

INTO THE WOODS: THE PHYLOGEOGRAPHY AND SPECIES DISTRIBUTION
OF WILD YEAST FROM FORESTS

by

JACQUELINE JOYE PEÑA

(Under the Direction of Douda Bensasson)

ABSTRACT

There is an urgent need to understand how global climate change affects natural populations. Combined approaches in ecology and evolution are proving effective at determining how past climate change has structured current populations for plants and animals, but this is less clear for fungi. In this dissertation, I leveraged wild budding yeast as a model system to study how climate and human migration may affect natural populations. First, I examined over 300 genomes of wild *Saccharomyces cerevisiae* from trees to determine if wild tree-associated populations are genetically distinct from domesticated lineages and estimated the approximate times when wild lineages diverged. This revealed population substructure within European, Japanese, and North American forest habitats, and the divergence of wild forest lineages roughly coincided with significant events since the last glacial maximum. Major yeast migration events out of Asia and into North America and Europe occurred during the onset of agriculture and human use of fermentation, suggesting footprints of human migration in wild *S. cerevisiae* population structure. Second, I conducted an extensive field survey to isolate a broad range of sympatric wild yeasts, mainly from oak trees to test for associations

between forest yeast species ranges and climate. This yielded over 450 wild yeast strains across a natural temperature gradient along the eastern United States. The most abundant species isolated were *S. cerevisiae*, *S. paradoxus*, and *L. thermotolerans*. For all three species, I found that yeast prevalence is associated with climate , and for one species, *S. paradoxus*, it is additionally associated with tree age: older trees harbor more *S. paradoxus*. For wild *S. cerevisiae*, the results suggested a thermal optimum from field results. This optimum is consistent with past thermal growth profiles estimated in laboratory experiments, and with the geographic distribution of isolates from wild forest lineages in past studies. Together, these studies suggest that climate and human migration have shaped the phylogeography and species distribution of wild yeast from forests, improving our understanding of how past climate change may have structured natural fungal populations.

INDEX WORDS: Population structure, phylogenetics, species range limits, climate change, human migration

INTO THE WOODS: THE PHYLOGEOGRAPHY AND SPECIES DISTRIBUTION
OF WILD YEAST FROM FORESTS

by

JACQUELINE JOYE PEÑA

B.S., Boise State University, 2016

M.S., Utah State University, 2018

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2025

©2025

Jacqueline J. Peña

All Rights Reserved

INTO THE WOODS: THE PHYLOGEOGRAPHY AND SPECIES DISTRIBUTION
OF WILD YEAST FROM FORESTS

by

JACQUELINE JOYE PEÑA

Major Professor: Douda Bensasson

Committee: Michelle Momany
Jill T. Anderson
Kelly A. Dyer
Marin T. Brewer

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2025

DEDICATION

To the yeasts, without them this work would not exist. Learning about your ecology and evolution has been a pleasure.

ACKNOWLEDGEMENTS

I am living my ancestors' wildest dreams because academic spaces were never meant for someone like me – a Latina. I thank God, family, friends, pets, and mentors for helping me grow into the scientist, especially the person I am today. Without my family and community's support, pursuing a PhD would be impossible.

First, I want to acknowledge the Indigenous lands where my research was performed, which includes but is not limited to the Cherokee, Yuchi, and Muscogee Indigenous lands. Cherokee land includes Georgia, Alabama, South Carolina, Tennessee, Kentucky, West Virginia, and Virginia. Yuchi land includes Kentucky, Tennessee, Alabama, Georgia, South Carolina, and North Carolina. Muscogee land includes Georgia, Alabama, and Florida.

Second, I thank my family for all their support and love throughout this tremendous journey. My paternal grandparents and parents taught me that higher education is the way out of poverty and debt. I want to acknowledge my roots – I come from a long line of hardworking and resourceful ancestors. I am grateful for my dad, who taught me how to be resilient and for being my most prominent hype man. My dad never lost faith in me, even when I doubted my own capabilities. I thank my mom for being my rock and always knowing how to lift my spirits, especially during the dark times. I thank my sister for being the sunshine in my life, she knows how to make me smile and encourages me to be brave. I thank my husband, Prince Duepa, for being my best friend and always supporting me in all aspects of my life. I love how he makes me laugh and

reminds me to live in the moment. I appreciate him letting me geek out about my research and encouraging me to pursue my research dreams. Lastly, I am grateful for my pets and passed away pets: Kit, Cleo, Kobe, and Ottis. Kit has given me so much joy, and Cleo has been a cute little menace. Kobe and Ottis were childhood dogs that brought me so much joy when I came home, and they are deeply missed.

Third, I thank all my academic mentors who have helped me scientifically and personally. My research advisor, Douda Bensasson, has supported me in many ways. I am genuinely grateful to have learned from a brilliant scientist and kind person. Her guidance on scientific rigor, evolutionary biology, ecology, coding, and all aspects of "professor finishing school" were instrumental to my career. Douda always knew how to keep me focused, and she created a comfortable learning environment where I could bring my whole self to the lab. For their mentorship and guidance, I would like to thank my advisory committee: Michelle Momany, Jill Anderson, Kelly Dyer, and Marin Brewer. I have learned so much from these brilliant women in science! I thank past and present members of the Bensasson lab for support and excellent scientific feedback: Eduardo Scopel Ferreira da Costa, Jenna Hamlin, Audrey Ward, Sydney McCall, Miranda McKibben, Rosemary Wills, Diana Ambrocio, Linda Habershaw, Domenic Won, Frema Owusu-Ansah, Oliver Nemeth, and Grady Waple. I thank my lab neighbors, Team Momany, for their camaraderie and insightful scientific advice: Alex Mela, Earl Kang, Brent Shuman, Bran Celia-Sanchez, Justina Stanislaw, and Agmsofi Uddin Mahamud. I thank Paola Barriga, Peggy Brickman, John Burke, Shu-mei Chang, Anny Chung, Wolfgang Lukowitz, Andrea Sweigart, Tanisha Williams, and Tati Russo-tait for their mentorship and support. I especially want to thank Angela Birkes, Walter Schmidt,

former Dean Suzanne Barbour, and Dean Ron Walcott for supporting me in my early PhD years and beyond. I thank Lisa Moncrief, Chelsea Hargroves, Reagan Scott, and Jeannie Nguyen for keeping me organized and handling departmental logistics.

Fourth, I am genuinely grateful for all the communities I've joined or created. Life has been more enjoyable with the friends I've made. I am grateful to the Plant Biology Graduate Student Association (PBGSA), Mycology Graduate Student Organization (MGSO), Fungal Group, PBIO1210L TA community, Diversify STEM Coffee Hour community, Bridges to the Doctorate (B2D), Louis Stokes Alliance for Minority Participation (LSAMP), HHMI Gilliam community, and Advocates for Anti-Racism in Life Sciences (AARLS). Then there's my trivia gang: Rachel and Andrew Perez-Undell, Phil and Jenn Benz, Patrick and Brittany Smallwood; my stitch 'n' bitch group: Mia Rochford, Chazz Jordan, Rachel Perez-Undell, Ashley Earley, Brent Shuman, Summerlin Courchaine and Bran Celia-Sanchez; and lastly other friends that I've made: Margot Popecki, Felicia Asem, Neely Palacio, Callie Oldfield, Mary Washburn, KB McCrum, Riley Thoen, Justine Valadez, Anna Wyngaarden, Brandon Mangum, Clayton Hale, Rhys Eshleman, and Emma Chandler.

For chapter three: I thank Tyler Arboretum, Duke Forest Teaching and Research Laboratory, Bankhead National Forest, Thompson Mills Forest, the University of Georgia, the Institute of Oceanography, and the Florida State Parks Service for allowing me and the Bensasson lab to conduct fieldwork.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	xii
LIST OF FIGURES	xv
CHAPTER	
1. INTRODUCTION AND LITERATURE REVIEW	1
1.1. Global climate change is altering natural populations.....	1
1.2 Fungi are underrepresented in climate change research.....	2
1.3 Wild budding yeast – models for studying how climate affects natural populations	3
1.4 Dissertation outline and specific aims.....	4
1.5 References.....	6
2. FOOTPRINTS OF HUMAN MIGRATION IN THE POPULATION STRUCTURE OF WILD BAKER'S YEAST	11
2.1 Abstract.....	12
2.2 Introduction.....	13
2.3 Materials and methods	15
2.3.1 Yeast strains and genome data.....	15
2.3.2 Read mapping and base calling.....	16
2.3.3 Quality filtering steps.....	17

2.3.4 Population structure and genetic admixture.....	18
2.3.5 Population substructure within Europe.....	20
2.3.6 <i>In silico</i> chromosome painting.....	20
2.3.7 Time divergence analysis.....	21
2.4 Results.....	24
2.4.1 Tree habitats harbor numerous genetically distinct lineages	24
2.4.2 Population substructure in the forest niches of North America and Japan.....	25
2.4.3 Fine-scale population structure of wild <i>S. cerevisiae</i> from Europe.....	26
2.4.4 Out-of-Asia migration of forest yeast since the last glacial maximum..	27
2.4.5 Occasional strains in Europe resemble present-day North American lineages.....	28
2.5 Discussion	30
2.5.1 Genetic isolation in forest niches.....	30
2.5.2 Global spread of <i>S. cerevisiae</i> forest populations since the last glacial maximum.....	32
2.5.3. Ongoing migration between human and tree environments.....	35
2.5.4 Footprints of human activity in the genomes of European tree strains... <td>36</td>	36
2.6 Conclusion.....	37
2.7 Acknowledgements.....	38
2.8 References.....	38
2.9 Data accessibility and benefit-sharing.....	48
2.10 Author contributions.....	48

