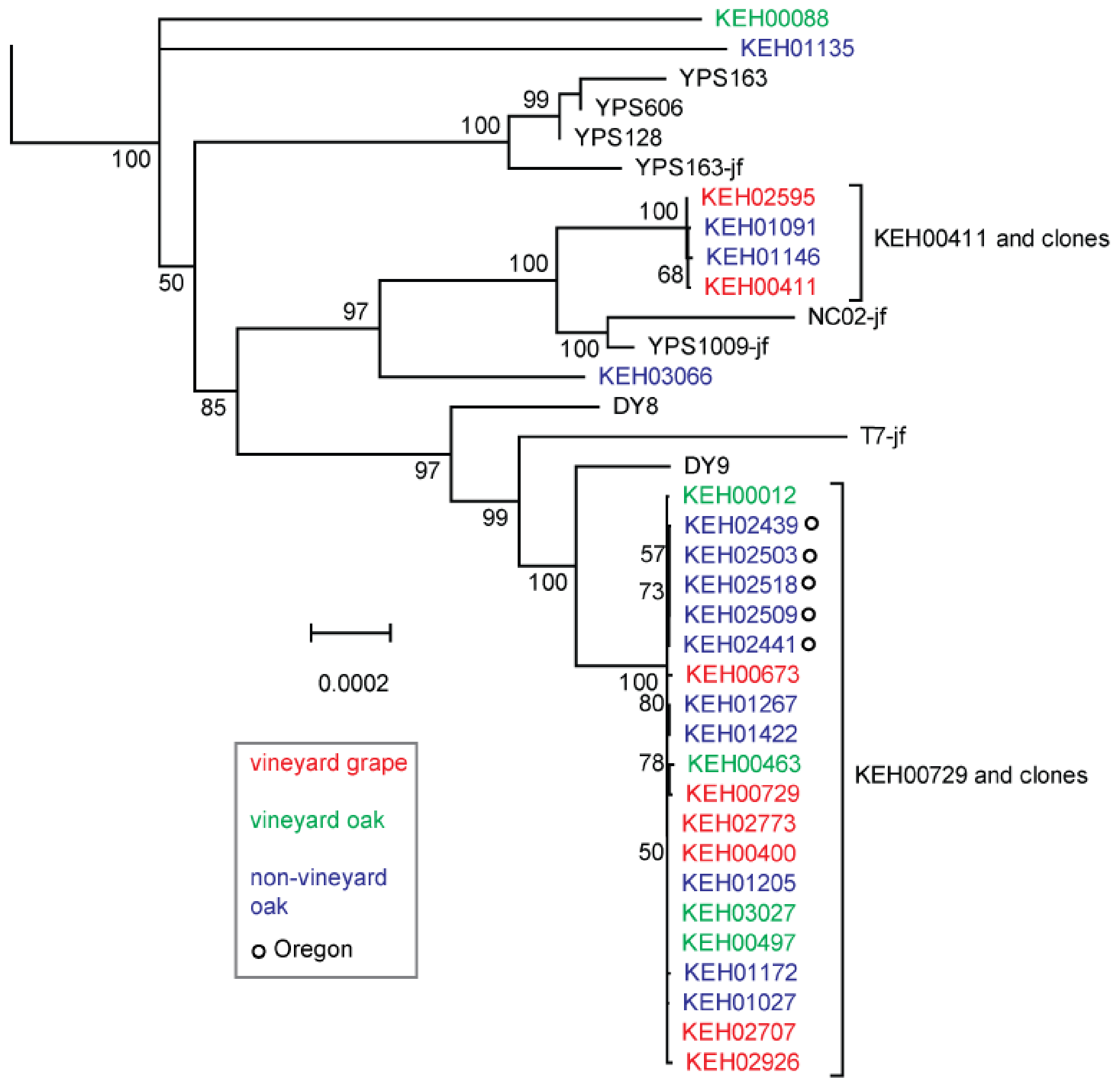


Figure 4-4. Neighbor Joining phylogeny of *S. cerevisiae* wild strains.

The wild North American lineage from the bootstrap consensus tree (1,000 replicates) of 115 taxa based on pairwise genetic distances (nucleotide p-value) at 220,996 positions. Bootstrap values less than 50 are not shown. The tree is drawn to scale.

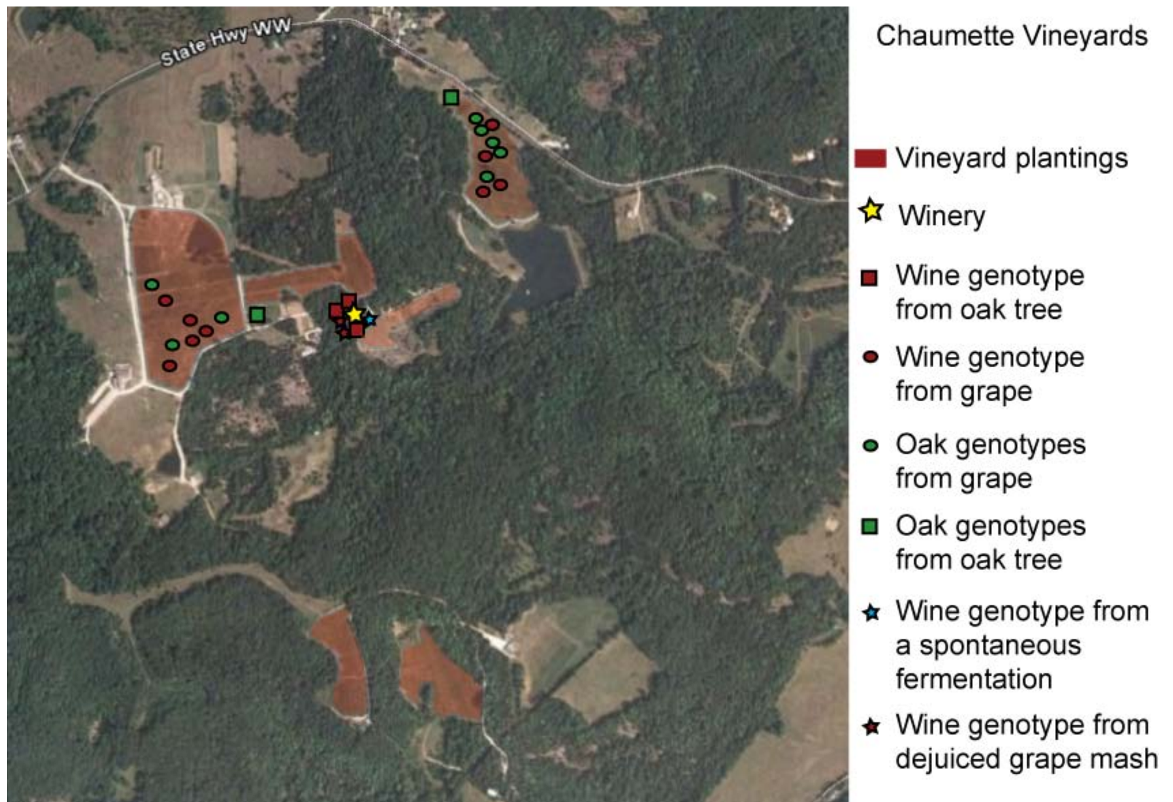


Figure 4-5. Geographical distribution of *S. cerevisiae* genotypes within Chaumette Vineyards.

Table 4-12. Nucleotide diversity (π) within and between wine and wild strains of *S. cerevisiae*.

	# strains	4fold degenerate sites (23614)	2fold degenerate sites (36230)	Non-degenerate sites (124382)	All coding sites (183915)	Non-coding sites (26651)	All sites (210566)	π_N / π_S
All *S. cerevisiae* strains	115	0.699	0.510	0.132	0.278	0.454	0.300	0.189
All wine strains	32	0.115	0.091	0.041	0.060	0.089	0.064	**0.357**
All wild strains	35	0.375	0.242	0.049	0.128	0.176	0.134	0.131
Wine and wild strains - overall	67	0.668	0.447	0.109	0.246	0.380	0.263	0.163
Wine and wild strains - within subpopulations	67	0.245	0.166	0.045	0.094	0.133	0.099	0.184
Wine and wild strains – between populations	67	0.423	0.281	0.064	0.152	0.247	0.164	0.151

Nucleotide diversity is π * 100, calculated using MEGA4.0 (Tamura et al. 2007) based on pairwise comparisons of nucleotide p-value.

π_N / π_S is the ratio of nucleotide diversity at nondegenerate (N) to 4-fold degenerate (S) sites.

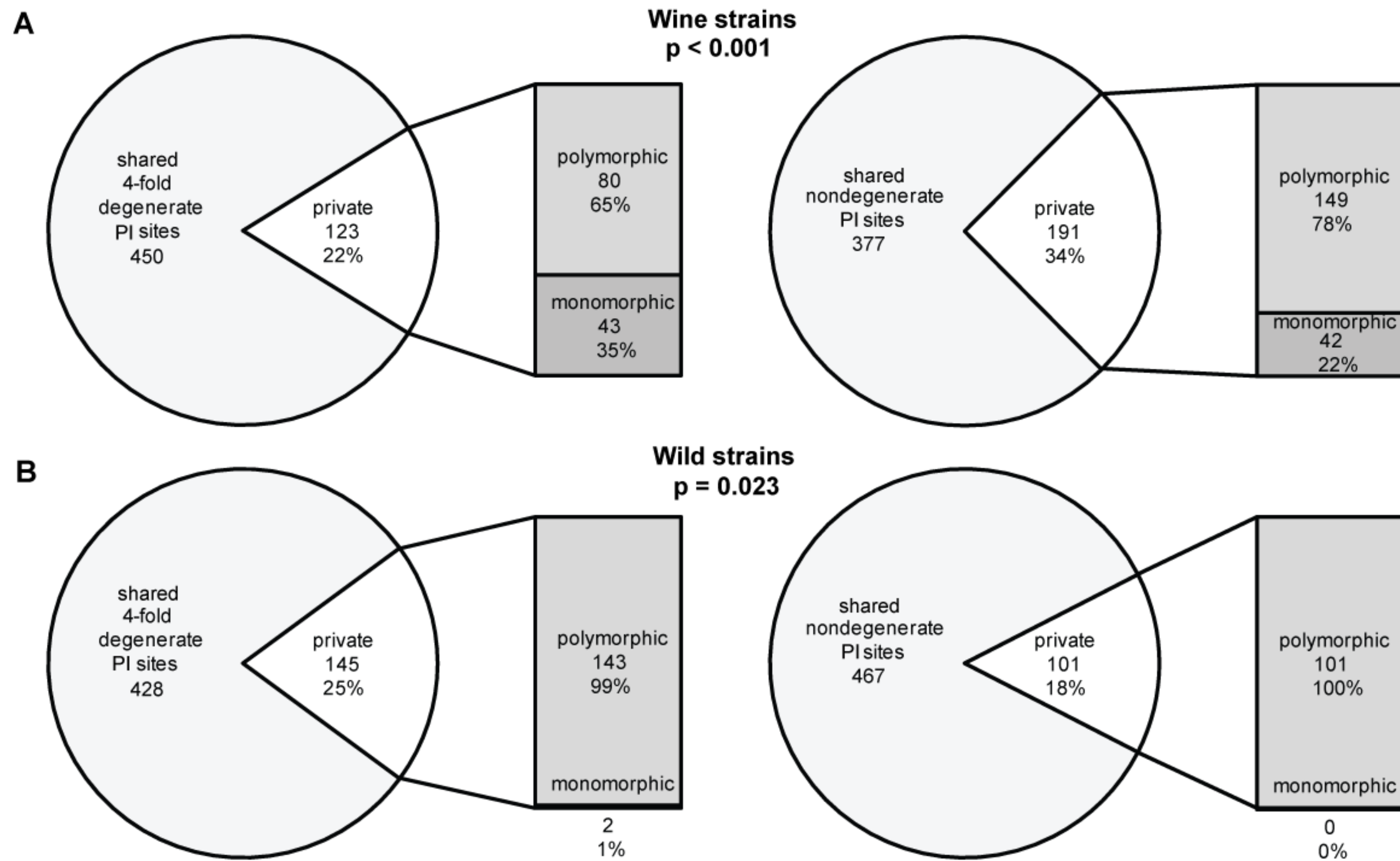


Figure 4-6. The proportion of private alleles at 4-fold degenerate and nondegenerate sites in wine and wild strains of *S. cerevisiae.*

Parsimony informative sites for wine, wild, and SGRP 'clean lineages' (Liti et al. 2009a) were included for analysis. P values are between 4-fold degenerate and nondegenerate sites in (A) wine strains and (B) wild strains of *S. cerevisiae.*

Table 4-13. The proportion of private alleles in wine and wild strains of *S. cerevisiae.*

Lineage	Substitution type	Proportion of private alleles[1]	Proportion of monomorphic private alleles[2]
Wine	4-fold degenerate	0.215	0.350
	Nondegenerate	0.336	0.220
		$p < 0.001$	**$p = 0.014$**
Wild	4-fold degenerate	0.253	0.014
	Nondegenerate	0.178	0.000
		$p = 0.002$	$p = 0.514$

[1] p values are calculated using Fisher's exact test between the number of shared and private alleles at 4-fold degenerate and non-degenerate sites.

[2] p values are calculated using Fisher's exact test between the number of polymorphic and monomorphic private alleles at 4-fold degenerate and non-degenerate sites.

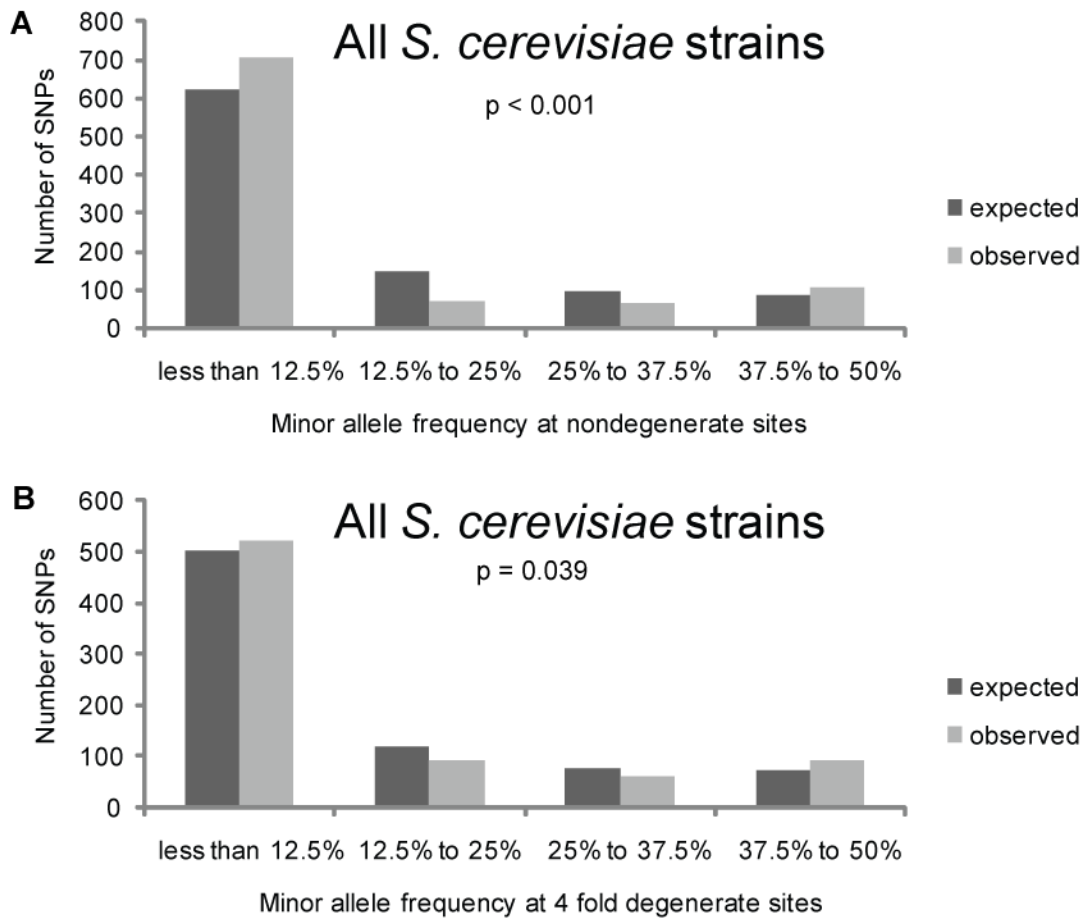


Figure 4-7. Expected and observed minor allele frequencies for *S. cerevisiae.*

Minor allele frequencies were calculated for parsimony informative sites. Clonal strains were removed for a total of 23 strains. Expected allele frequencies under neutral evolution were calculated using Watterson's θ.

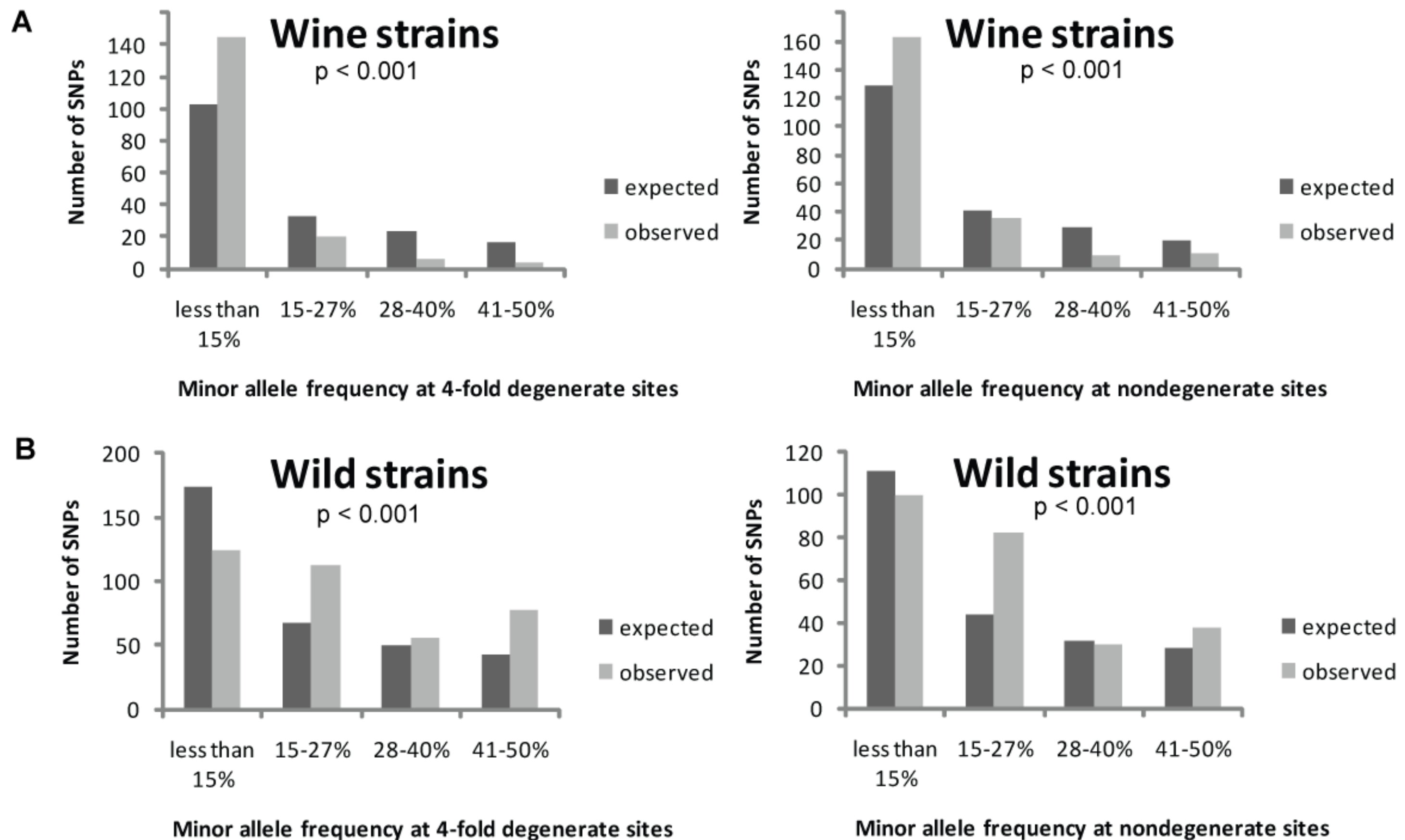


Figure 4-8. Expected and observed minor allele frequencies in wine and wild strains of *S. cerevisiae.*

Minor allele frequencies were calculated for parsimony informative sites. Clonal strains were removed for a total of (A) 23 wine strains and (B) 12 wild strains. Expected allele frequencies under neutral evolution were calculated using Watterson's θ.

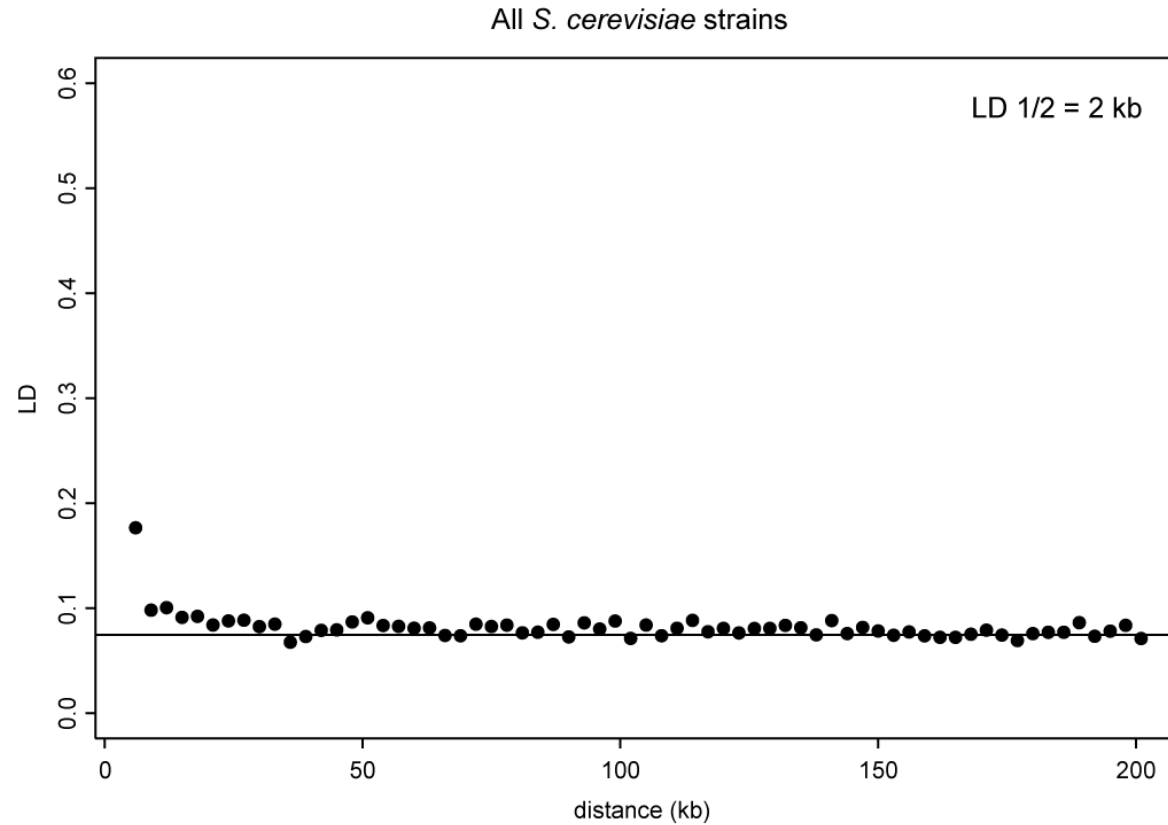


Figure 4-9. Linkage disequilibrium in *S. cerevisiae*.

Linkage disequilibrium for 115 *S. cerevisiae* strains is measured as r^2. LD ½ is the distance at which the maximum value for r2 in a 500 bp window decreases by ½.The average r^2 for a randomly sampled set of physically unlinked loci is indicated with a horizontal line.