3. CLIMATE CAN PREDICT THE SPECIES RANGES OF SYMPATRIC YEASTS	
FROM FORESTS	54
3.2 Introduction.....	54
3.3 Methods.....	56
3.3.1 Collecting environmental samples in the eastern United States.....	56
3.3.2 Microclimate and longitudinal sampling from Georgia.....	57
3.3.3 Culturing and identification of wild yeast species.....	58
3.3.4 Testing for associations between yeast prevalence and ecological or climatic factors.....	60
3.3.5 Worldwide presence and absence data for wild yeast species.....	63
3.4 Results.....	64
3.4.1 A broad range of yeast species from North American forests and marsh	66
3.4.2 Yeast prevalence on bark samples and sampling strategy.....	66
3.4.3 <i>Saccharomyces cerevisiae</i> prevalence is associated with summer temperature.....	68
3.4.4 <i>Lachancea fermentati</i> prevalence is associated with summer temperature and precipitation.....	70
3.4.5 <i>Saccharomyces paradoxus</i> prevalence is associated with temperature, precipitation, and tree age.....	71
3.5 Discussion	73
3.5.1 Salt marshes have a different species composition compared to forests.....	74

3.5.2 Climate and the species ranges of budding yeast from forests.....	75
3.5.3 The ecology of <i>S. paradoxus</i> is complicated.....	77
3.6 References.....	78
3.7 Author contributions.....	86
4 CONCLUSION AND FUTURE DIRECTIONS.....	99
4.1 References.....	101

APPENDICES

A. SUPPLEMENTAL INFORMATION FOR CHAPTER 2.....	103
B. SUPPLEMENTAL INFORMATION FOR CHAPTER 3.....	139

LIST OF TABLES

	Page
Table S2.1: Description of <i>S. cerevisiae</i> strains used in this study, quality filter applied (i.e. passed, heterozygous,..) , genome data, geographical origin, region (i.e. Asia, Europe,..), ecological origin, ecological category (wild, tree-associated, and human-associated), monosporic derivative (yes or no), intraspecies contamination (yes or no), average genome coverage, published clade assignment, clade assignment for this study, genome-wide levels of heterozygosity, and reference for strain source and genome source.....	101
Table S2.2: Description of <i>S. cerevisiae</i> tree-associated and reference panel strains that made it past quality filters for population structure analyses.	101
Table S2.3: Description of tree-associated <i>S. cerevisiae</i> strains from Europe for population substructure analyses (N = 51).	101
Table S2.4 Descripiton of <i>S. cerevisiae</i> strains randomly selected to represent 25 genetically distinct clades defined in this study to be used as a backbone phylogeny (N = 75).....	101
Table S2.5: Chromosome painting results for backbone <i>S. cerevisiae</i> strains used for time divergence analyses, the primary clade assignment based on chromosome painting analysis (cladeName), the percent primary clade assignment (percentCladeSimilarity), the secondary clade assignment	

(secondMostCommonNearestClade), the percent secondary clade assignment (percentSecondMostCommonNearestClade), and quality filter applied.	101
Table S2.6: Description of genomic regions (60-90 kb) selected for each strain and chromosome for time divergence analysis, showing strain order in Figure S11, chromosome number, chromosome location (coordinates), and notes.	101
Table S2.7: Approximate time estimates at which modern tree-associated lineages arose (i) outside Asia (Out-of-Asia); (ii) in North America (North American oak split); and (iii) in Europe (Wine-European oak split).	101
Table S2.8: Historic temperatures at field sites with trees harboring <i>Saccharomyces</i> <i>cerevisiae</i>	101
Table 3.1: Yeast isolated from different habitats	86
Table S3.1: Environmental samples collected between 2019 to 2022 from the United States used for bark position models.	137
Table S3.2: Daily temperature recorded between November 2021 - November 2022 every four hours from 5 <i>Quercus alba</i> trees at Thompson Mills Forest in Braselton, Georgia.....	137
Table S3.3: Yeast species identified using DNA sequences queried against the nucleotide database restricted to Ascomycete fungi (txid4890) using blastn algorithm at NCBI.....	137
Table S3.4: Summary of yeast isolated from oak bark samples partitioned by tree hosts from the United States (N = 102 hosts) and Europe (N = 92 hosts; Robinson et al., 2016) used for climate models.....	137

Table S3.5: Presence and absence data for <i>S. cerevisiae</i> and <i>L. fermentati</i> associated with plants from published studies used for climate envelope maps.....	137
---	-----

LIST OF FIGURES

	Page
Figure 2.1: Trees harbor numerous genetically distinct <i>S. cerevisiae</i> lineages with population substructure in North America and Japan.....	49
Figure 2.2: Fine-scale population structure of wild <i>S. cerevisiae</i> from Europe.	50
Figure 2.3: Out-of-Asia migration of forest yeast since the Last Ice Age.....	51
Figure 2.4: Occasional strains in Europe resemble present-day North American lineages.....	52
Figure S2.1: Genome-wide levels of heterozygosity of <i>Saccharomyces cerevisiae</i> partitioned by tree-associated (N = 172) (A) and non-tree-associated strains (N = 881) (B) after excluding monosporic derivatives.....	102
Figure S2.2: Genome-wide levels of heterozygosity for <i>Saccharomyces cerevisiae</i> of tree-associated (N = 172) (A) and non-tree-associated strains (N = 881) (B) partition by published clades after excluding monosporic derivatives.....	103
Figure S2.3: Loglikelihood values from ADMIXTURE analyses for each cluster (K) across five replicate runs among 313 <i>S. cerevisiae</i> tree-associated and reference panel strains (non-tree-associated) strains.....	104
Figure S2.4: Population structure and admixture of <i>Saccharomyces cerevisiae</i> tree-sampled strains and reference panel strains after excluding strains with heterozygosity > 0.001.....	105
Figure S2.5: Phylogenetic relationships of <i>Saccharomyces cerevisiae</i>	107

Figure S2.6: Loglikelihood values from ADMIXTURE analyses for each cluster (K) across five replicate runs among 51 wild <i>S. cerevisiae</i> from European woodlands.....	108
Figure S2.7: The population structure and admixture of wild <i>Saccharomyces cerevisiae</i> from Europe.....	109
Figure S2.8: Chromosome painting for 77 strains used as a backbone phylogeny.....	110
Figure S2.9: Histograms of within-clade and between-clade comparisons for North American A-C (A-C).....	119
Figure S2.10: Histograms of within-clade and between-clade comparisons for Wine/European (A) and European oak (B).....	120
Figure S2.11: in silico chromosome painting results of backbone phylogeny strains grouped by chromosome of strains that made it past quality filters for time divergence analysis.....	121
Figure S2.12: Time divergence results of selected 60-90 kb genomic regions of backbone phylogeny strains for each chromosome.....	128
Figure 3.1: Environmental sample locations across the eastern United States from September 2019 to October 2022.	87
Figure 3.2: Yeast isolation frequency from <i>Quercus alba</i> and <i>Q. virginiana</i> across all field sites visited in 2021.	88
Figure 3.3: The proportion of yeast isolates from <i>Quercus alba</i> and <i>Q. virginiana</i> across multiple years in Georgia.....	89

Figure 3.4: <i>Saccharomyces cerevisiae</i> isolation frequency increases with increasing temperature (°C) of the warmest month (Tmax) for 196 <i>Quercus</i> sp. from Europe (A) and eastern United States (B)	90
Figure 3.5: Optimal summer temperature (°C) can predict the species range of wild <i>S. cerevisiae</i> associated with plants and the distribution of wild lineages.....	91
Figure 3.6: The combined effect of temperature and precipitation for the isolation frequency of <i>L. fermentati</i> isolation frequency for 102 <i>Quercus</i> sp. from northern and southern United States.....	93
Figure 3.7: There are no regions of Europe with high summer temperature and precipitation suitable for <i>L. fermentati</i> that are similar to regions in the United States.....	94
Figure 3.8: <i>Saccharomyces paradoxus</i> frequency increases on older trees (years) for 196 <i>Quercus</i> sp. from southern and northern Europe (A) and the eastern United States (B).....	95
Figure 3.9: The combined effect of temperature (°C) of the warmest month (Tmax) (A) and precipitation (mm) of the warmest quarter for the isolation frequency of <i>S. paradoxus</i> from 196 <i>Quercus</i> sp. from Europe and the United States	96
Figure S3.1: Mean monthly temperature (°C) where tree hosts are likely to experience the (A) hottest (4:00 pm) and (B) coldest (4:00 am) temperatures during a 24-hour cycle from the north and south sides of <i>Quercus alba</i> at Thompson Mills Forest in Braselton, Georgia.	138

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Global climate change is altering natural populations

Climate change is causing elevated temperature, elevated CO₂, and altered precipitation, which is rapidly altering ecosystems (Sage, 2019). A research priority is predicting how species, communities, and ecosystems respond to ongoing and future climate change (Boult & Evans, 2021; Moritz & Agudo, 2013). Climate change impacts the ecology of species by shifting species distributions to higher elevations (Lafontaine et al., 2018) or to northern latitudes (Osland et al., 2021). These shifts result in changes in population size (Kleinheselink & Adler, 2018) and mismatches between seasonal cues and phenology (Buckley & Kingsolver, 2019). These alterations, in turn, may affect phenotypic evolution, genetic differentiation, and speciation (Hendry, 2016; Rodríguez-Verdugo et al., 2017), where warming temperatures could accelerate or hinder adaptation (McGaughran et al., 2021). Therefore, we need interdisciplinary approaches for mitigation strategies to identify what makes a species vulnerable to climate change and locate these vulnerable populations (Boult & Evans, 2021; Pacifici et al., 2015).

Combined approaches from ecology and evolution are necessary to understand how populations will respond to future climate scenarios (Araújo et al., 2019; Lafontaine et al., 2018; McGaughran et al., 2021; Pacifici et al., 2015) because evolution and ecology can occur on the same time-scale (Carroll et al., 2007; Fussmann et al., 2007; Hendry, 2016; Pelletier et al., 2009).

1.2 Fungi are underrepresented in climate change research

Linking ecological modeling (correlative and mechanistic) with evolution is proving to be effective in assessing the vulnerability of several plants, birds, and mammals (Pacifici et al., 2015), but fungi are less studied (Willis, 2018). Fungi encompasses diverse ecosystems (Bahram & Netherway, 2021; Lofgren & Stajich, 2021; Peay et al., 2016) and interact with plants, insects, and mammals. A few species are also used by humans for food and beverage production (Legras et al., 2018; Willis, 2018) and in medicine (e.g., Penicillin) (Willis, 2018). The dark side of fungal interactions with other eukaryotes can result in infectious diseases such as Chytridiomycosis in amphibians, Aspergillosis in humans, and many plant blights (Willis, 2018). With ongoing global climate change, fungal diseases might emerge more often, threatening human health and food security (Case et al., 2022; Nnadi & Carter, 2021; Willis, 2018), yet fungi are understudied compared to animals and plants.