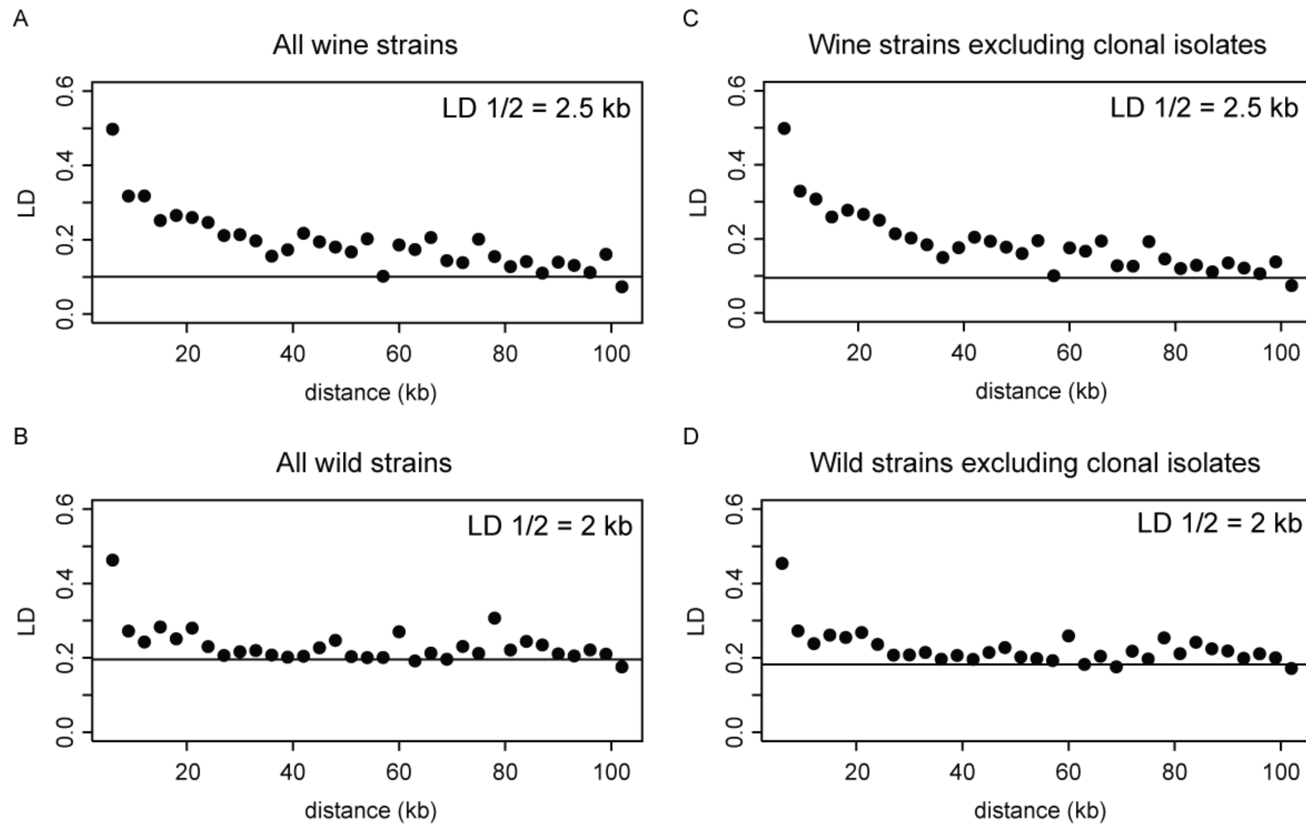


Figure 4-10. Linkage disequilibrium in wine and wild *S. cerevisiae* populations.

Linkage disequilibrium for (A) All wine strains, (B) all wild strains, (C) wine strains excluding clonal isolates and (D) wild strains excluding clonal isolates is measured as r^2. LD ½ is the distance at which the maximum value for r2 in a 500 bp window decreases by ½.The average r^2 for a randomly sampled set of physically unlinked loci is indicated with a horizontal line

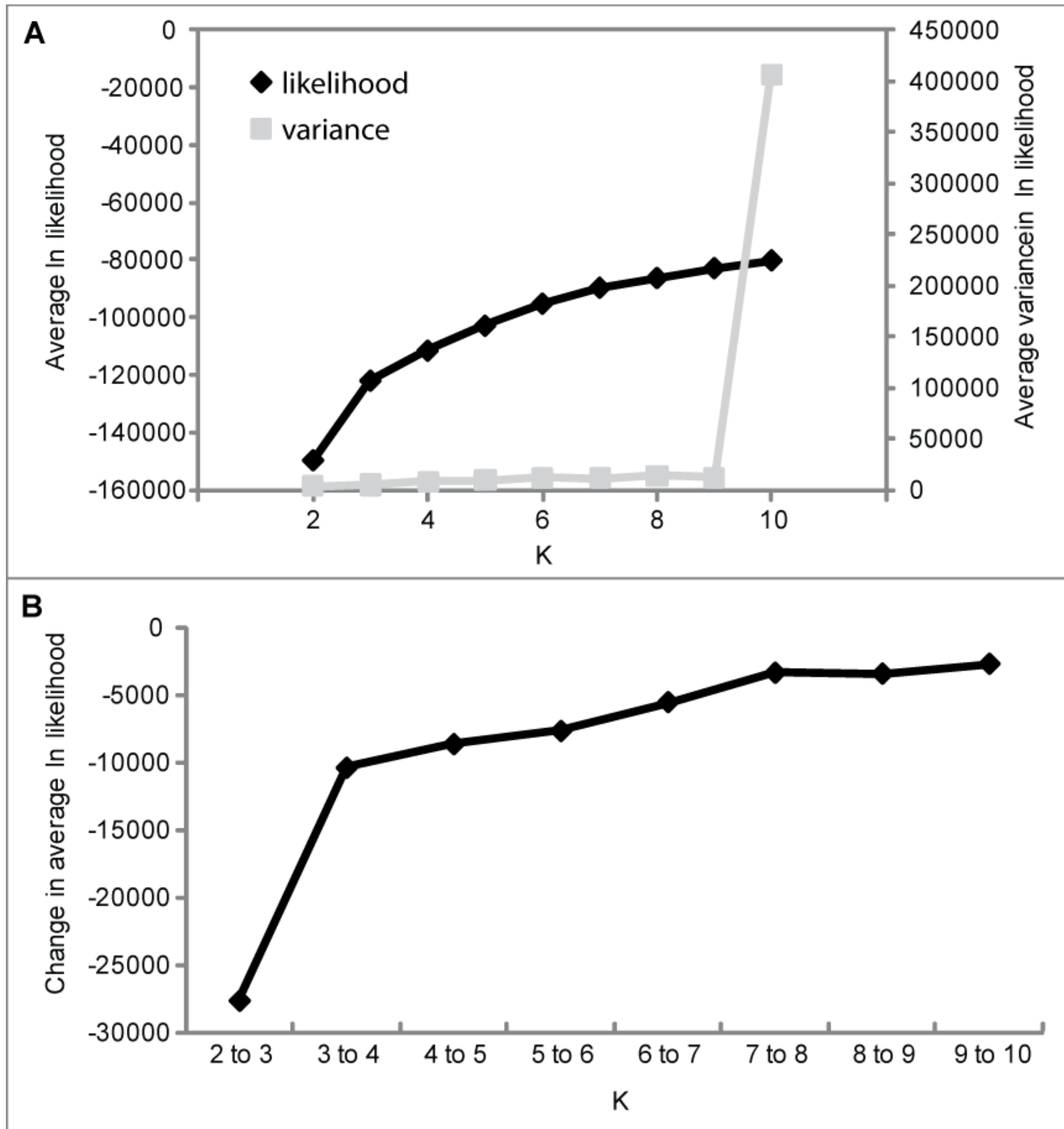


Figure 4-11. Average likelihood and variance for population structure simulations of *S. cerevisiae.*

Average likelihood and variance values (A) and the change in average likelihood (B) from 10 replicate STRUCTURE simulations (Pritchard et al. 2000) for each number of inferred populations (K).

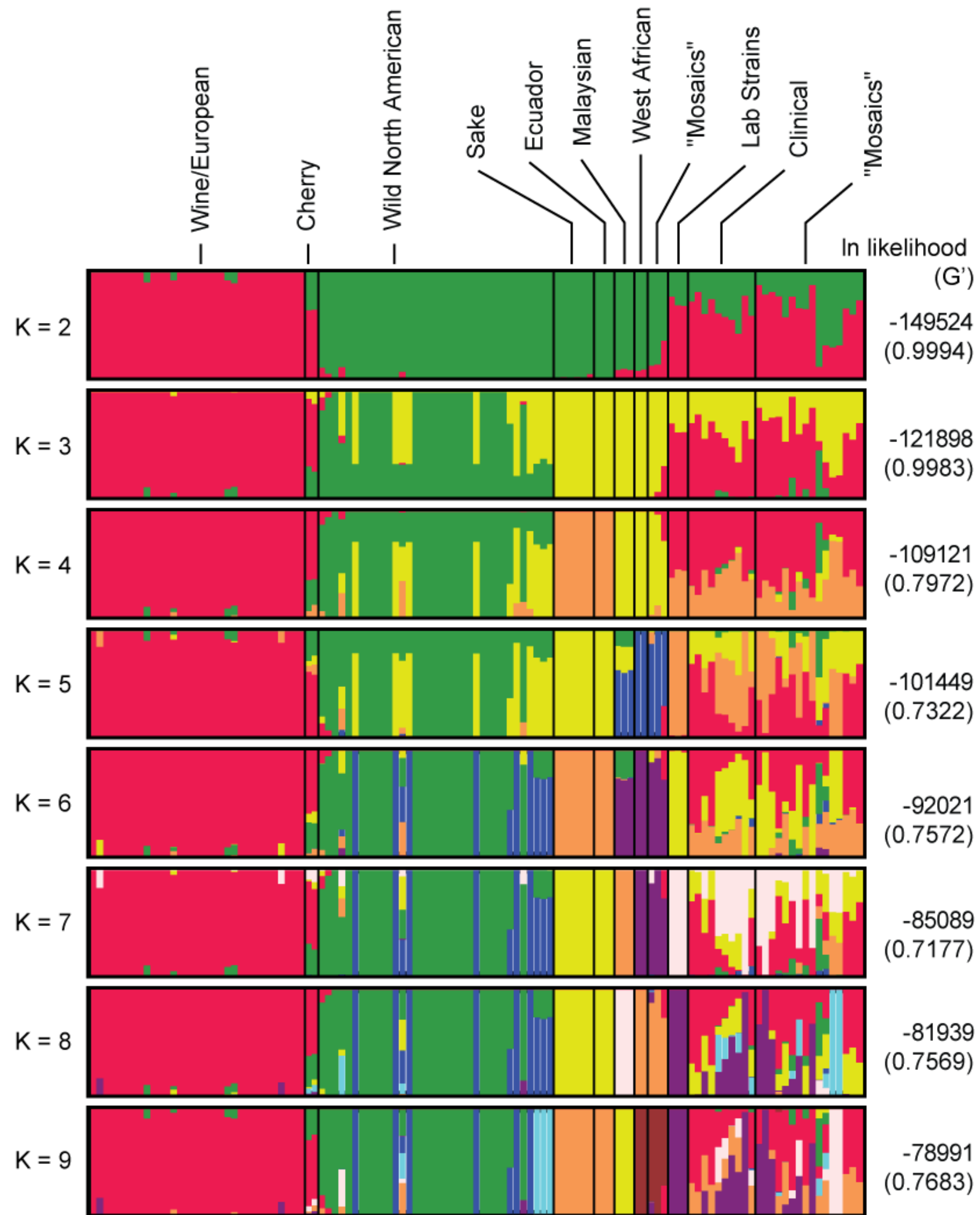


Figure 4-12. Inferred population structure of *S. cerevisiae*.

Population structure for 2 through 9 populations (K) was inferred using Structure (Pritchard et al. 2000). Ten replicate simulations were generated for each K. G' is the similarity coefficient between replicate simulations calculated using CLUMPP (Jakobsson & Rosenberg 2007).

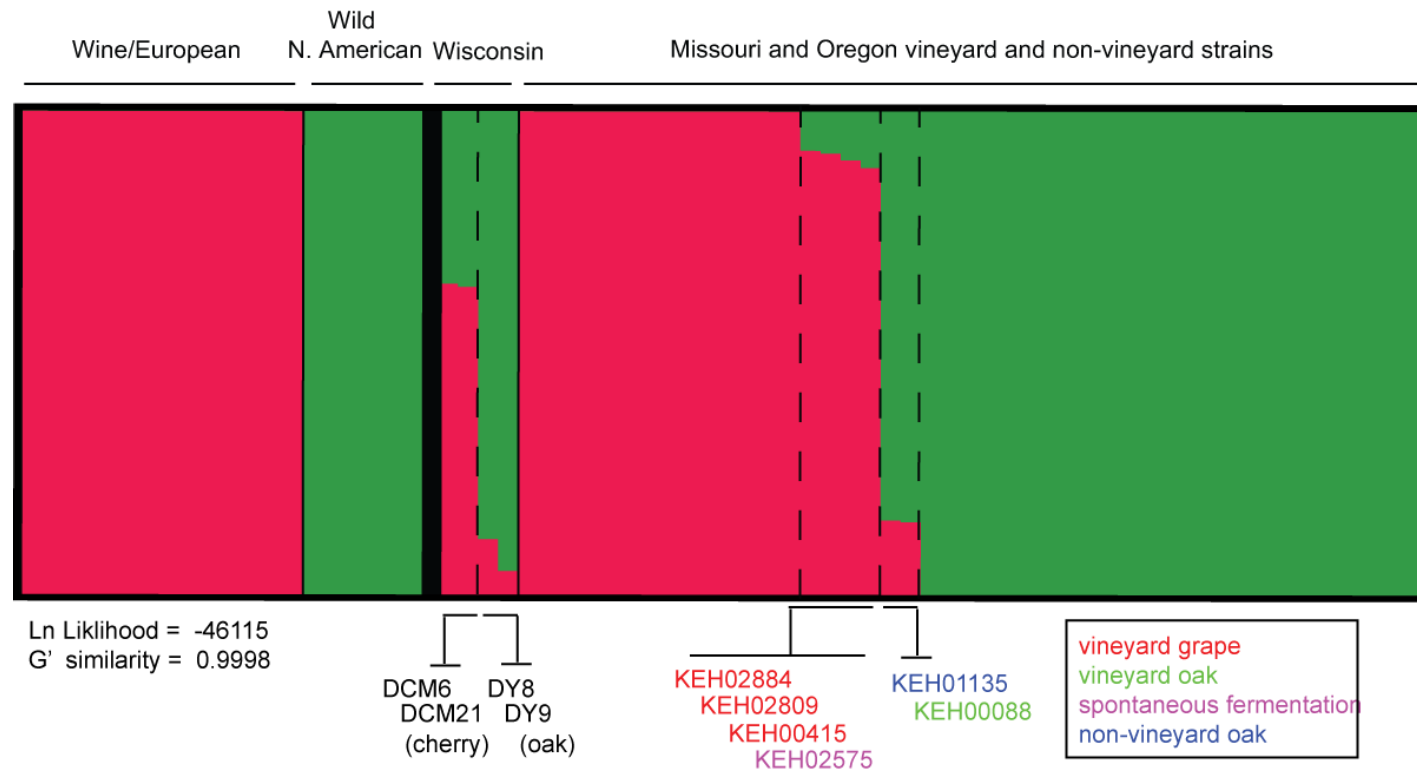


Figure 4-13. Admixture between wine and wild genotypes of *S. cerevisiae.*

Admixture was inferred using STRUCTURE (Pritchard et al. 2000) by assigning previously identified and genome sequenced strains to two populations (left of the thick black bar), and inferring the inferred ancestry of strains sampled in this study (right of the thick black bar). The similarity coefficient (G') from ten replicate runs was determined using CLUMPP (Jakobsson & Rosenberg 2007).