Furthermore, fungal microbe biogeography receives even less attention, historically attributed to the Baas Becking hypothesis that "everything is everywhere, but the environment selects." Now it is increasingly evident that fungal microbes can range from endemism to global distributions (Mozzachiodi et al., 2022; Peay et al., 2016; Stajich, 2017; Taylor et al., 2006), likely caused by geographical isolation and climate (Peay et al., 2016). Current challenges are resolving fungal species that have been misidentified that was solely based on morphology (Lofgren & Stajich, 2021) and large-scale databases that have incomplete or missing ecological information (Boekhout et al., 2022; Harrison et al., 2024; Mozzachiodi et al., 2022). In addition, some species are difficult to culture because of their tight interactions with other organisms (Willis, 2018).

We are at the tip of an iceberg in capturing the extent of fungal biodiversity. With ongoing climate change, we could lose this biodiversity before understanding it.

1.3 Wild budding yeast – models for studying how climate affects natural populations

Budding yeast are fungi that originated 317-523 million years ago (MYA) and are remarkably diverse; at least as genetically diverse as plant and animal lineages (Shen et al., 2018). Budding yeast (subphylum Saccharomycotina) are found in diverse environments (aquatic and terrestrial) and have specialized ecological niches (Boekhout et al., 2022; Buzzini et al., 2017; Dujon & Louis, 2017; Harrison et al., 2024; Mozzachiodi et al., 2022). A hallmark feature of the evolution of budding yeast is the gain and loss of metabolic traits, which not only shaped the evolution of Saccharomycotina (Opulente et al., 2024; Shen et al., 2018), but may explain why yeast can inhabit diverse habitats (Lachance, 2003). Furthermore, temperature preferences for growth likely has shape the biogeography of yeasts, especially for *Saccharomyces* species, including *Saccharomyces cerevisiae* (Gonçalves et al., 2011; Langdon et al., 2020; Leducq et al., 2014; Nespolo et al., 2020; Peris et al., 2023; Robinson et al., 2016; Salvadó et al., 2011a; Sampaio & Gonçalves, 2008; Spurley et al., 2022; Sylvester et al., 2015). With rising temperatures, changes or fluctuations are likely to affect the species distributions of wild yeast.

Saccharomyces cerevisiae has been an extraordinary model organism for eukaryotes, and its use has led to scientific discoveries in genetics and cellular biology. It is tractable for biotechnology and computational genomics because of its short generation time and small genome (12 megabase pairs) (Scannell et al., 2011). We are now in an

excellent position to extend wild *S. cerevisiae* and its non-domesticated sister species, *S. paradoxus* as models for microbial ecology (Cavalieri et al., 2022; Tsai, 2022).

Additionally, human-associated yeasts have also been isolated from temperate forests (Mozzachiodi et al., 2022), and their evolution could help us understand how climate change may affect microfungal populations in the wild. This is becoming more pressing since human pathogenic yeast, such as *Candida albicans*, *Nakaseomyces glabrata*, *Candida parapsilosis*, and *Candida tropicalis* can be isolated from wild habitats (Bensasson et al., 2019; Opulente et al., 2019; Robinson et al., 2016), and are listed as fungal priority pathogens by the World Health Organization (WHO) (*WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action*, 2022).

1.4 Dissertation outline and specific aims

For this dissertation, I studied natural populations of wild yeasts using a combined approach from ecology (field collections and correlative species distribution models) and evolution (phylogenomics and population genomics) to understand how climate and human migration have shaped wild yeast populations.

Chapter 2 illustrates the population genomic structure and phylogenomic relationships of wild *S. cerevisiae* from trees. I used over 300 genomes associated with trees to determine if tree-associated populations are genetically distinct from domesticated lineages and estimated the timing when forest lineages diverged. This included 31 new genome sequences of strains from Florida, Georgia, and Pennsylvania (see Chapter 3 for field and isolation methods). I discovered that Europe, Japan, and North America harbor numerous genetically distinct forest lineages with fine-scaled

population substructure. Divergence of these forest lineages occurred since the last glacial maximum and coincided with the spread of fermentation and agriculture. Lastly, I found a few strains from Europe resemble contemporary North American lineages, which may be attributed to how humans responded to the Great French Wine Blight. This demonstrates that forest populations are not entirely exempt from human activity. If humans affect yeast in natural environments, examining the evolution of other human-associated yeast species, especially human pathogenic species in wild environments, may be important.

In Chapter 3, I identified the ecological and climatic factors that predict the species ranges of *S. cerevisiae* from forests. I extended this analysis to other wild yeast species isolated alongside *S. cerevisiae*: *Saccharomyces paradoxus* and *Lachancea fermentati*. The Bensasson Lab and I have sampled environmental samples from angiosperms (tree bark, flowers, and fruit), gymnosperms (tree bark and exudate), marsh sediment, arthropods, and basidiomycetes fruiting bodies between 2019-2022 from 8 field sites along the eastern United States. This has led to the culturing and identification of over 450 yeast strains across the Saccharomycotina subphylum, with *S. cerevisiae*, *S. paradoxus*, *L. fermentati*, and *L. thermotolerans*, being the most common species isolated. We have isolated two commensal pathogenic yeast species, *Nakaseomyces glabrata* (*Candida glabrata*) and *Candida parapsilosis*, which are listed as high priority pathogens for research and public health action by the World Health Organization (*WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action, 2022*).

Additionally, we have isolated several industrial yeast species used for biotechnology: *Schizosaccharomyces japonicus*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, and *Wickerhamomyces anomalus*. We used generalized linear models to find that *S. cerevisiae*, *S. paradoxus*, and *L. fermentati* isolation frequencies are associated with temperature and precipitation, suggesting that these climate variables may limit their ranges. Our extensive field survey suggests that climate has affected wild yeast within forest environments and that their climate associations could help us predict how fungi will respond to climate change.

1.5 References

- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J.-L., Serra, M., Dequin, S., Couloux, A., Guy, J., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular Ecology*, 24(21), 5412–5427. <https://doi.org/10.1111/mec.13341>
- Alsammar, H., & Delneri, D. (2020). An update on the diversity, ecology and biogeography of the *Saccharomyces* genus. *FEMS Yeast Research*, 20(3). <https://doi.org/10.1093/femsyr/foaa013>
- Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858. <https://doi.org/10.1126/sciadv.aat4858>
- Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*, 93(7), 1527–1539. <https://doi.org/10.1890/11-1930.1>
- Arora, P., Singh, P., Wang, Y., Yadav, A., Pawar, K., Singh, A., Padmavati, G., Xu, J., & Chowdhary, A. (2021). Environmental Isolation of *Candida auris* from the Coastal Wetlands of Andaman Islands, India. *mBio*, 12(2), e03181-20. <https://doi.org/10.1128/mbio.03181-20>
- Bahram, M., & Netherway, T. (2021). Fungi as mediators linking organisms and ecosystems. *FEMS Microbiology Reviews*, 46(2), fuab058-. <https://doi.org/10.1093/femsre/fuab058>

- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of Natural Hybridization in Brazilian Wild Lineages of *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, 8(2), 317–329. <https://doi.org/10.1093/gbe/evv263>
- Bensasson, D., Dicks, J., Ludwig, J. M., Bond, C. J., Elliston, A., Roberts, I. N., & James, S. A. (2019). Diverse Lineages of *Candida albicans* Live on Old Oaks. *Genetics*, 211(1), 277–288. <https://doi.org/10.1534/genetics.118.301482>
- Boekhout, T., Amend, A. S., Baidouri, F. E., Gabaldón, T., Geml, J., Mittelbach, M., Robert, V., Tan, C. S., Turchetti, B., Vu, D., Wang, Q.-M., & Yurkov, A. (2022). Trends in yeast diversity discovery. *Fungal Diversity*, 114(1), 491–537. <https://doi.org/10.1007/s13225-021-00494-6>
- Boult, V. L., & Evans, L. C. (2021). Mechanisms matter: Predicting the ecological impacts of global change. *Global Change Biology*. <https://doi.org/10.1111/gcb.15527>
- Boynton, P. J., & Greig, D. (2014). The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast*, 31(12), 449–462. <https://doi.org/10.1002/yea.3040>
- Buckley, L. B., & Kingsolver, J. G. (2019). Environmental variability shapes evolution, plasticity and biogeographic responses to climate change. *Global Ecology and Biogeography*, 28(10), 1456–1468. <https://doi.org/10.1111/geb.12953>
- Buzzini, P., Lachance, M. A., & Yurkov, A. (Eds.). (2017). *Yeasts in Natural Ecosystems: Diversity*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-62683-3>
- Carroll, S. P., Hendry, A. P., Reznick, D. N., & Fox, C. W. (2007). Evolution on ecological time-scales. *Functional Ecology*, 21(3), 387–393. <https://doi.org/10.1111/j.1365-2435.2007.01289.x>
- Casadevall, A., Kontoyiannis, D. P., & Robert, V. (2019). On the Emergence of *Candida auris*: Climate Change, Azoles, Swamps, and Birds. *mBio*, 10(4), 10.1128/mbio.01397-19. <https://doi.org/10.1128/mbio.01397-19>
- Case, N. T., Berman, J., Blehert, D. S., Cramer, R. A., Cuomo, C., Currie, C. R., Ene, I. V., Fisher, M. C., Fritz-Laylin, L. K., Gerstein, A. C., Glass, N. L., Gow, N. A. R., Gurr, S. J., Hittinger, C. T., Hohl, T. M., Iliev, I. D., James, T. Y., Jin, H., Klein, B. S., ... Cowen, L. E. (2022). The future of fungi: Threats and opportunities. *G3*, 12. <https://doi.org/10.1093/g3journal/jkac224>

- Cavalieri, D., Valentini, B., & Stefanini, I. (2022). Going wild: Ecology and genomics are crucial to understand yeast evolution. *Current Opinion in Genetics & Development*, 75, 101922. <https://doi.org/10.1016/j.gde.2022.101922>
- Charron, G., Leducq, J.-B., Bertin, C., Dubé, A. K., & Landry, C. R. (2014). Exploring the northern limit of the distribution of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* in North America. *FEMS Yeast Research*, 14(2), 281–288. <https://doi.org/10.1111/1567-1364.12100>
- Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., Huang, A. A., Dudley, A. M., & Fay, J. C. (2013). Genomic Sequence Diversity and Population Structure of *Saccharomyces cerevisiae* Assessed by RAD-seq. *G3: Genes/Genomes/Genetics*, 3(12), 2163–2171. <https://doi.org/10.1534/g3.113.007492>
- 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. <https://doi.org/10.1371/journal.pone.0005317>
- Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., & Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nature Communications*, 9(1), 2690. <https://doi.org/10.1038/s41467-018-05106-7>
- Dujon, B. A., & Louis, E. J. (2017). Genome Diversity and Evolution in the Budding Yeasts (Saccharomycotina). *Genetics*, 206(2), 717–750. <https://doi.org/10.1534/genetics.116.199216>
- Fay, J. C., Liu, P., Ong, G. T., Dunham, M. J., Cromie, G. A., Jeffery, E. W., Ludlow, C. L., & Dudley, A. M. (2019). A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. *PLOS Biology*, 17(3), e3000147. <https://doi.org/10.1371/journal.pbio.3000147>
- Fussmann, G. F., Loreau, M., & Abrams, P. A. (2007). Eco-evolutionary dynamics of communities and ecosystems. *Functional Ecology*, 21(3), 465–477. <https://doi.org/10.1111/j.1365-2435.2007.01275.x>
- Gavin, D. G., Fitzpatrick, M. C., Gugger, P. F., Heath, K. D., Rodríguez-Sánchez, F., Dobrowski, S. Z., Hampe, A., Hu, F. S., Ashcroft, M. B., Bartlein, P. J., Blois, J. L., Carstens, B. C., Davis, E. B., Lafontaine, G. de, Edwards, M. E., Fernandez, M., Henne, P. D., Herring, E. M., Holden, Z. A., ... Williams, J. W. (2014). Climate refugia: Joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, 204(1), 37–54. <https://doi.org/10.1111/nph.12929>

- Gonçalves, P., Valério, E., Correia, C., Almeida, J. M. G. C. F. de, & Sampaio, J. P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric *Saccharomyces* Species. *PLoS ONE*, 6(6), e20739. <https://doi.org/10.1371/journal.pone.0020739>
- Günther, C. S., & Goddard, M. R. (2019). Do yeasts and *Drosophila* interact just by chance? *Fungal Ecology*, 38, 37–43. <https://doi.org/10.1016/j.funeco.2018.04.005>
- Han, D.-Y., Han, P.-J., Rumbold, K., Koricha, A. D., Duan, S.-F., Song, L., Shi, J.-Y., Li, K., Wang, Q.-M., & Bai, F.-Y. (2021). Adaptive Gene Content and Allele Distribution Variations in the Wild and Domesticated Populations of *Saccharomyces cerevisiae*. *Frontiers in Microbiology*, 12, 631250. <https://doi.org/10.3389/fmicb.2021.631250>
- Harrison, M., Opulente, D. A., Wolters, J. F., Shen, X., Zhou, X., Groenewald, M., Hittinger, C. T., Rokas, A., & LaBella, A. L. (2024). Exploring Saccharomycotina Yeast Ecology Through an Ecological Ontology Framework. *Yeast*. <https://doi.org/10.1002/yea.3981>
- Hartig, F., Lohse, L., & leite, M. de S. (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (Version 0.4.7) [Computer software]. <https://cran.r-project.org/web/packages/DHARMa/index.html>
- Hendry, A. P. (2016). *Eco-evolutionary Dynamics*. Princeton University Press.
- Hyma, K. E., & Fay, J. C. (2013). Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. *Molecular Ecology*, 22(11), 2917–2930. <https://doi.org/10.1111/mec.12155>
- Jessica Gurevitch; Samuel M. Scheiner; Gordon A. Fox. (2020). *The Ecology of Plants* (3rd ed.). Oxford University Press. <https://global.oup.com/ushe/product/the-ecology-of-plants-9781605358291>
- Johnson, L. J., Koufopanou, V., Goddard, M. R., Hetherington, R., Schäfer, S. M., & Burt, A. (2004). Population Genetics of the Wild Yeast *Saccharomyces paradoxus*. *Genetics*, 166(1), 43–52. <https://doi.org/10.1534/genetics.166.1.43>
- Kleinheslink, A. R., & Adler, P. B. (2018). The response of big sagebrush (*Artemisia tridentata*) to interannual climate variation changes across its range. *Ecology*, 99(5), 1139–1149. <https://doi.org/10.1002/ecy.2191>
- Kowallik, V., Miller, E., & Greig, D. (2015). The interaction of *Saccharomyces paradoxus* with its natural competitors on oak bark. *Molecular Ecology*, 24(7), 1596–1610. <https://doi.org/10.1111/mec.13120>

- Kuehne, H. A., Murphy, H. A., Francis, C. A., & Sniegowski, P. D. (2007). Allopatric Divergence, Secondary Contact, and Genetic Isolation in Wild Yeast Populations. *Current Biology*, 17(5), 407–411. <https://doi.org/10.1016/j.cub.2006.12.047>
- Kurtzman, C. P., & Suzuki, M. (2010). Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera *Babjeviella*, *Meyerozyma*, *Millerozyma*, *Priceomyces*, and *Scheffersomyces*. *Mycoscience*, 51(1), 2–14. <https://doi.org/10.1007/S10267-009-0011-5>
- Lachance, M.-A. (2003). The Phaff school of yeast ecology. *International Microbiology*, 6(3), 163–167. <https://doi.org/10.1007/s10123-003-0129-9>
- Lafontaine, G. de, Napier, J. D., Petit, R. J., & Hu, F. S. (2018). Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology*, 99(7), 1530–1546. <https://doi.org/10.1002/ecy.2382>
- Langdon, Q. K., Peris, D., Eizaguirre, J. I., Opulente, D. A., Buh, K. V., Sylvester, K., Jarzyna, M., Rodríguez, M. E., Lopes, C. A., Libkind, D., & Hittinger, C. T. (2020). Postglacial migration shaped the genomic diversity and global distribution of the wild ancestor of lager-brewing hybrids. *PLoS Genetics*, 16(4), e1008680. <https://doi.org/10.1371/journal.pgen.1008680>
- Leducq, J.-B., Charron, G., Samani, P., Dubé, A. K., Sylvester, K., James, B., Almeida, P., Sampaio, J. P., Hittinger, C. T., Bell, G., & Landry, C. R. (2014). Local climatic adaptation in a widespread microorganism. *Proceedings of the Royal Society B: Biological Sciences*, 281(1777), 20132472. <https://doi.org/10.1098/rspb.2013.2472>
- Lee, K. M., & Coop, G. (2019). Population genomics perspectives on convergent adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1777), 20180236. <https://doi.org/10.1098/rstb.2018.0236>
- Lee, T. J., Liu, Y.-C., Liu, W.-A., Lin, Y.-F., Lee, H.-H., Ke, H.-M., Huang, J.-P., Lu, M.-Y. J., Hsieh, C.-L., Chung, K.-F., Liti, G., & Tsai, I. J. (2022). Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages. *Genome Research*, 32(5), 864–877. <https://doi.org/10.1101/gr.276286.121>
- Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldon, T., Schuller, D., Sampaio, J. P., & Dequin, S. (2018). Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. *Molecular Biology and Evolution*, 35(7), 1712–1727. <https://doi.org/10.1093/molbev/msy066>

Lofgren, L. A., & Stajich, J. E. (2021). Fungal biodiversity and conservation mycology in light of new technology, big data, and changing attitudes. *Current Biology*, 31(19), R1312–R1325. <https://doi.org/10.1016/j.cub.2021.06.083>

McGaughan, A., Laver, R., & Fraser, C. (2021). Evolutionary Responses to Warming. *Trends in Ecology & Evolution*. <https://doi.org/10.1016/j.tree.2021.02.014>

Micheal J Crawley. (2014). *Statistics: An Introduction Using R, 2nd Edition* (Second). John Wiley & Sons.

Moritz, C., & Agudo, R. (2013). The Future of Species Under Climate Change: Resilience or Decline? *Science*, 341(6145), 504–508. <https://doi.org/10.1126/science.1237190>

Mozzachiodi, S., Bai, F., Baldrian, P., Bell, G., Boundy-Mills, K., Buzzini, P., Čadež, N., Cubillos, F. A., Dashko, S., Dimitrov, R., Fisher, K. J., Gibson, B., Gouliamova, D., Greig, D., Heistinger, L., Hittinger, C. T., Jecmenica, M., Koufopanou, V., Landry, C. R., ... Boynton, P. (2022). Yeasts from temperate forests. *Yeast*. <https://doi.org/10.1002/yea.3699>

Naumov, G. I., Naumova, E. S., & Sniegowski, P. D. (1997). Differentiation of European and Far East Asian populations of *Saccharomyces paradoxus* by allozyme analysis. *International Journal of Systematic Bacteriology*, 47(2), 341–344. <https://doi.org/10.1099/00207713-47-2-341>

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.