Table 4-14. Inferred ancestry of admixed *S. cerevisiae* isolates.

ID	Source	Inferred wine/European ancestry	Inferred wild North American ancestry
DCM6	Cherry, WI	0.64 (0.60)	0.36 (0.30)
DCM21	Cherry, WI	0.64 (0.58)	0.36 (0.24)
DY8	Oak, WI	0.11 (0.12)	0.89 (0.82)
DY9	Oak, WI	0.05 (0.05)	0.95 (0.95)
KEH02884	Grape mash, MO	0.88 (0.89)	0.12 (0.09)
KEH02809	Grape, MO	0.91 (0.92)	0.09 (0.08)
KEH0415	Grape, MO	0.90 (0.91)	0.10 (0.09)
KEH02575	Spontaneous fermentation, MO	0.92 (0.92)	0.08 (0.04)
KEH01135	Non-vineyard oak, MO	0.15 (0.00)	0.85 (0.61)
KEH0088	Vineyard oak, MO	0.15 (0.00)	0.85 (0.55)

Ancestry was inferred using STRUCTURE (Pritchard et al. 2000). The inferred ancestry values are averages for 10 replications, evaluated using CLUMPP (Jakobsson & Rosenberg 2007). Values in parentheses are the inferred ancestry from K=7 in the unsupervised analysis.

American

Hawaiian

UWOPS91 917 1

0.001

European?

European

Far Eastern

Figure 4-14. Neighbor-joining phylogeny of *S. paradoxus*.

The bootstrap consensus tree (1,000 replicates) of 66 taxa based on pairwise genetic distances (nucleotide p-value) at 281,944positions. Missing and ambiguous data were removed for pairwise comparisons only. The tree is drawn to scale.

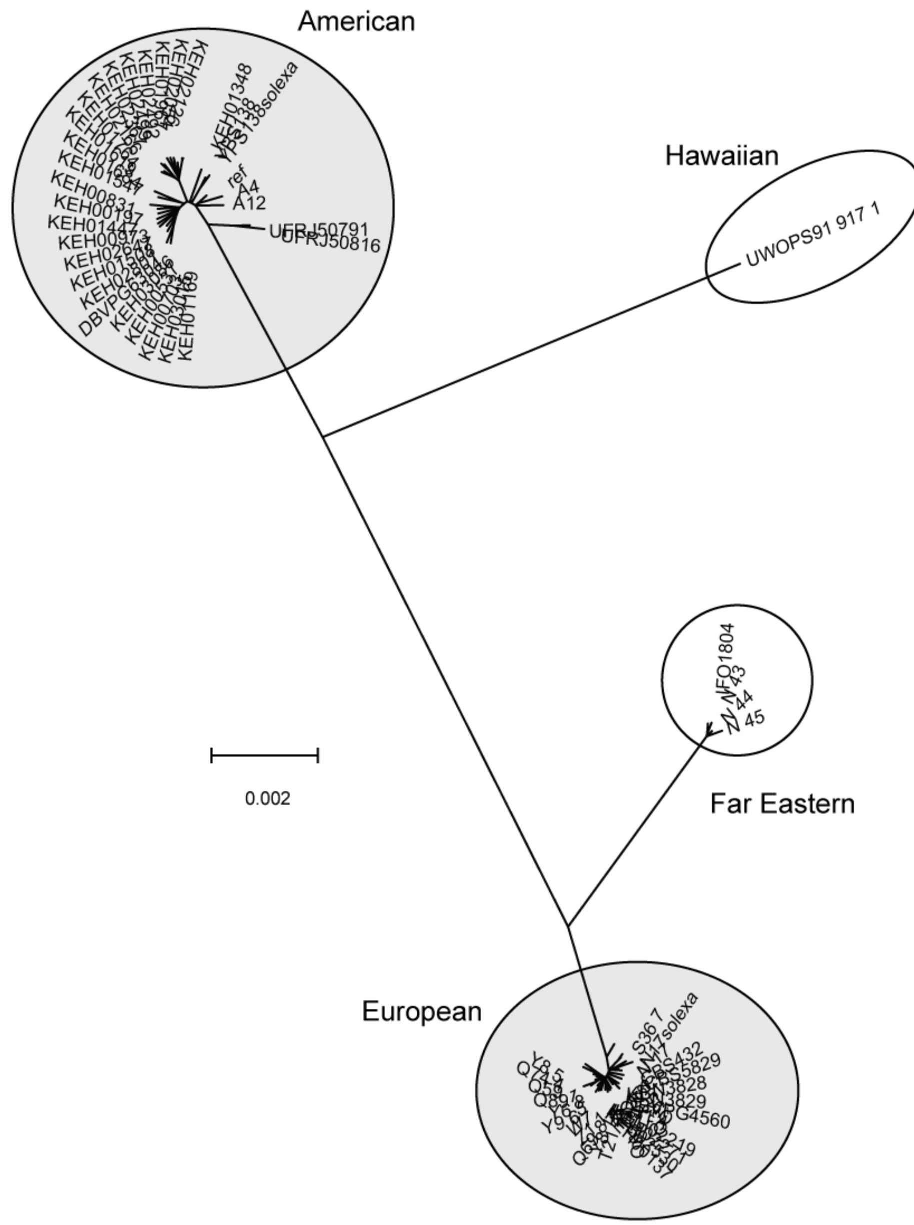


Figure 4-15. Neighbor-joining phylogeny of *S. paradoxus.*

The bootstrap consensus tree (1,000 replicates) of 66 taxa based on pairwise genetic distances (nucleotide p-value) at 96,753 positions. All positions with missing and ambiguous data were removed. The tree is drawn to scale.

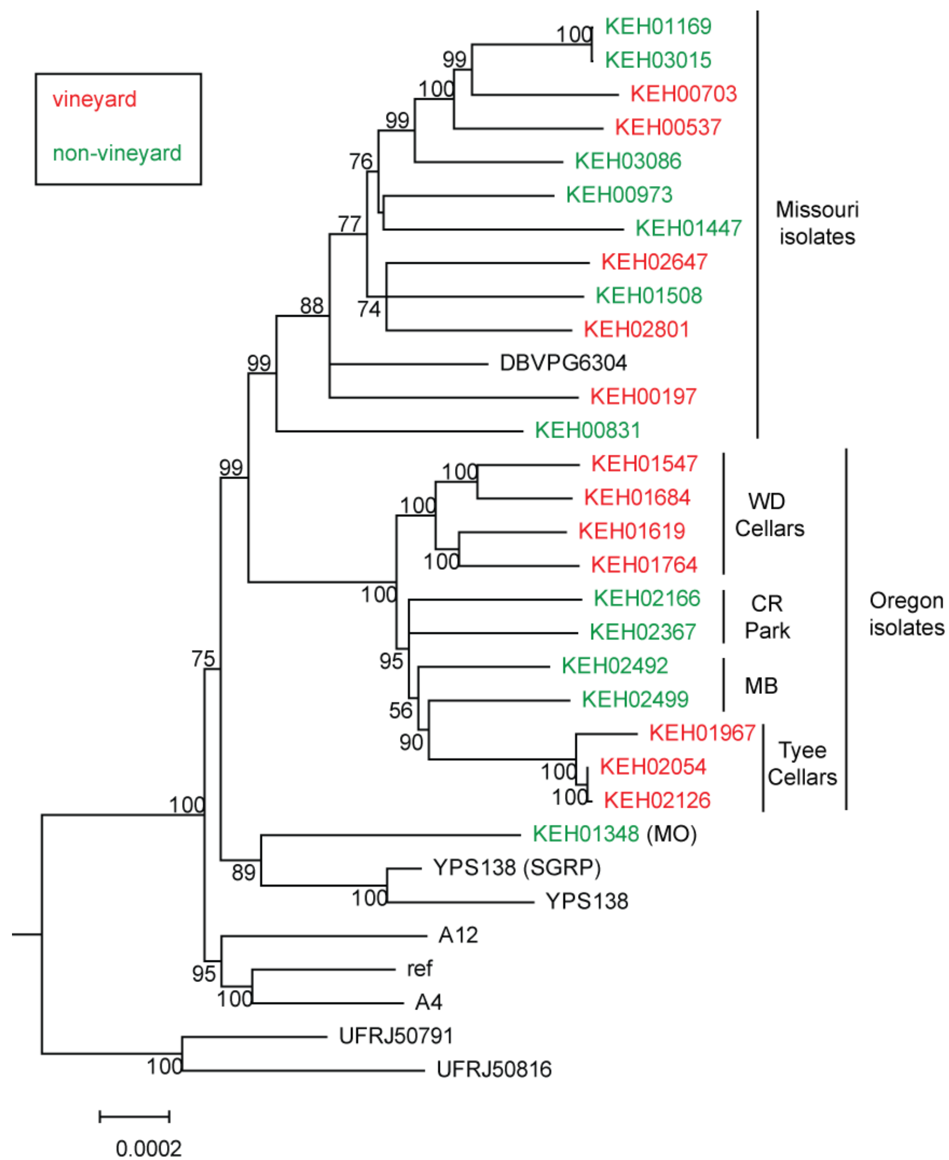


Figure 4-16. Neighbor-joining phylogeny of American *S. paradoxus* isolates.

The American *S. paradoxus* lineage from the bootstrap consensus tree (1,000 replicates) of 66 taxa based on pairwise genetic distances (nucleotide p-value) at 281,944positions. Missing and ambiguous data were removed for pairwise comparisons only. The tree is drawn to scale.

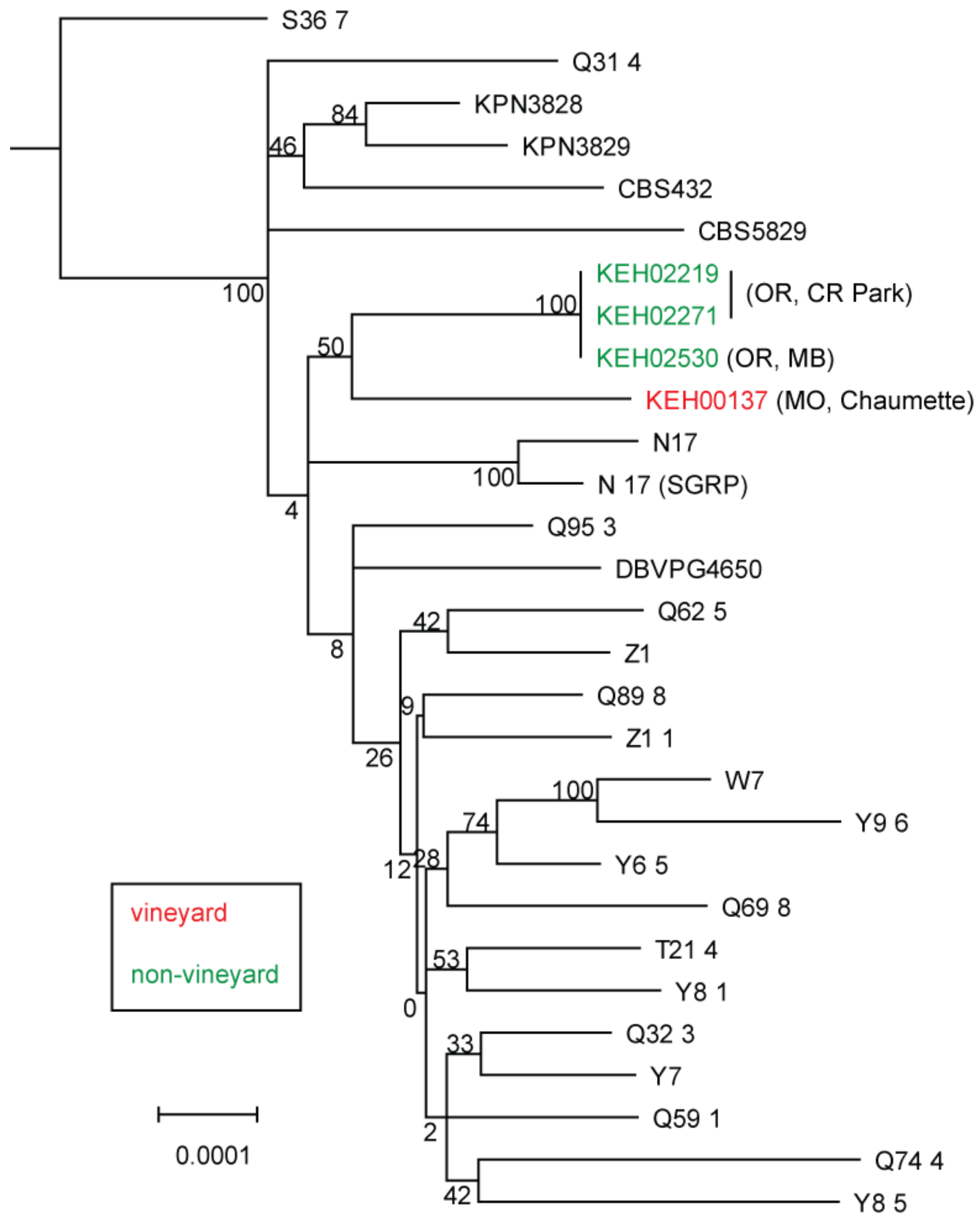


Figure 4-17. Neighbor-joining phylogeny of European *S. paradoxus.*

The European *S. paradoxus* lineage from the bootstrap consensus tree (1,000 replicates) of 66 taxa based on pairwise genetic distances (nucleotide p-value) at 281,944positions. Missing and ambiguous data were removed for pairwise comparisons only. The tree is drawn to scale.

Table 4-15. Nucleotide diversity in *S. paradoxus.*

	# of strains	**π * 100**
American	30	0.167
European	28	0.074
Far Eastern	4	0.057
Total	63	1.413
Within populations	63	0.099 (7%)
Between populations	63	1.314 (93%)

Nucleotide diversity is π * 100, calculated using MEGA4.0 (Tamura et al. 2007) based on pairwise comparisons of nucleotide p-value.

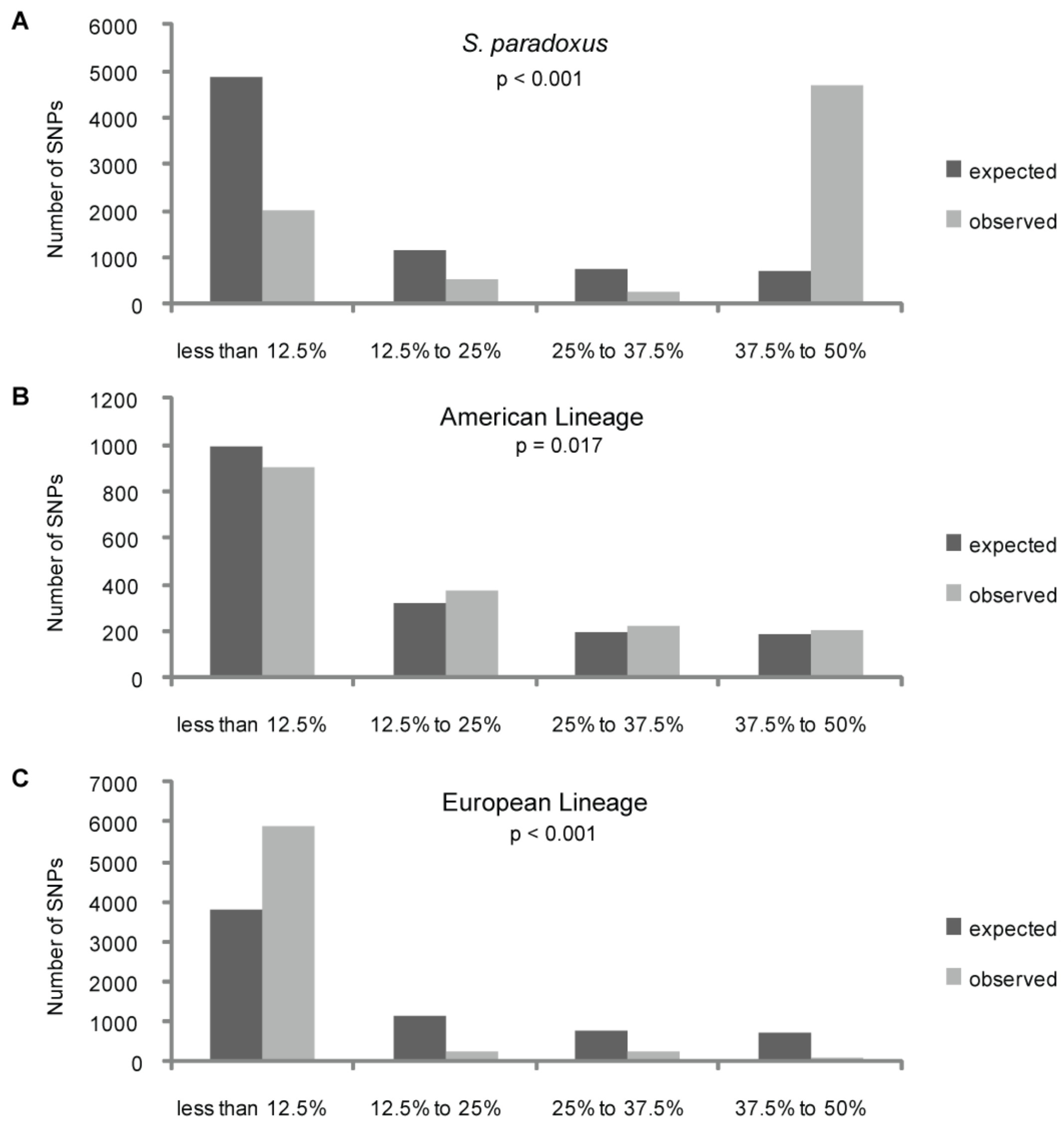


Figure 4-18. Minor allele frequencies in *S. paradoxus.*

Minor allele frequencies were calculated for parsimony informative sites (A) across all *S. paradoxus* strains, (B) within the American lineage, and (C) within the European lingeage. Expected allele frequencies under neutral evolution were calculated using Watterson's θ.

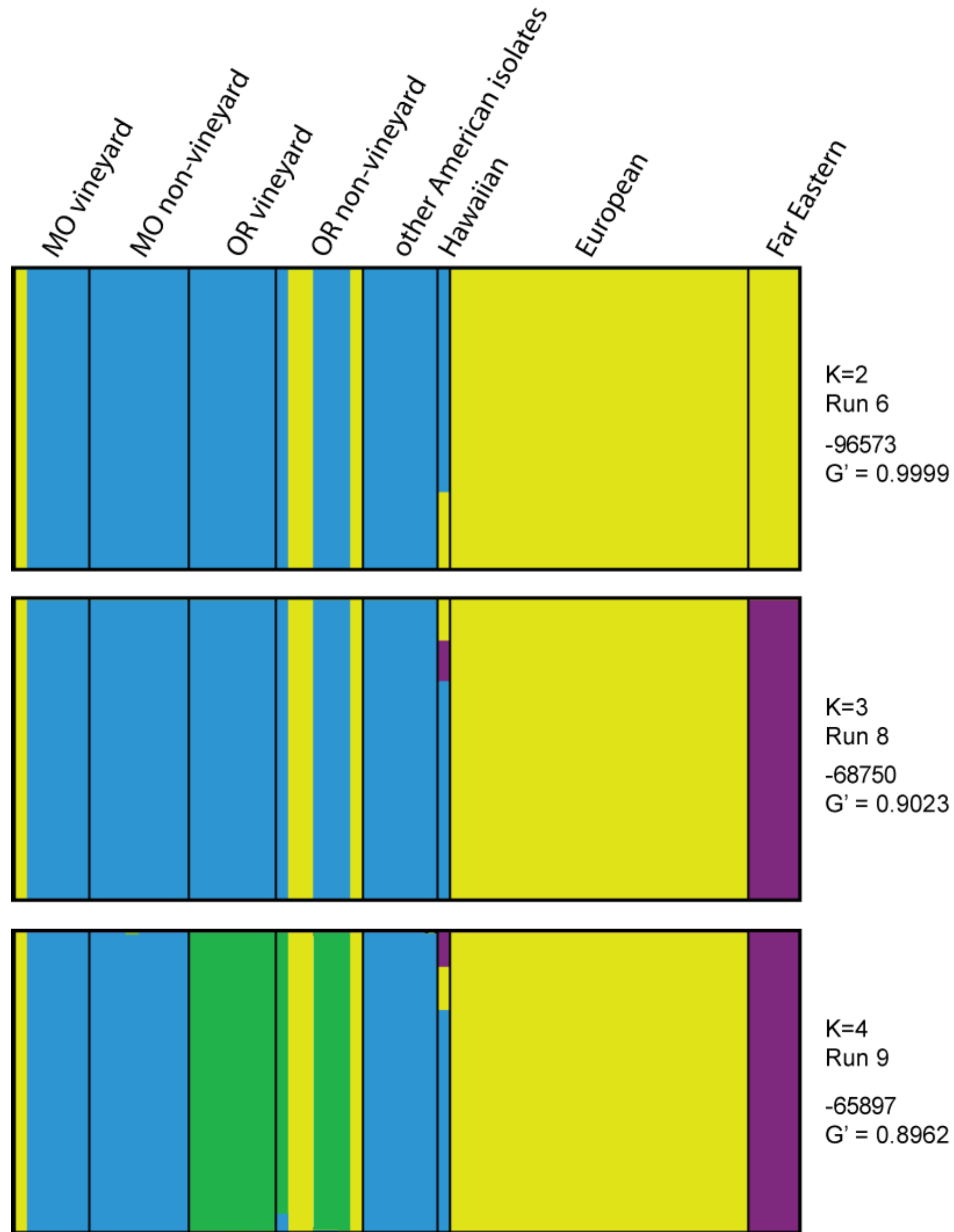


Figure 4-19. Population structure in *S. paradoxus.*

Population structure for 2 through 4 populations (K) was inferred using Structure (Pritchard et al. 2000b). Ten replicate simulations were generated for each K. G' is the similarity coefficient between replicate simulations calculated using CLUMPP (Jakobsson & Rosenberg 2007).

Chapter 5 : Conclusions, Inferences and Future Directions

Domestication Phenotypes in S. cerevisiae

This study was initiated to address hypotheses regarding the domestication of the fungus *S. cerevisiae* for winemaking. Although domestication has been well characterized in many crop (Doebley et al. 2006), and animal species (Goodrich & Wiener 2005), there is only a single previous case (*Aspergillus oryzae)* that explores both the phenotypic and genotypic signatures of domestication in a fungal species (Rokas 2009). This dissertation represents the first examination of the hypothesis of domestication of *S. cerevisiae* wine strains in a combined phenotypic, genotypic and ecological context.

In chapters one and two we tested the hypothesis that like plant and animal domestications, phenotypes exist that correlate with the genetic differentiation of wine and wild strains. In chapter one we tested whether local habitat specific adaptation has altered the relative fitness of wine and wild strains. In chapter two we tested the hypothesis that human perception for flavor and aroma characteristics could act as a selectable phenotype that distinguishes strains and could thus contribute to domestication. Although we find no evidence for fitness differences correlated with the domestication of *S. cerevisiae* wine strains, this study is the first that we know of to provide evidence that the genetic divergence between wine and wild strains of *S. cerevisiae* is accompanied by divergence in wine aroma and flavor attributes. The enrichment of several desirable wine attributes and/or the elimination of several undesirable wine attributes imply that wine strains may have been intentionally or inadvertently domesticated for the production of quality grape wine.

Our results demonstrate that humans could have selected for differences in wine attributes between putatively domesticated wine strains and wild strains of *S. cerevisiae.* Although we did not specifically test which attributes humans prefer, it is easy to understand how sulfurous aromas and flavors (e.g. cabbage, wet dog) could be highly unpleasurable while fruity and floral aromas are not. However, human preference can vary widely, which brings up the

question: Is it possible that our ancestors could have agreed enough on what made a good quality wine to select on yeast derived aromas and flavors? The concept of selection for aroma and flavor compounds is not unique to *S. cerevisiae.* Many domesticated and cultivated species have diverged similarly from their wild counterparts due to human selection. For example, studies have documented flavor and aroma differences due to secondary metabolite spectrum differences between wild and cultivated strawberries (Aharoni et al. 2004) and also for presence of aromas in basmati rice (Kovach et al. 2009). Some species have also been selected for more subtle changes, changes that often coincide with the specific preference of a culture or region. An example of this concept of culture selection is the preference for glutinous rice in southeast Asia, where selection for varieties with differences in starch content contribute to this texture (Olsen & Purugganan 2002). Other domesticated species have been selected to have striking differences from their wild counterparts, so much so that the wild species is considered unpalatable or unmarketable, for example the selection for sweet almonds, which has resulted in almonds that do not contain the chemicals that create hydrogen cyanide after chewing (Zohary & Hopf 2000).