Nespolo, R. F., Villarroel, C. A., Oporto, C. I., Tapia, S. M., Vega-Macaya, F., Urbina, K., Chiara, M. D., Mozzachiodi, S., Mikhalev, E., Thompson, D., Larrondo, L. F., Saenz-Agudelo, P., Liti, G., & Cubillos, F. A. (2020). An Out-of-Patagonia migration explains the worldwide diversity and distribution of *Saccharomyces eubayanus* lineages. *PLOS Genetics*, 16(5), e1008777. <https://doi.org/10.1371/journal.pgen.1008777>

Nnadi, N. E., & Carter, D. A. (2021). Climate change and the emergence of fungal pathogens. *PLoS Pathogens*, 17(4), e1009503. <https://doi.org/10.1371/journal.ppat.1009503>

Opulente, D. A., LaBella, A. L., Harrison, M.-C., Wolters, J. F., Liu, C., Li, Y., Kominek, J., Steenwyk, J. L., Stoneman, H. R., VanDenAvond, J., Miller, C. R., Langdon, Q. K., Silva, M., Gonçalves, C., Ubbelohde, E. J., Li, Y., Buh, K. V., Jarzyna, M., Haase, M. A. B., ... Hittinger, C. T. (2024). Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts. *Science*, 384(6694), eadj4503. <https://doi.org/10.1126/science.adj4503>

- Opulente, D. A., Langdon, Q. K., Buh, K. V., Haase, M. A. B., Sylvester, K., Moriarty, R. V., Jarzyna, M., Considine, S. L., Schneider, R. M., & Hittinger, C. T. (2019). Pathogenic budding yeasts isolated outside of clinical settings. *FEMS Yeast Research*, 19(3). <https://doi.org/10.1093/femsyr/foz032>
- Osland, M. J., Stevens, P. W., Lamont, M. M., Brusca, R. C., Hart, K. M., Waddle, J. H., Langtimm, C. A., Williams, C. M., Keim, B. D., Terando, A. J., Reyier, E. A., Marshall, K. E., Loik, M. E., Boucek, R. E., Lewis, A. B., & Seminoff, J. A. (2021). Tropicalization of temperate ecosystems in North America: The northward range expansion of tropical organisms in response to warming winter temperatures. *Global Change Biology*. <https://doi.org/10.1111/gcb.15563>
- Pacifci, M., Foden, W. B., Visconti, P., Watson, J. E. M., Butchart, S. H. M., Kovacs, K. M., Scheffers, B. R., Hole, D. G., Martin, T. G., Akçakaya, H. R., Corlett, R. T., Huntley, B., Bickford, D., Carr, J. A., Hoffmann, A. A., Midgley, G. F., Pearce-Kelly, P., Pearson, R. G., Williams, S. E., ... Rondinini, C. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, 5(3), Article 3. <https://doi.org/10.1038/nclimate2448>
- Peay, K. G., Kennedy, P. G., & Talbot, J. M. (2016). Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology*, 14(7), Article 7. <https://doi.org/10.1038/nrmicro.2016.59>
- Pelletier, F., Garant, D., & Hendry, A. P. (2009). Eco-evolutionary dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1523), 1483–1489. <https://doi.org/10.1098/rstb.2009.0027>
- Peña, J. J., Scopel, E. F. C., Ward, A. K., & Bensasson, D. (2025). Footprints of Human Migration in the Population Structure of Wild Baker's Yeast. *Molecular Ecology*, e17669(n/a), e17669. <https://doi.org/10.1111/mec.17669>
- Peris, D., Ubbelohde, E. J., Kuang, M. C., Kominek, J., Langdon, Q. K., Adams, M., Koshalek, J. A., Hulfachor, A. B., Opulente, D. A., Hall, D. J., Hyma, K., Fay, J. C., Leducq, J.-B., Charron, G., Landry, C. R., Libkind, D., Gonçalves, C., Gonçalves, P., Sampaio, J. P., ... Hittinger, C. T. (2023). Macroevolutionary diversity of traits and genomes in the model yeast genus *Saccharomyces*. *Nature Communications*, 14(1), 690. <https://doi.org/10.1038/s41467-023-36139-2>
- Peter, J., Chiara, M. D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), 339. <https://doi.org/10.1038/s41586-018-0030-5>
- Pontes, A., Hutzler, M., Brito, P. H., & Sampaio, J. P. (2020). Revisiting the Taxonomic Synonyms and Populations of *Saccharomyces cerevisiae*—Phylogeny,

Phenotypes, Ecology and Domestication. *Microorganisms*, 8(6), Article 6.
<https://doi.org/10.3390/microorganisms8060903>

Robinson, H. A., Pinharanda, A., & Bensasson, D. (2016). Summer temperature can predict the distribution of wild yeast populations. *Ecology and Evolution*, 6(4), 1236–1250. <https://doi.org/10.1002/ece3.1919>

Rodríguez-Verdugo, A., Buckley, J., & Stapley, J. (2017). The genomic basis of eco-evolutionary dynamics. *Molecular Ecology*. <https://doi.org/10.1111/mec.14045>

Sage, R. F. (2019). Global change biology: A primer. *Global Change Biology*, 26(1), 3–30. <https://doi.org/10.1111/gcb.14893>

Salvadó, Z., Arroyo-López, F. N., Guillamón, J. M., Salazar, G., Querol, A., & Barrio, E. (2011b). Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Applied and Environmental Microbiology*, 77(7), 2292–2302. <https://doi.org/10.1128/aem.01861-10>

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. <https://doi.org/10.1128/AEM.02396-07>

Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics*, 40, 415–436.

Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Čadež, N., Libkind, D., ... Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, 175(6), 1533–1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>

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. <https://doi.org/10.1111/j.1567-1364.2002.tb00048.x>

Song, G., Dickins, B. J. A., Demeter, J., Engel, S., Dunn, B., & Cherry, J. M. (2015). AGAPE (Automated Genome Analysis PipelinE) for Pan-Genome Analysis of *Saccharomyces cerevisiae*. *PLOS ONE*, 10(3), e0120671. <https://doi.org/10.1371/journal.pone.0120671>

- Spurley, W. J., Fisher, K. J., Langdon, Q. K., Buh, K. V., Jarzyna, M., Haase, M. A. B., Sylvester, K., Moriarty, R. V., Rodriguez, D., Sheddan, A., Wright, S., Sorlie, L., Hulfachor, A. B., Opulente, D. A., & Hittinger, C. T. (2022). Substrate, temperature, and geographical patterns among nearly 2000 natural yeast isolates. *Yeast*, 39(1–2), 55–68. <https://doi.org/10.1002/yea.3679>
- Stajich, J. E. (2017). Fungal Genomes and Insights into the Evolution of the Kingdom. *Microbiology Spectrum*, 5(4). <https://doi.org/10.1128/microbiolspec.funk-0055-2016>
- 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. [https://doi.org/10.1016/S1567-1356\(03\)00171-5](https://doi.org/10.1016/S1567-1356(03)00171-5)
- Sylvester, K., Wang, Q.-M., James, B., Mendez, R., Hulfachor, A. B., & Hittinger, C. T. (2015). Temperature and host preferences drive the diversification of *Saccharomyces* and other yeasts: A survey and the discovery of eight new yeast species. *FEMS Yeast Research*, 15(3). <https://doi.org/10.1093/femsyr/fov002>
- Taylor, J. W., Turner, E., Townsend, J. P., Dettman, J. R., & Jacobson, D. (2006). Eukaryotic microbes, species recognition and the geographic limits of species: Examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1475), 1947–1963. <https://doi.org/10.1098/rstb.2006.1923>
- Tilakaratna, V., & Bensasson, D. (2017). Habitat Predicts Levels of Genetic Admixture in *Saccharomyces cerevisiae*. *G3: Genes, Genomes, Genetics*, 7(9), 2919–2929. <https://doi.org/10.1534/g3.117.041806>
- Tsai, I. J. (2022). The teenage years of yeast population genomics trace history, admixing and getting wilder. *Current Opinion in Genetics & Development*, 75, 101942. <https://doi.org/10.1016/j.gde.2022.101942>
- Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A., & Bai, F.-Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular Ecology*, 21(22), 5404–5417. <https://doi.org/10.1111/j.1365-294x.2012.05732.x>
- WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action* (1st ed). (2022). World Health Organization.
- Willis, K. J. (2018). *State of the World's Fungi*. <https://stateoftheworldsfungi.org/>
- Zhang, H., Skelton, A., Gardner, R. C., & Goddard, M. R. (2010). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand:

Evidence for migration from Europe and interspecies hybrids. *FEMS Yeast Research*, 10(7), 941–947. <https://doi.org/10.1111/j.1567-1364.2010.00681.x>

CHAPTER 2

FOOTPRINTS OF HUMAN MIGRATION IN THE POPULATION STRUCTURE OF
WILD BAKER'S YEAST¹

¹Peña, J. J., Scopel, E. F., Ward, A. K., & Bensasson, D. *Molecular Ecology*. e17669
Reprinted here with permission of the publisher.

2.1 Abstract

Humans have a long history of fermenting food and beverages that led to domestication of the baker's yeast, *Saccharomyces cerevisiae*. Despite their tight companionship with humans, yeast species that are domesticated or pathogenic can also live on trees. Here we used over 300 genomes of *S. cerevisiae* from oaks and other trees to determine whether tree-associated populations are genetically distinct from domesticated lineages and estimate the timing of forest lineage divergence. We found populations on trees are highly structured within Europe, Japan, and North America. Approximate estimates of when forest lineages diverged out of Asia and into North America and Europe coincide with the end of the last ice age, the spread of agriculture, and the onset of fermentation by humans. It appears that migration from human-associated environments to trees is ongoing. Indeed, patterns of ancestry in the genomes of three recent migrants from the trees of North America to Europe could be explained by the human response to the Great French Wine Blight. Our results suggest that human-assisted migration affects forest populations, albeit rarely. Such migration events may even have shaped the global distribution of *S. cerevisiae*. Given the potential for lasting impacts due to yeast migration between human and natural environments, it seems important to understand the evolution of human commensals and pathogens in wild niches.

Keywords: phylogeography, genetic admixture, ecological genomics, yeast ecology, wine yeast, *Saccharomyces cerevisiae*

2.2 Introduction

Since the last ice age, humans have transitioned from a hunter-gatherer to a sedentary lifestyle and developed new technologies for preserving food including fermentation (McGovern, 2003). The baker's yeast, *Saccharomyces cerevisiae*, is a driver of such fermentations and is used to produce beer, wine, sake, cocoa, and coffee (Marsit et al., 2017). The earliest archaeological evidence of fermented rice, honey, and fruit in China dates to 7,000 BCE (McGovern et al., 2004). Fermented beer was first discovered in ancient Sumerian vessels from 6,000 BCE (Michel et al., 1992) and there is evidence for wine production between 6,000 and 4,000 BCE in Iran, the Caucasus, and Mesopotamia (Pretorius, 2000; McGovern, 2003). Wine production then spread throughout the Mediterranean and was prevalent across Europe and Northern Africa by 500 BCE (Pretorius, 2000).

Today, the population genetics of *S. cerevisiae* shows imprints of domestication with several genetic lineages associated with distinct domestication events (Almeida et al., 2015; Duan et al., 2018; Fay et al., 2019; Fay & Benavides, 2005; Gallone et al., 2016; Gayevskiy et al., 2016; Legras et al., 2007; Legras et al., 2018; Liti et al., 2009; Peter et al., 2018; Schacherer et al., 2009). Deep sampling for wild strains from Chinese and Taiwanese forests revealed high levels of lineage diversity compared to all other lineages, and the current consensus is that East Asian forests likely harbored the ancestral source populations that gave rise to all global *S. cerevisiae* lineages (Duan et al., 2018; Lee et al., 2022; Wang et al., 2012). Even outside Asia, sampling of *S. cerevisiae* from natural environments shows a wild side to this human-associated yeast species; there are

wild genetic lineages that are distinct from known domesticated lineages (Almeida et al., 2015; Cromie et al., 2013; Fay & Benavides, 2005; Han et al., 2021; Liti et al., 2009; Peter et al., 2018; Tilakaratna & Bensasson, 2017).

How much has human activity affected the ecology and evolution of wild populations of human-associated yeast species? Human pathogenic yeast such as *Candida albicans*, *Nakaseomyces glabrata*, *Candida parapsilosis*, and *Candida tropicalis* can be isolated from trees and other plant habitats (Bensasson et al., 2019; Opulente et al., 2019; Robinson et al., 2016) and other forest yeast species are associated with humans (Boynton & Greig, 2014; Mozzachiodi *et al.* 2022). Here, we make use of the extensive genome data available for *S. cerevisiae* and focus on the population structure and phylogenetic relationships of wild *S. cerevisiae* from a single ecological niche. By studying strains from oaks and other trees, we characterize populations in an ancestral niche while avoiding the complications of genetic admixture more commonly seen in *S. cerevisiae* from fruit, flowers, and insects (Hyma & Fay, 2013; Tilakaratna & Bensasson, 2017). Specifically, we identified tree-associated populations and estimated the timing of wild yeast migration events. We show that (i) tree-associated *S. cerevisiae* populations are highly structured, (ii) the worldwide spread of forest populations out of Asia probably occurred since the last glacial maximum, and lastly (iii) human-assisted migration is ongoing and may include migration from the USA to Europe since the Great French Wine Blight.

2.3 Materials and Methods

2.3.1 Yeast strains and genome data

Whole-genome sequences for strains sampled from trees were compiled from publicly available data ($N = 295$; Table S1) (Almeida et al., 2015; Barbosa et al., 2016; Bergström et al., 2014; Duan et al., 2018; Fay et al., 2019; Gayevskiy et al., 2016; Han et al., 2021; Pontes et al., 2020; Skelly et al., 2013; Song et al., 2015; Strope et al., 2015; Yue et al., 2017). We defined *S. cerevisiae* tree-sampled strains as those isolated from tree bark, exudate and leaves from trees or litter, and we also included strains from any soil. Metadata was compiled for each genome sequence to include geographical origin, ecological substrate, and previously reported genetic clade associated with the strain (Table S1). New whole-genome sequence data was generated for strains from trees in Indiana and Kentucky ($N = 7$; Osburn et al., 2018), North Carolina ($N = 9$; Diezmann & Dietrich 2009), Europe ($N = 3$; Robinson et al., 2016) and for new *S. cerevisiae* strains from the bark of white oak (*Quercus alba*) and live oak (*Q. virginiana*) from Georgia, Florida, Pennsylvania, and North Carolina ($N = 15$; Bensasson lab). DNA was extracted from single yeast colonies using Promega Wizard® Genomic DNA purification kit following the manufacturer's protocol for yeast except that only 75 units of lyticase (Sigma) were used in an overnight incubation at 37°C. For the generation of genome data from 22 strains from the Bensasson and Osburn labs, paired-end Illumina libraries were generated by the Georgia Genomics and Bioinformatics Core using the purePlex DNA Library Preparation Kit (GGBC Project #5256) or the Nextera DNA-Seq Library Protocol (GGBC Project #5881). Paired end sequencing was performed on the Illumina

NextSeq2000 platform (2×150 bp). The remaining strains were sequenced at the University of Manchester as described in Almeida et al. (2015). Genome data is available on NCBI-SRA under project number PRJNA1090965 and includes a further 8 strains from Pennsylvania that are monosporic derivatives of previously studied strains (Table S1; Sniegowski et al., 2002).

To examine how strains from trees are related to other yeast strains, we constructed a reference panel of strains to represent published clades (1,030 strains from 42 clades; Duan et al., 2018; Peter et al., 2018). These reference-panel strains were isolated from the human body (clinical), fermentation (e.g., wine and beer), baking, bioethanol, crops (e.g., sugar cane), decaying wood, fruit, flowers, insects, mushrooms, and water (e.g., sewers and oceans).

2.3.2 Read mapping and base calling

Paired-end and single-end genomic Illumina reads were downloaded from the European Bioinformatics Institute (<https://www.ebi.ac.uk/>) or generated in this study. Reads were mapped to the *S. cerevisiae* reference genome, S288c (SacCer_Apr2011/sacCer3 from UCSC), using Burrows-Wheeler Aligner (bwa-mem, version 0.7.17; Li & Durbin, 2009). We used SAMtools to sort, index, and compress bam files and generated a consensus sequence using the mpileup function with the -I option to exclude indels (version 1.6; Li et al., 2009). Next, we used the BCFtools call function with the -c option to generate a consensus sequence (version 1.9) (Li et al., 2009) and converted from vcf format to fastq format in SAMtools using the vcftools.pl vcf2fq command. Lastly, base calls with a phred-scaled quality score of less than 40 were treated

as missing data (calls were converted to “N”) using seqtk seq -q 40 in SAMtools. In practice, this probably leads to an overall error rate lower than 1 in 10,000 because high-depth Illumina sequence usually yields consensus base calls that are not close to the Q40 cutoff. For example, after applying a Q40 cutoff to our past Illumina sequencing of *Candida albicans* we observed an error rate below 1 in 100,000 (< 169 errors in 14 Mbp; Bensasson et al., 2019) and use of the Q40 threshold on Sanger sequence from *S. cerevisiae* yielded a similarly low error rate (< 1 in 180,000 bp; Bensasson 2011). None of the analyses discussed here would be affected by such a low error rate.

2.3.3 Quality filtering steps

For population structure analysis, we removed genome sequences if they were from tree-sampled strains already represented in the dataset ($N = 27$), from strains with no geographical information ($N = 1$), or if they had average read depth below $30\times$ ($N = 18$). Additionally, genome data were visualized to check for intraspecies cross-contaminated using vcf2allelePlot.pl (Bensasson et al., 2019; Scopel et al., 2021): a genome was removed ($N = 18$) if some reads produced unexpected SNP calls at a frequency over 1%. For the remaining strains sampled from trees, we estimated genome-wide levels of heterozygosity using vcf2allelePlot.pl (Bensasson et al., 2019) which estimates the number of heterozygous point substitutions divided by the length of the high-quality genome sequence (phred-scaled quality score over 40). Most strains from trees are homozygous (Figure S1). We removed 35 tree strains that are heterozygous (heterozygosity over 0.001) because they are difficult to represent in downstream

phylogenetic analyses and may be interclade hybrids. After filtering, 236 tree-sampled strains remained for population genetic analyses (Table S1 and Figure S2).

For the reference panel strains, we applied the same quality filtering steps and examined levels of genome-wide heterozygosity, removing 16 out of 42 published clades because all individuals were heterozygous (Figure S1 and S2).

2.3.4 Population structure and genetic admixture

Whole-genome alignments were generated by concatenating the alignments for all 16 chromosomes into a single multiple-alignment file. Strains from trees were compared to reference strains after random selection of three strains per clade from the 26 published clades that remained after quality filtering (Table S2). One strain (BJ6) was randomly assigned to both CHN-IV and Far East Asia clades ($N = 77$ strains). Ambiguity codes or lowercase base calls were converted to N's, and ends were filled to align to the same length. A neighbor-joining tree from genetic distances estimated by pairwise comparison of all genome sequences was constructed using MEGA-CC (version 10.0.5; Kumar et al., 2012). We used a Tamura-Nei substitution model (Tamura & Nei, 1993) with a gamma distribution and 100 bootstrap replicates. Gaps or missing data were discarded from each pairwise sequence comparison. For visualization, the neighbor-joining tree was rotated using ape (version 5.6.2; Paradis & Schliep, 2018) and further visualized using ggtree (version 3.4.4; Xu et al., 2021; Yu et al., 2017, 2018).

Population structure and individual ancestry were estimated from SNP allele frequencies using ADMIXTURE (version 1.3.0; Alexander et al., 2009). Genome data for all strains was merged into a single alignment in variant call format (vcf) using BCFtools

and mitochondrial DNA was removed. Non-variant sites were filtered out using the min-ac 1 function in BCFtools, which retains variants with at least one non-reference allele. Low-quality reads with a Phred-scaled quality score under 40 were removed using the minQ option in VCFtools (version 0.1.16; Danecek et al., 2011). Then we converted the single alignment vcf file to text formatted and binary files using PLINK (version 1.9; Purcell et al., 2007) for downstream analysis with ADMIXTURE. We assigned strains to distinct populations or genetic clusters (K) through repeated runs of ADMIXTURE. Runs assumed different numbers of genetic clusters from 4 to 40 with five replicates per K. We selected the run with the highest log likelihood value for each K and visualized population structure across different K's (Figure S3). We used CLUMPAK, specifically 'Distruct', to align ancestry proportions (Q matrices) across different values of K (Kopelman et al., 2015). ADMIXTURE results were visualized as stacked bar plots using the pophelper R package (version 3.2.1; Francis, 2017; Figure S4). Distinct genetic clusters were verified if they showed monophyletic clades with at least 95% bootstrap support in a neighbor-joining tree (Figure 1A and Table S2). We selected the run with the highest number of verified clusters or clades based on monophyletic groups in the neighbor-joining phylogeny and whether strains grouped by geography.