Within *S. cerevisiae,* we found the most phenotypic variation in wine attributes within the wine strains themselves, even though the genetic diversity in this group is much lower than in wild strains. Even so, several sulfur-related attributes (e.g. cabbage, wet dog) that are produced by wild strains are conspicuously absent in wine strains. It is possible that these sulfurous attributes were widely considered ‘unpalatable,’ akin to bitter almonds in the United States. Increased variation within wine strains could have been caused by selection for varying wine attributes following selective elimination of sulfurous attributes. It is also possible that the removal of sulfurous attributes simply allowed humans to perceive the more subtle aroma and flavor attributes that were already present in the background. In either case, humans today have access to a broad array of wine aroma and flavor compounds (Noble et al. 1987) that are produced in part due to the yeast strains used for fermentation. Such an increase in phenotypic diversity following domestication is not uncommon. One of the best examples of this is the domestic dog;

the popularity of breeding dogs as pets has resulted in an explosion in the phenotypic variation (Serpell 1995).

In addition to differentiation in wine aroma and flavor characteristics, we also described differences in resistance to copper sulfate between wine and wild strains of *S. cerevisiae.* Although this phenotype has been hypothesized to be an adaptation to vineyard life, we have found that the resistance to copper sulfate does not correlate with vineyard life for other, non *Saccharomyces* yeasts. Curiously, after demonstrating that vineyard populations of *S. cerevisiae* are made up of two distinct genotypes (wine and wild), we find that copper sulfate resistance correlates perfectly with genotype (wine) rather than habitat (vineyard), although the sample size was very low (8 wild genotypes and 2 wine genotypes). This observation, along with the distribution of resistance in non-*Saccharomyces* species populations (Chapter 3) suggests that either copper sulfate resistance is not a specific adaptation for survival in vineyards, that the strength of selection is variable between species, or that selective pressures in these vineyards have changed.

From our new collections of vineyard and non-vineyard yeast species, we provide evidence that copper sulfate resistance within vineyards is variable, and that copper sulfate resistance may be a useful trait for identifying strains with a 'wine' genotypic background. The lack of copper sulfate resistant strains in one Missouri vineyard (Mt. Pleasant), may indicate the absence or relatively low frequency of wine genotypes in this vineyard, in contrast with Chaumette vineyard in Missouri, from which we isolated both wine and wild genotypes, and which shows a high proportion of copper sulfate resistant *S. cerevisiae* strains. A likely contributing difference between these two vineyards is that Chaumette ferments wine on the vineyard premises, whereas Mt. Pleasant ferments wine at an off-site facility. If the lack of copper sulfate resistant isolates truly does suggest the recent introduction of European strains, the pattern would support the idea that wine genotypes migrate out of winery facilities and onto grapes and

adjacent oak trees. This result also reinforces the idea that migration of wine strains is limited, or they are not fit enough to survive outside of the winery. Either interpretation suggests that winery proximity could greatly influence the structure of local *S. cerevisiae* populations.

California wineries are some of the oldest wineries in the United States, and a large proportion of them (~10%) utilize spontaneous fermentations rather than commercial starter strains (Mortimer 2000). One study (Mortimer 2000) surveyed the genetic structure of 239 strains from spontaneous fermentations in California wineries. They reported unusual patterns in a large number of strains, including increased heterozygosity, and variation in copper sulfate resistance. Although the genotypic relation of these isolates to the 'wine' and 'wild' populations from this study are unknown, the pattern of variation in California vineyards could indicate the presence of 'wine' strains (copper sulfate resistant), 'wild' strains (copper sulfate sensitive), and gene flow between them (heterozygosity for copper sulfate resistance). Their results may indicate a large presence of wild genotypes in spontaneous fermentations, with substantial levels of gene flow, supporting the observation of gene flow between 'wine' and 'wild' strains in this study. These results, given the clear phenotypic differentiation that we observed between wine and wild strains for wine attributes raise additional questions regarding the relative differences in fitness of hybrid and pure genotypes in competition in grape wine fermentations and also the contribution of 'wild' genotypes to the flavor and aroma of spontaneous fermentations.

When species are domesticated through genetic bottlenecks, genetic variation can be stripped away through neutral processes, rather than through selection. In most domesticated crop species, researchers routinely utilize the genetic variation from wild populations in order to reintroduce desirable traits and expand the range of abiotic and biotic stress resistance into domesticated species (Meilleur & Hodgkin 2004). The utilization of wild germplasm in crop species may predict future improvements in the wine industry. Currently, researchers are very interested in the contribution of different yeast strains to wine attributes (Pretorius 2000), and

studies are underway to explore the use of nontraditional *Saccharomyces* species for desirable winemaking properties (Majdak et al. 2002; Orlic et al. 2007; Orlić et al. 2010). Like other relationships between wild and domesticated species, it is possible that wild strains of *S. cerevisiae* may harbor variation that could modify or enhance the specific wine attributes that humans value, or that we did not realize we desired.

The Genetics Signatures of Domestication in S. cerevisiae

In chapters three and four we tested for the presence of genetic signatures of domestication in the context of broad sampling of *S. cerevisiae* at a local ecological scale from vineyard and non-vineyard locations in North America. The previous population genetic data used to infer the domestication of wine strains of *S. cerevisiae* was based on a global sampling of strains that were isolated over several decades without sufficient representation to account for local population structure. Although inferences of population structure and demography can be strongly influenced by this type of sampling scheme (Ross-Ibarra et al. 2008), we show that the same genetic patterns used to infer domestication can be recapitulated at a local scale. That is, we observe two distinct populations that exist sympatrically and correspond to European 'wine' genotypes and North American 'wild' genotypes. However, when including increased sampling at a local scale, our analysis demonstrates more complex patterns of population demography. Some of our results suggest patterns that are not common in plant or animal domestication study systems and suggest that the domestication of fungi should be considered from a different perspective.

Although *S. cerevisiae* is not typically considered a pathogen of crops, it shares many similar attributes to fungal plant pathogens, including the ability to propagate either through clonal or sexual reproduction. Therefore, the population dynamics of plant fungal pathogens may be useful for understanding the population genetic patterns observed in *S. cerevisiae*. Previous studies have demonstrated that the development of agro-systems can have a significant impact

on the population structure of crop associated fungal pathogens (Stukenbrock & McDonald 2008; McDonald & Linde 2002), including changes in their population size, level of diversity, rate of divergence and possibly system of mating (Stukenbrock & McDonald 2008; McDonald & Linde 2002). Our study demonstrates that a European population of *S. cerevisiae* is preferentially associated with vineyard grapes and oak trees. An analogous example for this sort of genetic specificity may come from pathogens of other fruit species. For example, the domestication of the apple (*Malus domestica)* is hypothesized to have contributed to genetic differentiation within a pathogen species, the apple scab fungus *Venturia inaequalis.* The genetic patterns observed appear to be related to the increased adoption and cultivation of apples (Gladieux et al. 2010).

In contrast to wine populations of *S. cerevisiae,* wild North American strains harbor high amounts of genetic diversity, but we show that this diversity is characterized by few clonal genotypes. In fungal pathogen populations, high levels of genetic diversity can be maintained through the presence of several divergent clonal genotypes (genomic diversity), rather than through the presence of large interbreeding populations. For example, *Magnaporthe oryzae,* one of the most important pathogens of rice, as well as *Phytophthora infestans,* the causal agent of potato blight, exhibit a highly clonal population structure (Stukenbrock & McDonald 2008). In both *M. oryzae* and *P. infestans,* genetic diversity, as opposed to genomic diversity, is higher in the hypothesized center of origin whereas the spread of rice and potato cultivation is hypothesized to have allowed for the global dispersal of clones (Stukenbrock & McDonald 2008). If wild North American clonal genotypes of *S. cerevisiae* have been influenced by similar demographic patterns, we might expect that their presence in North America is due to dispersal, and that their center of origin is not in North America.

Wild *S. cerevisiae* isolates have previously been postulated to represent either escapees from human associated fermentations (Mortimer 2000; Naumov 1996), or the natural populations from which wine strains were derived (Fay & Benavides 2005). While both hypotheses highlight

the direct role of humans in the domestication of *S. cerevisiae*, one area that has not been as well explored is the indirect role that humans played in the evolution of *S. cerevisiae* through the domestication of the grapevine*.* Similar to the influence crop domestication has had on associated fungal pathogens, it is possible that the domestication of the grapevine has played a key role in shaping diversity within *Saccharomyces* species. The domestication of the wild grapevine is hypothesized to have occurred around 7500-8000 years ago and represents a major demographic transition for *Vitis vinfera* and perhaps, *S. cerevisiae.*

One of the potential scenarios to explain the domestication of *S. cerevisiae* for wine making includes the possibility that during the domestication of grape vines, specific strains of *S. cerevisiae* were preferentially associated with wild grapevines. Under this hypothesis, the close ties between wild grapevines and *S. cerevisiae* could have initially narrowed the genetic pool of *S. cerevisiae* accessible for human to select on and contributed to its use as the predominant species for winemaking. This possibility does not preclude the further domestication of specific wine strains of *S. cerevisiae* for desirable wine making properties, but does provide a hypothesis regarding the clonal nature of wild N. American strains, along with the apparent absence of a North American 'wine' population, and could also explain the apparent domination of *S. cerevisiae* in spontaneous fermentations (Johnston 1990). While this hypothesis cannot be tested with the data included in this study, an analysis of global samples of *S. cerevisiae* isolated from wild and domesticated grapes could elucidate the potential for co-domestication of these two species.

A counterpoint to the population structure and distribution of wild North American *S. cerevisiae* is the population structure and distribution of N. American *S. paradoxus.* The partial sympatry and drastically different population structure between these two species leads one to consider what factors contribute to the observed differentiation. However, much like the complete distribution of *Saccharomyces* in the wild, the mechanisms responsible for habitat differentiation

between *Saccharomyces* species have not been well studied. One well documented difference between the species is that *S. paradoxus* has a lower optimal growth temperature than *S. cerevisiae*, potentially contributing to co-occurrence on oak trees (Sweeney et al. 2004). Additionally, there may be other aspects of growth ability that influence the distribution of these species. Anecdotally, the only site in Oregon (MB) from which wild *S. cerevisiae* was isolated differed from other Oregon locations mostly in the level of humidity, suggesting that differences in desiccation tolerance may play a role in habitat differentiation between the species. Although sampling studies such as this use enrichment procedures that can't provide accurate density measurements, our results suggest that 'wild' *S. cerevisiae* and *S. paradoxus* have different ecological amplitudes, consistent with the observed differences in genetic structure and diversity. It is possible that the ways in which humans have shaped diversity within *S. cerevisiae* have been underestimated, and that like fungal crop pathogens, the genetic diversity of 'wild' populations of *S. cerevisiae* has been shaped by humans indirectly.

Short generation times, amenability to genetic manipulation, and wealth of data about the physiology and genetic architecture of *S. cerevisiae* have made it a robust model organism. This study has further illuminated the ecological and evolutionary forces acting to shape diversity within this species and its close relative *S. paradoxus.* As this dissertation has shown, the processes shaping variation within *Saccharomyces* are clearly complex. However, it is this complexity that makes them particularly suitable for testing hypotheses regarding the interaction of migration, drift, selection and mating systems in both natural and human associated populations. In summary, we have further advanced our understanding of the population genetics and natural history of *S. cerevisiae* and *S. paradoxus*, and for the first time tested specific hypotheses regarding the potential for human selection in shaping phenotypic variation within *S. cerevisiae.*

Literature Cited

A. Majdak et al., 2002. Comparison of Wine Aroma Compounds Produced by *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* Strains. *Food Technol. Biotechnol.*, 40(2), 103-109.

Aa, E. et al., 2006. Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 6(5), 702-715.

Abbott, R.J. & Gomes, M.F., 1989. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity*, 62(3), 411-418.

Aharoni, A. et al., 2004. Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *The Plant Cell Online*, 16(11), 3110.

Altschul, S.F. et al., 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410.

Backhus, L.E. et al., 2001. Functional genomic analysis of a commercial wine strain of *Saccharomyces cerevisiae* under differing nitrogen conditions. *FEMS Yeast Research*, 1(2), 111-125.

Baird, N.A. et al., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE*, 3(10), e3376.

Barbosa, C. et al., 2009. Nitrogen addition influences formation of aroma compounds, volatile acidity and ethanol in nitrogen deficient media fermented by Saccharomyces cerevisiae wine strains. *Journal of Bioscience and Bioengineering*, 108(2), 99-104.

de Barros Lopes, M. et al., 2002. Evidence for multiple interspecific hybridization in

Saccharomyces sensu stricto species. *FEMS Yeast Research*, 1(4), 323-331.

Beck, J.B., Schmuths, H. & Schaal, B.A., 2007. Native range genetic variation in *Arabidopsis thaliana* is strongly geographically structured and reflects Pleistocene glacial dynamics. *Molecular Ecology*, 17(3), 902-915.