Phylogenomic relationships among strains were further examined using a maximum likelihood tree after excluding strains that showed possible recent genetic admixture when K = 30 (Figure S5 and Table S2). Admixed strains were defined as individuals whose percent ancestry from a single population is less than 90% in ADMIXTURE results for tree-sampled and reference panel strains (Table S2). It is possible however that some of the mixed ancestry invoked by the ADMIXTURE

software could reflect ancestral polymorphisms or poorly sampled population subdivision. We included only three reference strains from CHN-IV and Far East Asia, which appear to be the same clade (Figure 1). We used a genome-wide alignment of SNPs mapped to the S228c reference genome to construct a phylogenomic tree with IQtree, ultrafast bootstrapping (version 1.6.12) (Minh et al., 2013; Nguyen et al., 2015), and a general time reversible model with a gamma distribution to estimate site heterogeneity. The maximum likelihood tree was visualized using ape and ggtree in R (version 4.2.2).

2.3.5 Population substructure within Europe

Many *S. cerevisiae* strains have good quality genome data from European trees ($N = 51$ strains; Table S3) and have not been tested for population substructure. We used the same methods to analyze population substructure among European tree-sampled strains. ADMIXTURE was run by varying K from 2 to 8, and each K was repeated five times (Figure S6 and S7). Then, we constructed a phylogenomic tree using maximum likelihood estimation to infer phylogenetic relationships after removing one strain showing mixed ancestry at $K = 4$ (ZP541).

2.3.6 *In silico* chromosome painting

To identify genomic segments that could show gene flow between populations, we used a chromosome painting approach with faChrompaint.pl and a 30 kb window size (Bensasson et al., 2019). This *in silico* chromosome painting approach compares non-overlapping sliding windows of sequence to a panel of predefined clades. We identified 25 genetically distinct clades from population structure and phylogenomic analyses

(Figure S4 and S5) and randomly selected three strains per clade to use as a “backbone” panel ($N = 75$ backbone strains; Figure S8 and Table S4). For a strain of interest, each 30 kb window was compared to a multiple sequence alignment of backbone strains, then “painted” a color representing the clade of the most similar backbone sequence. Genomic regions that were diverged from all other backbone sequences (proportion of differing sites over 0.003) were painted white. This divergence threshold was decided based on within-clade pairwise comparisons for North American and European clades: over 90% of 30 kb windows showed a proportion of differing sites below 0.003 (< 0.3%), and most between-clade pairwise comparisons showed divergence over 0.003 (Figure S9 and S10). Exclusion of the most diverged regions reduces the chance that genetic similarity between lineages could be the result of incomplete lineage sorting. The genomic regions with low similarity to any other strain are the ones most likely to contain ancestral polymorphism that have not yet have fixed. There were few such regions (Figure S8).

2.3.7 Time divergence analysis

Time divergence analyses was performed on a single non-admixed locus (30-60 kb) from each backbone strain per chromosome (Figure S11). *In silico* chromosome painting analyses of backbone strains confirmed whether each backbone strain matched its primary clade assignment from allele frequency analyses using ADMIXTURE (Figure S8). For downstream time divergence analyses, we removed three backbone strains with less than 50% primary clade assignment using a chromosome painting approach, five backbone strains with over 10% secondary clade assignment and only included a single strain from the outgroup CHN-IX/Taiwanese clade (Table S5).

To estimate the timing of divergence events in the absence of purifying selection, we considered only nucleotide sites at which synonymous changes could occur. We extracted and concatenated 435 genes, with 11-39 genes for each 30-60 kb locus, after excluding 29 genes with introns, 42 genes that overlapped with other genes, and 4 genes with low quality sequence for at least 50% of the alignment. Using MEGA-CC, we extracted 4-fold and 2-fold degenerate sites for each locus and concatenated them into 3,507 to 19,016 bp alignments with 469 - 2,245 variable and 211 - 1,362 informative sites each. Numbers of informative sites were estimated using SeaView (Gouy et al., 2010).

A neighbor-joining tree was constructed for each locus on each chromosome using the same methods previously mentioned and the phylogeny was rooted using the CHN-IX/Taiwanese strain (EN14S01) as the outgroup. Neighbor-joining trees and multiple sequence alignments were used to estimate time trees per chromosome using MEGA-CC with the RelTime-ML option (Tamura et al., 2018) using a Tamura-Nei substitution model (Tamura & Nei, 1993) with the default setting to consider all sites for branch length calculations (Figure S12).

Using neighbor-joining distance trees and estimates of the *S. cerevisiae* mutation rate, we estimated the approximate timing at which modern tree-associated lineages (i) migrated out of Asia; (ii) into North America; and (iii) the separation of Wine/European and European oak lineages from trees. We used genetic distance to estimate the time (T) to the most recent common ancestor (MRCA) in generations per year: $T_{\text{MRCA}} = k / \mu / \text{generations per year}$; where k is the genetic distance to the MRCA of strains in a clade and μ is the point mutation rate per bp. We used the mutation rate of 1.67×10^{-10} point

substitutions per site per generation, which was estimated from hundreds of point mutations after genome sequencing of mutation accumulation lines of diploid *S. cerevisiae* (Zhu et al., 2014). Under controlled laboratory settings at 30°C, wild diploid *S. cerevisiae* and its sister species *Saccharomyces paradoxus* have an average generation (doubling) time of 65 minutes in the presence of glucose and 125 minutes on nutrient-poor growth media (Kaya et al., 2021). Although glucose, fructose and sucrose are present in the bark of trees that harbor yeast (Sampaio & Gonçalves, 2008) it is likely less available than in the lab. In regions where *S. cerevisiae* were easily sampled from trees, historic temperatures were usually below 30°C (Table S8). With these considerations in mind, we made a rough estimate of the number of generations per year for yeast on trees assuming: (i) a 90 minute generation time because while the tree niche is likely less nutrient rich than laboratory growth media, some sugars are available; (ii) 12 hours growth per day to account for no growth at lower night time temperatures; (iii) and no growth for six months of the year to account for low temperatures in winter. The resulting estimate is an average of 4 generations per day or 1,460 generations per year. This is much lower than the number of generations possible at 30°C in nutrient rich media in laboratory conditions: 22 per day; 8,086 generations per year. It is also lower than the estimate of 2,920 generations per year used to estimate the age of the wine-associated lineage (Fay & Benavides, 2005), which may be less affected by cold winters and nutrient poor conditions.

2.4 Results

2.4.1 Tree habitats harbor numerous genetically distinct lineages

Using genome data for 236 strains from oaks and other trees, we examined population structure among wild *S. cerevisiae* in this niche. Phylogenetic analyses of tree-sampled strains and a reference panel of 77 strains from published clades (Duan et al., 2018; Peter et al., 2018) revealed several genetically distinct lineages that only occur on trees from China, Europe, Japan, North America, Russia, and Taiwan (Figure 1, Table S2 and Figure S5). These include previously studied wild lineages such as 'CHN-IX', 'CHN-II,' and 'European oak' (Duan et al., 2018; Peter et al., 2018), and more tree-associated lineages (see below).

Before applying filters to an initial sample of 328 strains isolated from trees with good genome data ($> 30\times$ read depth), we also observed numerous strains from clades that are usually associated with humans (Table S1). For example, there were 29 strains from the 'Wine/European' lineage from 21 field sites in 4 continents; 24 from the 'Asian fermentation' lineage from 19 sites in 5 continents, 12 'South African Beer' clade strains from 3 South African field sites (Han et al., 2021); 6 strains from the 'Mixed Origin' clade associated with baking and clinical strains (Peter et al., 2018); and occasional strains from 'African honey wine', 'French dairy' and 'African palm wine' (Table S1). Some of these were too heterozygous for further study (Table S1, Figure S2).

2.4.2 Population substructure in the forest niches of North America and Japan

Phylogenetic analyses further revealed population substructure within North America and Japan, where each region has multiple genetically distinct populations (Figure 1A). Analysis of allele frequencies using ADMIXTURE confirmed that there are several genetically distinct populations in North America and Japan (Figure 1B). More specifically, there are at least four tree-associated (wild) American *S. cerevisiae* lineages in the eastern United States (Figure 1C) that are well-supported across phylogenetic and ADMIXTURE analyses (Figure 1, S4, and S5). Most wild strains from Pennsylvania are from a previously described lineage (Liti et al., 2009), and we refer to it here as 'North American A' (Table S1 and Table S2). There are two more North American wild lineages: 'North American B' and 'North American C', covering a broader geographical region than the North American A lineage (Figure 1C). North American B strains are from Georgia, North Carolina, a singleton strain from Michigan, and a singleton strain from Pennsylvania (Table S1 and Table S2). The North American C lineage occurs in strains from the southeastern United States (Kentucky, Georgia, and Florida). Lastly, wild strains from Ecuador and Brazil cluster with North American strains from Pennsylvania and New Jersey. We are coining this lineage as 'American wild' to reflect its Pan-American geography (Figure 1 and Figure S5) though some of these strains from trees were previously assigned to an 'Ecuadorian' clade (Peter et al., 2018) or a 'Brazil 1' clade (Barbosa et al., 2016).

Some tree strains have mixed genetic ancestry from deeply diverged American lineages (within the red box in Figure 1A and B), which is expected given their

overlapping geographic distributions (Figure 1C). It is perhaps more surprising that these lineages have remained distinct despite gene flow. The strains we and others have sampled are not from primary forests, which have almost vanished from North America (Potapov et al., 2017). A possible explanation is therefore that these forest lineages came into contact only recently because they are newly arrived in their current locations.

In Japan, there are at least two wild *S. cerevisiae* populations: one from Hiruzen Highland, 'Japan A', and one from Chiba Prefecture, 'Japan B' (Figure 1 and Figure S5). Japan A is diverged from all North American lineages and Japan B (Figure 1 and S5). Japan B is most like the North American A lineage (Figure 1 and S5).