Bisson, L.F. & Karpel, J.E., 2010. Genetics of Yeast Impacting Wine Quality. *Annual Review of Food Science and Technology*, 1, 139–162.

Blanco, P, Orriols, I. & Losada, A., 2010. Survival of commercial yeasts in the winery environment and their prevalence during spontaneous fermentations. *Journal of Industrial Microbiology & Biotechnology*. doi: 10.1007/s10295-010-0818-2.

Boulton, R.B. et al., 1996. *Principles and Practices of Winemaking* 1st ed., Springer.

Callejon, R. et al., 2010. Volatile and sensory profile of organic red wines produced by different selected autochthonous and commercial *Saccharomyces cerevisiae* strains. *Analytica Chimica Acta*, 660(1-2), 68-75.

Carrau, F.M. et al., 2008. Production of fermentation aroma compounds by *Saccharomyces cerevisiae* wine yeasts: effects of yeast assimilable nitrogen on two model strains. *FEMS Yeast Research*, 8(7), 1196-1207.

Carson, H.L., Knapp, E.P. & Phaff, H.J., 1956. The Yeast Flora of The Natural Breeding Sites of Some Species of *Drosophila*. *Ecology*, 37(3), 538-544.

Cavalieri, D. et al., 2003. Evidence for *S. cerevisiae* Fermentation in Ancient Wine. *Journal of Molecular Evolution*, 57, S226-S232.

Cruz, F., Vila, C. & Webster, M.T., 2008. The Legacy of Domestication: Accumulation of Deleterious Mutations in the Dog Genome. *Mol Biol Evol*, 25(11), 2331-2336.

Csoma, H. et al., 2010. Biological diversity of *Saccharomyces* yeasts of spontaneously fermenting wines in four wine regions: Comparative genotypic and phenotypic analysis. *International Journal of Food Microbiology*, 140(2-3), 239-248.

de Deken, R.H., 1966. The Crabtree Effect: A Regulatory System in Yeast. *J Gen Microbiol*, 44(2), 149-156.

Diezmann, S. & Dietrich, F.S., 2009. *Saccharomyces cerevisiae*: Population Divergence and Resistance to Oxidative Stress in Clinical, Domesticated and Wild Isolates. *PLoS ONE*, 4(4), e5317.

Doebley, J.F., Gaut, B.S. & Smith, B.D., 2006. The Molecular Genetics of Crop Domestication. *Cell*, 127(7), 1309-1321.

Doniger, S.W. et al., 2008. A Catalog of Neutral and Deleterious Polymorphism in Yeast. *PLoS Genet*, 4(8), e1000183.

Dujon, B., 2010. Yeast evolutionary genomics. *Nat Rev Genet*, 11(7), 512-524.

e-CFR: Title 27: Alcohol, Tobacco and Firearms, Available at: http://ecfr.gpoaccess.gov/cgi/t/text/text-idx?c=ecfr&sid=506cf0c03546efff958847134c5527d3&rgn=div5&view=text&node=27:1.0.1.1.19&idno=27#27:1.0.1.1.19.12.295.7 [Accessed July 27, 2010].

Estevez, P., Gil, M.L. & Falque, E., 2004. Effects of seven yeast strains on the volatile composition of Palomino wines. *International Journal of Food Science and Technology*, 39(1), 61-69.

Etschmann, M. et al., 2008. Production of the aroma chemicals 3-(methylthio)-1-propanol and 3-(methylthio)-propylacetate with yeasts. *Applied Microbiology and*

Biotechnology, 80(4), 579-587.

Eyre-Walker, A. et al., 1998. Investigation of the bottleneck leading to the domestication of maize. *Proceedings of the National Academy of Sciences of the United States of America*, 95(8), 4441 -4446.

Ezeronye, O.U. & Legras, J., 2009. Genetic analysis of *Saccharomyces cerevisiae* strains isolated from palm wine in eastern Nigeria. Comparison with other African strains. *Journal of Applied Microbiology*. 106(5), 1409-1761.

Ezeronye, O. & Legras, J., 2009. Genetic analysis of *Saccharomyces cerevisiae* strains isolated from palm wine in eastern Nigeria. Comparison with other African strains. *Journal of Applied Microbiology*, 106(5), 1569-1578.

Ezov, T.K. et al., 2006. Molecular-genetic biodiversity in a natural population of the yeast *Saccharomyces* cerevisiae from "Evolution Canyon": microsatellite polymorphism, ploidy and controversial sexual status. *Genetics*, 174(3), 1455-68.

Fan, X., 2005. Nature, Cause, and Control of Irradiation-Induced Off-Odor in Ready-to-Eat Meat Products. In *Process and Reaction Flavors*. ACS Symposium Series. American Chemical Society, pp. 208-221. Available at: [Accessed July 27, 2010].

Fay, J.C. & Benavides, J.A., 2005. Evidence for Domesticated and Wild Populations of *Saccharomyces cerevisiae*. *PLoS Genetics*, 1(1), 66-71.

Fay, J.C. et al., 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biology*, 5(4), R26-R26.

Fleet, G.H., 1993. *Wine Microbiology and Biotechnology*, CRC Press.

Fogel, S. et al., 1983. Gene amplification in yeast: CUP1 copy number regulates copper resistance. *Current Genetics*, 7(5), 347-355.

Gangl, H. et al., 2009. Exceptional fermentation characteristics of natural hybrids from *Saccharomyces cerevisiae* and *S. kudriavzevii*. *New Biotechnology*, 25(4), 244-251.

Garijo, P. et al., 2008. The occurrence of fungi, yeasts and bacteria in the air of a Spanish winery during vintage. *International Journal of Food Microbiology*, 125(2), 141-5.

Gealy, D., 2005. Gene movement between rice (*Oryza sativa*) and weedy rice (*Oryza sative*): a US temperate rice perspective. In *Crop Ferality and Volunteerism*. Boca Raton, FL: CRC {ress, pp. 323-354.

Geiser, D.M., Pitt, J.I. & Taylor, J.W., 1998. Cryptic speciation and recombination in the aflatoxin-producing fungus Aspergillus flavus. *Proceedings of the National Academy of Sciences of the United States of America*, 95(1), 388 -393.

Gerke, J.P., Chen, C.T.L. & Cohen, B.A., 2006. Natural Isolates of *Saccharomyces cerevisiae* Display Complex Genetic Variation in Sporulation Efficiency. *Genetics*, 174(2), 985-997.

Gladieux, P. et al., 2010. Evolution of the population structure of *Venturia inaequalis*, the apple scab fungus, associated with the domestication of its host. *Molecular Ecology*, 19(4), 658–674.

Glushakova, A. et al., 2007. Massive isolation and identification of *Saccharomyces*

paradoxus yeasts from plant phyllosphere. *Microbiology*, 76(2), 205-210.

Goddard, M.R. et al., 2010. A distinct population of *Saccharomyces cerevisiae* in New Zealand: evidence for local dispersal by insects and human-aided global dispersal in oak barrels. *Environmental Microbiology*, 12(1), 63-73.

Goffeau, A. et al., 1996. Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287), 546, 563-567.

González, S.S. et al., 2006. Natural hybrids from *Saccharomyces cerevisiae, Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Research*, 6(8), 1221-1234.

Goodrich, J. & Wiener, P., 2005. A walk from the wild side: The genetics of domestication of livestock and crops. *BioEssays*, 27(5), 574-576.

Gressel, J., 2005. *Crop ferality and volunteerism*, CRC Press.

Groth, C., Hansen, J. & Piskur, J., 1999. A natural chimeric yeast containing genetic material from three species. *International journal of systematic bacteriology*, 49 Pt 4, 1933-8.

Hazelwood, L.A. et al., 2008. The Ehrlich Pathway for Fusel Alcohol Production: a Century of Research on *Saccharomyces cerevisiae* Metabolism. *Appl. Environ. Microbiol.*, 74(12), 3920.

Heard, G. & Fleet, G., 1990. A convenient microtitre tray procedure for yeast identification. *Journal of Applied Microbiology*, 68(5), 447-451.

Herjavec, S. et al., 2003. The influence of some commercial *Saccharomyces cerevisiae* strains on the quality of Chardonnay wines. *Food Technology and Biotechnology*,

41(1), 77-81.

Hesseltine, C.W., 1965. A Millennium of Fungi, Food, and Fermentation. *Mycologia*, 57(2), 149-197.

Howell, K.S. et al., 2004. Variation in 4-mercapto-4-methyl-pentan-2-one release by *Saccharomyces cerevisiae* commercial wine strains. *FEMS Microbiology Letters*, 240(2), 125-129.

Izawa, T., 2007. Adaptation of flowering-time by natural and artificial selection in *Arabidopsis* and rice. *Journal of Experimental Botany*, 58(12), 3091 -3097.

Jakobsson, M. & Rosenberg, N.A., 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801-1806.

Johnson, L.J. et al., 2004. Population Genetics of the Wild Yeast *Saccharomyces paradoxus*. *Genetics*, 166(1), 43-52.

Johnston, J., 1990. Brewing and distilling yeasts. In *Yeast Technology*. Berlin, Germany: Springer-Verlag, pp. 55-104.

Katou, T. et al., 2008. Brewing characteristics of haploid strains isolated from sake yeast Kyokai No. 7. *Yeast*, 25(11), 799-807.

Katou, T. et al., 2009. QTL mapping of sake brewing characteristics of yeast. *Journal of Bioscience and Bioengineering*, 107(4), 383-393.

Kim, S. et al., 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet*, 39(9), 1151-1155.

Koufopanou, V. et al., 2006. The spatial scale of genetic differentiation in a model

organism: the wild yeast *Saccharomyces paradoxus*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1475), 1941-1946.

Kovach, M.J. et al., 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences*, 106(34), 14444 -14449.

Kuehne, H.A. et al., 2007. Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations. *Current Biology: CB*, 17(5), 407-411.

Kvitek, D.J., Will, J.L. & Gasch, A.P., 2008. Variations in Stress Sensitivity and Genomic Expression in Diverse *S. cerevisiae* Isolates. *PLoS Genet*, 4(10), e1000223.

Langmead, B. et al., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3), R25.

Legras, J.L. et al., 2007. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular ecology*, 16(10), 2091-102.

Li, H. et al., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

Li, S. et al., 2010. Yeast species associated with wine grapes in China. *International Journal of Food Microbiology*, 138(1-2), 85-90.

Lindberg, J. et al., 2007. Selection for tameness modulates the expression of heme related genes in silver foxes. *Behav Brain Funct*, 3, 18.

Liti, G. et al., 2009a. Population genomics of domestic and wild yeasts. *Nature*, 458(7236), 337-341.

Londo, J.P. & Schaal, B.A., 2007. Origins and population genetics of weedy red rice in

the USA. *Molecular Ecology*, 16(21), 4523-4535.

Lopandic, K. et al., 2007. Genetically different wine yeasts isolated from Austrian vine-growing regions influence wine aroma differently and contain putative hybrids between *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. *FEMS Yeast Research*, 7(6), 953-965.

Loscos, N. et al., 2007. Release and Formation of Varietal Aroma Compounds during Alcoholic Fermentation from Nonfloral Grape Odorless Flavor Precursors Fractions. *Journal of Agricultural and Food Chemistry*, 55(16), 6674-6684.

Lu, J. et al., 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends in Genetics*, 22(3), 126-131.

Mable, B.K. & Adam, A., 2007. Patterns of genetic diversity in outcrossing and selfing populations of *Arabidopsis lyrata*. *Molecular Ecology*, 16(17), 3565-3580.

Maharjan, R. et al., 2006. Clonal Adaptive Radiation in a Constant Environment. *Science*, 313(5786), 514-517.

Masneuf-Pomarède, I. et al., 2006. Influence of fermentation temperature on volatile thiols concentrations in Sauvignon blanc wines. *International Journal of Food Microbiology*, 108(3), 385-390.

McCullough, M.J. et al., 1998. Intergenic Transcribed Spacer PCR Ribotyping for Differentiation of *Saccharomyces* Species and Interspecific Hybrids. *J. Clin. Microbiol.*, 36(4), 1035-1038.

McDonald, B.A. & Linde, C., 2002. The population genetics of plant pathogens and breeding strategies for durable resistance. *Euphytica*, 124(2), 163–180.

McGovern, P.E. et al., 2004. Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17593-8.

McGovern, P.E., 2003. *Ancient wine : the search for the origins of viniculture*, Princeton University Press.

Meilleur, B.A. & Hodgkin, T., 2004. In situ conservation of crop wild relatives: status and trends. *Biodiversity and Conservation*, 13(4), 663-684.

Mendes-Ferreira, A. et al., 2009. The production of hydrogen sulphide and other aroma compounds by wine strains of *Saccharomyces cerevisiae* in synthetic media with different nitrogen concentrations. *Journal of Industrial Microbiology & Biotechnology*, 36(4), 571-583.

Molina, A. et al., 2009. Differential synthesis of fermentative aroma compounds of two related commercial wine yeast strains. *Food Chemistry*, 117(2), 189-195.

Mortimer, R.K., 2000. Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome research*, 10(4), 403-9.

Mortimer, R. & Polsinelli, M., 1999. On the origins of wine yeast. *Research in Microbiology*, 150(3), 199-204.

Murphy, H.A. et al., 2006. Mate choice assays and mating propensity differences in natural yeast populations. *Biology Letters*, 2(4), 553–556.

Murphy, H.A. & Zeyl, C.W., 2010. Yeast Sex: Surprisingly High Rates of Outcrossing between Asci. , 5(5), e10461..

Nardi, T. et al., 2006. A rapid method for differentiating *Saccharomyces* sensu stricto

strains from other yeast species in an enological environment. *FEMS Microbiology Letters*, 264(2), 168-173.

Naumov, G.I., Naumova, E.S. & Sniegowski, P.D., 1998. *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. *Canadian journal of microbiology*, 44(11), 1045–1050.

Naumov, G.I., Gazdiev, D.O. & Naumova, E.S., 2003. The Finding of the Yeast Species *Saccharomyces bayanus* in Far East Asia. *Microbiology*, 72(6), 738-743.

Naumov, G.I. et al., 2002. *Saccharomyces bayanus* var. uvarum in Tokaj wine-making of Slovakia and Hungary. *Applied Microbiology and Biotechnology*, 59(6), 727-730.

Naumov, G.I., Naumova, E.S. & Sniegowski, P.D., 1997. Differentiation of European and Far East Asian Populations of *Saccharomyces paradoxus* by Allozyme Analysis. *Int J Syst Bacteriol*, 47(2), 341-344.

Naumov, G. et al., 2000. Association of *Saccharomyces bayanus* var. uvarum with some French wines: genetic analysis of yeast populations. *Research in Microbiology*, 151(8), 683-691.

Naumov, G., 1996. Genetic identification of biological species in the *Saccharomyces* sensu stricto complex. *Journal of Industrial Microbiology & Biotechnology*, 17(3-4), 295-302.

Naumova, E.S. et al., 2005. The molecular genetic differentiation of cultured *Saccharomyces* strains. *Microbiology*, 74(2), 179-187.

Nguyen, H. & Gaillardin, C., 2005. Evolutionary relationships between the former species *Saccharomyces uvarum* and the hybrids *Saccharomyces bayanus* and

Saccharomyces pastorianus; reinstatement of *Saccharomyces uvarum* (Beijerinck) as a distinct species. *FEMS Yeast Research*, 5(4-5), 471-483.

Noble, A.C. et al., 1987. Modification of a Standardized System of Wine Aroma Terminology. *Am. J. Enol. Vitic.*, 38(2), 143-146.

Nordborg, M. et al., 2005. The Pattern of Polymorphism in *Arabidopsis thaliana*. *PLoS Biol*, 3(7), e196.

Nurgel, C. et al., 2002. Influence of *Saccharomyces cerevisiae* strains on fermentation and flavor compounds of white wines made from cv. Emir grown in Central Anatolia, Turkey. *Journal of Industrial Microbiology & Biotechnology*, 29(1), 28-33.

Olsen, K.M. & Purugganan, M.D., 2002. Molecular Evidence on the Origin and Evolution of Glutinous Rice. *Genetics*, 162(2), 941-950.

Orlić, S. et al., 2010. A comparative study of the wine fermentation performance of *Saccharomyces paradoxus* under different nitrogen concentrations and glucose/fructose ratios. *Journal of Applied Microbiology*, 108(1), 73-80.

Orlic, S. et al., 2007. Influence of indigenous *Saccharomyces paradoxus* strains on Chardonnay wine fermentation aroma. *Int J Food Sci Tech Int J Food Sci Tech*, 42(1), 95-101.

Orlić, S. et al., 2010. Diversity and oenological characterization of indigenous *Saccharomyces cerevisiae* associated with Žilavka grapes. *World Journal of Microbiology and Biotechnology*, 26(8), 1483-1489.

Park, H. & Bakalinsky, A.T., 2000. SSU1 mediates sulphite efflux in *Saccharomyces*

cerevisiae. *Yeast (Chichester, England)*, 16(10), 881-8.

Pasteur, L., 1866. *âEtudes sur le vin ses maladies, causes qui les provoquent, procâedâes nouveaux pour le conserver et pour le vieillir*, Paris: Imprimerie impâeriale,

Phaff, H.J. & Knapp, E.P., 1956. The taxonomy of yeasts found in exudates of certain trees and other natural breeding sites of some species of *Drosophila*. *Antonie van Leeuwenhoek*, 22(1), 117-130.

Pramateftaki, P.V., Lanaridis, P. & Typas, M.A., 2000. Molecular identification of wine yeasts at species or strain level: a case study with strains from two vine-growing areas of Greece. *Journal of applied microbiology*, 89(2), 236-48.

Pretorius, I.S., 2000. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast*, 16(8), 675-729.

Pritchard, J.K., Stephens, M. & Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-59.

Querol, A., JiméNez, M. & Huerta, T., 1990. Microbiological and Enological Parameters during Fermentation of Musts from Poor and Normal Grape-Harvests in the Region of Alicante (Spain). *Journal of Food Science*, 55(6), 1603-1606.

R Development Core Team, 2009. *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org.

Redzepovic, S. et al., 2002. Identification and characterization of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* strains isolated from Croatian vineyards. *Letters in Applied Microbiology*, 35(4), 305-310.

Regodón Mateos, J., Pérez-Nevado, F. & Ramírez Fernández, M., 2006. Influence of *Saccharomyces cerevisiae* yeast strain on the major volatile compounds of wine. *Enzyme and Microbial Technology*, 40(1), 151-157.

Replansky, T. et al., 2008. *Saccharomyces* sensu stricto as a model system for evolution and ecology. *Trends in Ecology & Evolution*, 23(9), 494-501.

Reuter, M., Bell, G. & Greig, D., 2007. Outbreeding in yeast in response to dispersal by an insect vector. *Current Biology*, 17(3), R81-R83.

Rokas, A., 2009. The effect of domestication on the fungal proteome. *Trends in Genetics*, 25(2), 60-63.

Romano, P. et al., 2003. Metabolic diversity of *Saccharomyces cerevisiae* strains from spontaneously fermented grape musts. *World Journal of Microbiology and Biotechnology*, 19(3), 311-315.

Rosenberg, N.A., 2003. distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137-138.

Ross-Ibarra, J. et al., 2008. Patterns of Polymorphism and Demographic History in Natural Populations of *Arabidopsis lyrata*. *PLoS ONE*, 3(6).

Rothstein, R.J., 1977. A genetic fine structure analysis of the suppressor 3 locus in Saccharomyces. *Genetics*, 85(1), 55-64.

Rothstein, R.J., Esposito, R.E. & Esposito, M.S., 1977. The effect of ochre suppression on meiosis and ascospore formation in *Saccharomyces*. *Genetics*, 85(1), 35-54.

Ruderfer, D.M. et al., 2006. Population genomic analysis of outcrossing and recombination in yeast. *Nat Genet*, 38(9), 1077-1081.

Sampaio, J.P. & Gonçalves, P., 2008. Natural Populations of *Saccharomyces kudriavzevii* in Portugal Are Associated with Oak Bark and Are Sympatric with *S. cerevisiae* and *S. paradoxus*. *Applied and Environmental Microbiology*, 74(7), 2144–2152.

Schacherer, J. et al., 2009. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature*, 458(7236), 342-345.

Schuller, D. et al., 2005. Ecological survey of *Saccharomyces cerevisiae* strains from vineyards in the Vinho Verde Region of Portugal. *FEMS Microbiology Ecology*, 51(2), 167-177.

Schuller, D. & Casal, M., 2006. The genetic structure of fermentative vineyard-associated *Saccharomyces cerevisiae* populations revealed by microsatellite analysis. *Antonie van Leeuwenhoek*, 91(2), 137-150.

Serpell, J., 1995. *The domestic dog: its evolution, behaviour, and interactions with people*, Cambridge University Press.

Shehata, A.M.E.T., Mrak, E.M. & Phaff, H.J., 1955. Yeasts Isolated from *Drosophila* and from Their Suspected Feeding Places in Southern and Central California. *Mycologia*, 47(6), 799-811.

Sniegowski, P.D., Dombrowski, P.G. & Fingerman, E., 2002. *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research*, 1(4), 299-306.

Spor, A. et al., 2009. Niche-driven evolution of metabolic and life-history strategies in natural and domesticated populations of *Saccharomyces cerevisiae*. *BMC*

Evolutionary Biology, 9(1), 296.

Spor, A. et al., 2008. “Ant” and “Grasshopper” Life-History Strategies in *Saccharomyces cerevisiae*. *PLoS ONE*, 3(2), e1579.

Spratt, B.G. & Maiden, M.C.J., 1999. Bacterial population genetics, evolution and epidemiology. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 354(1384), 701 -710.

Stone, H. & Sidel, J.L., 2004. *Sensory evaluation practices*, Academic Press.

Stukenbrock, E.H. & McDonald, B.A., 2008. The Origins of Plant Pathogens in Agro-Ecosystems. *Annual Review of Phytopathology*, 46(1), 75-100.

Sweeney, J.Y., Kuehne, H.A. & Sniegowski, P.D., 2004. Sympatric natural *Saccharomyces cerevisiae* and *S. paradoxus* populations have different thermal growth profiles. *FEMS Yeast Research*, 4(4-5), 521-525.

Swiegers, J. & Pretorius, I., 2007. Modulation of volatile sulfur compounds by wine yeast. *Applied Microbiology and Biotechnology*, 74(5), 954-960.

Swiegers, J.H. et al., 2009. The influence of yeast on the aroma of Sauvignon Blanc wine. *Food Microbiology*, 26(2), 204-211.

Swiegers, J.H. & Pretorius, I.S., 2005. Yeast Modulation of Wine Flavor. *Advances in applied microbiology*, Volume 57, 131-175.

Tamura, K. et al., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24(8), 1596-1599.

Templeton, A.R., 2006. *Population Genetics and Microevolutionary Theory* 1st ed., Wiley-Liss.

This, P., Lacombe, T. & Thomas, M.R., 2006. Historical origins and genetic diversity of wine grapes. *Trends in Genetics*, 22(9), 511-519.

Thomson, J.M. et al., 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet*, 37(6), 630-635.

Török, T. et al., 1996. Quest for wine yeasts—An old story revisited. *Journal of Industrial Microbiology & Biotechnology*, 17(3-4), 303-313.

Tsai, I.J. et al., 2008. Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proceedings of the National Academy of Sciences*, 105(12), 4957 -4962.

Ugliano, M. et al., 2009. Effect of Nitrogen Supplementation and *Saccharomyces* Species on Hydrogen Sulfide and Other Volatile Sulfur Compounds in Shiraz Fermentation and Wine. *Journal of Agricultural and Food Chemistry*, 57(11), 4948-4955.

Valero, E. et al., 2007. Biodiversity of *Saccharomyces* yeast strains from grape berries of wine-producing areas using starter commercial yeasts. *FEMS Yeast Research*, 7(2), 317-329.

Valero, E. et al., 2005. Dissemination and survival of commercial wine yeast in the vineyard: a large-scale, three-years study. *FEMS Yeast Research*, 5(10), 959-969.

Versavaud, A. et al., 1995. Genetic diversity and geographical distribution of wild *Saccharomyces cerevisiae* strains from the wine-producing area of Charentes, France. *Applied and environmental microbiology*, 61(10), 3521-9.

Vilela-Moura, A. et al., 2010. Effects of acetic acid, ethanol, and SO2 on the removal of

volatile acidity from acidic wines by two *Saccharomyces cerevisiae* commercial strains. *Applied Microbiology and Biotechnology*, 87(4), 1317-1326.

Watterson, G.A., 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7(2), 256-276.

Will, J.L. et al., 2010. Incipient Balancing Selection through Adaptive Loss of Aquaporins in Natural *Saccharomyces cerevisiae* Populations. *PLoS Genet*, 6(4), e1000893.

Winzeler, E.A. et al., 2003. Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics*, 163(1), 79-89.

Wondra, M. & Berovic, M., 2001. Analyses of aroma components of chardonnay wine fermented by different yeast strains. *Food Technology and Biotechnology*, 39(2), 141-148.

Young, L. et al., 2002. Compounds Responsible for the Odor of Dog Hair Coat. In *WSAVA 2002 Congress Proceedings*. WSAVA World Congress. Granada.

Yurkov, A.M., 2005. First Isolation of the Yeast *Saccharomyces paradoxus* in Western Siberia. *Microbiology*, 74(4), 459-462.

Zeder, M.A., 2006. Central questions in the domestication of plants and animals. *Evolutionary Anthropology: Issues, News, and Reviews*, 15(3), 105-117.

Zhang, H. et al., 2010. *S. paradoxus* and *S. cerevisiae* reside on oak trees in New Zealand: evidence for migration from Europe and inter-species hybrids. *FEMS Yeast Research*, no.doi: 10.1111/j.1567-1364.2010.00681.x.

Zohary, D. & Hopf, M., 2000. *Domestication of plants in the old world: the origin and*

spread of cultivated plants in West Asia, Europe, and the Nile Valley, Oxford University Press.