2.4.3. Fine-scale population structure of wild *S. cerevisiae* from Europe

European yeast from trees form two genetically distinct lineages: 'Wine/European' and 'European oak' (Figure 1 and Figure S5) that were previously known (Almeida et al., 2015; Tilakaratna & Bensasson 2017). Although the Wine/European lineage is usually recovered from wine fermentations, it sometimes occurs on trees, especially in vineyards (Gayevskiy et al., 2016; Hyma and Fay, 2013; Robinson et al., 2016). Initial analyses suggest population substructure within the European oak lineage (Figure 1, Figure S4). To better describe this substructure, we therefore ran separate ADMIXTURE and phylogenomic analyses for all strains that were isolated from European trees (N = 51; Table S3 and Figure 2). These analyses showed population substructure within the European oak lineage that correlates with geography (Figure 2, Figure S6 and S7). Specifically, there is evidence for five sub-lineages from: (i) Portugal and Spain, which we are coining 'Iberian oak' with support from both phylogenetic and ADMIXTURE

analyses (Figure 2); and phylogenetic analyses suggest other distinct populations in (ii) Italy, (iii) Montenegro, (iv) Greece and Hungary, and (v) the North Caucasus (Figure 2). Additionally, several oak trees harbor strains from the 'Wine/European' winemaking lineage (Figure 2).

Although there is fine-scale population structure within the European oak lineage, even the divergence between the European oak and Wine/European lineages is smaller than the divergences seen in Taiwan (Lee et al., 2022), China, Japan and North America (Figure 1A). The lack of deep divergence among European tree strains (N=53) does not seem due to a lack of sampling because the number of good quality genomes is not substantially higher for North America (N = 66) and in other continents deeply diverged lineages can occur in close proximity (Lee et al., 2022; Figure 1C). Instead, the shallow divergences within Europe and within the European oak lineage suggest that *S. cerevisiae* colonised Europe and European forests more recently than in other regions.

2.4.4 Out-of-Asia migration of forest yeast since the last glacial maximum

We used a relative rate approach (Tamura et al., 2018) to estimate the timing of lineage divergences that likely correspond to (i) migration events out of Asia, (ii) into North America, and (iii) the origin of a Wine lineage distinct from European forest lineages. After excluding admixture, phylogenetic analysis of individual loci (30-60 kb) from each chromosome reproduced most genetically distinct clades defined in this study (Table S6 and Figure S12). Using these phylogenetic trees and assuming a mutation rate of 1.67×10^{-10} point substitutions per site per generation (Zhu et al., 2014) and 4 generations per day (see Methods), we generated rough estimates of divergence times.

East Asia is the probable origin for *S. cerevisiae* (Duan et al., 2018; Han et al., 2021; Wang et al., 2012) and we estimate that non-Asian lineages diverged from those only found in Asia approximately 20 thousand years ago (mean = 19.8 kya, 95% CI 16.9 - 22.7 kya; Figure 3). Out of 16 individual loci, 14 loci showed a clade for European oak (Table S7 and Figure S12). Time divergence estimation for these loci suggests European oak and the domesticated Wine/European lineage diverged around 6 kya (mean = 6.1 kya, 95% CI 4.5 - 7.6 kya; Figure 3). Out of 16 loci, 9 loci clustered the 3 North American clades (A-C), sometimes with Japan B (5 loci) or Japan A lineages (2 loci). This mostly North American clade appears to diverge from others (Table S7 and Figure S12) around 12 kya (mean = 11.6 kya, 95% CI 9.4 - 13.7 kya; Figure 3). These estimates suggest global *S. cerevisiae* migrations occurred since the last glacial maximum (Figure 3).

2.4.5 Occasional strains in Europe resemble present-day North American lineages

There are three strains from European trees (ZP530, 2163, and EXF6780) that resemble North American strains (black stars in Figure 1) and differ from all other European lineages (Figure 2). These genomes are from two different investigations where (i) ZP530 was isolated from chestnut (*Castanea sativa*) from Marão, Campeã, Portugal (Almeida et al. 2015), (ii) EXF6780 was isolated from sessile oak (*Quercus petraea*) from Velike Lašče, Kobila hill, Slovenia (Almeida et al., 2015) and (iii) 2163 was isolated from Portuguese oak (*Quercus faginea*) from Castellon, Spain (Peter et al., 2018). In the phylogenetic analysis, these strains differ from their most closely-related clades (North American B and C in Figure 1A) and appear to show some genetic

admixture (Figure 1B). Did these strains arrive on European trees as a result of ancient migration or could their genetic distance from other North American strains be explained by recent admixture? To find out, we “painted” their chromosomes according to the clade of the most closely related strain (Figure 4A). EXF6780, ZP530, and 2163 were compared to our backbone phylogeny (Table S5), which revealed the strains are a mix of three lineages found in North America and that two strains have admixture from the lineage used to ferment grape wine (Figure 4).

To assess genetic distances from modern North American lineages while accounting for admixture, we selected genomic regions from chromosomes 2 and 4 for phylogenetic analysis that did not show admixture from multiple clades (Figure 4A). Phylogenies with the effects of admixture removed in this way showed that strains 2163, EXF6780, and ZP530 seem almost identical to modern North American lineages across (Figure 4B). It therefore seems most likely that these ‘American European’ strains arrived very recently in Europe, and that admixture among North American and Wine/European lineages (Figure 4) explains their genetic distance from modern strains in the whole-genome phylogeny (Figure 1A).

Intriguingly, all three of these American European strains show some regions with close sequence similarity to CHN-VIII (blue windows in Figure 4A). One potential explanation is that the CHN-VIII lineage also has some recent domesticated Wine/European ancestry because all reference Wine/European strains resemble CHN-VIII in many genomic regions (Figure S8). Chromosome painting shows that one of the 3 strains representing CHN-VIII is very similar to reference strains from North American

C, American Wild and Wine/European lineages (> 10%, Figure S8). Locus-by-locus phylogenetic analysis (Figure S11) suggests that the remaining two CHN-VIII strains are diverged from the European oak lineage, yet indistinguishable in large genomic regions from the Wine/European lineage (chromosomes 4-7 and 11) and North American C (chromosomes 2 and 12). CHN-VIII genomes are also similar to the American wild lineage (chromosomes 3 and 15), and show regions diverged from any other lineages (chromosomes 1, 8, 9 and 16). The degree of similarity between modern Wine/European and North American C strains suggest very recent gene flow from Europe. Consistent with this proposal, CHN-VIII occurs in apple orchards and secondary forests near Beijing and Wine/European strains have also been isolated from orchards in China (Duan et al., 2018; Wang et al., 2012).

2.5 Discussion

2.5.1 Genetic isolation in forest niches

Before molecular methods showed that *S. cerevisiae* lives on trees (Naumov et al., 1992, Sniegowski et al., 2002), people thought it lived only with humans and not in natural environments (Vaughan-Martini & Martini 1995). This resembles thinking for *Candida* pathogenic species before their recent discovery on trees and other plants (Bensasson et al., 2019; Opulente et al., 2019). Yet our results suggest that *S. cerevisiae* recently colonized woodlands many times and live there in relative isolation from humans. Phylogeographic analysis of forest *S. cerevisiae* populations shows that strains are recognizably from Iberian, French or Italian trees with further lineages occurring in Eastern Europe (Figures 1 and 2). There are at least four forest lineages in North America

with regional differences; for example, North American C occurs in the southern USA, and North American A in Pennsylvania. The higher divergence seen among North American lineages (A-C), suggests that their most recent common ancestor must considerably predate the common ancestor of European *S. cerevisiae* and therefore an earlier arrival (see below). The genetic similarity of North American and Japanese lineages (A and B) suggests that the migration into North America, after the earlier arrival of American Wild, came from a different Asian source perhaps close to Japan. There is also genetic (mtDNA) and archaeological evidence suggesting shared ancestry among the people of China, Japan and the Americas suggesting a human migration route from northern coastal China into the Americas (Li et al., 2023).

Past analyses also show high population structure in *S. cerevisiae* from the primeval forests of East Asia (Wang et al., 2012; Lee et al., 2022) and in its sister species, *S. paradoxus* (Hénault et al., 2017; Leducq et al., 2014). *Saccharomyces* yeast are probably not usually air dispersed (Mortimer, 2000), therefore, it is not surprising that they show more population structure than other fungal microbes. Our observations for the tree niche contrast with the broader distribution of domesticated and fruit-associated lineages (Almeida et al., 2015; Duan et al., 2018; Gallone et al., 2016; Gonçalves et al., 2016; Lee et al., 2022; Legras et al., 2007; Legras et al., 2018; Peter et al., 2018) and are consistent with the proposal that animal-assisted long-distance migration is relatively rare in forests (Magwene et al., 2011; Tilakaratna & Bensasson 2017).

The evidence for isolated forest populations includes phylogenetic analyses using whole genome, single chromosome, and single locus data (Figures 1A, 2A, 4B, S5, and

S12) in addition to analyses of allele frequencies (Figures 1B, 2C, S7). Monophyletic tree-associated clades were reproducible across most chromosomes at the tips of phylogenetic trees, suggesting the fixation of many alleles for each lineage (Figures 1A, 2A, 4B, S5, and S12). Why might *S. cerevisiae* show many phylogenetically distinct lineages within continents? According to past estimates, *S. cerevisiae* reproduces sexually only once in hundreds or thousands of generations (Magwene et al., 2011; Lee et al., 2022) and the same is true for *S. paradoxus* (Tsai et al., 2008). Even when meiosis does occur, *S. paradoxus* are almost always selfing (99% of sexual cycles, Tsai et al., 2008). Asexual reproduction of *S. cerevisiae* might lead to population bottlenecks and the local fixation of alleles by genetic drift.

2.5.2 Global spread of *S. cerevisiae* forest populations since the last glacial maximum

Mutation rate estimates for *S. cerevisiae* applied to phylogenetic analyses suggest that the expansion of forest *S. cerevisiae* out of Asia likely occurred in the last 20,000 years (Figure 3 and S12). Divergence among the forest lineages of Europe and America is less deep than among the lineages of Asia (Figure 1A), so our analyses support an East Asian species origin (Wang et al., 2012; Lee et al., 2022) with the caveat that better sampling in other regions such as Central Asia or Central Africa could also reveal high genetic diversity. Averages estimated from phylogenies of 16 loci suggest that *S. cerevisiae* lineages migrated out of Asia around 17 - 23 kya (Figure 3 and S12), which was when climate started warming after the last glacial maximum (Clark et al., 2009). The forest lineages that first diverged from East Asian populations include those occurring in South America (American wild, French Guiana human; Figure S12). The

fine scale population structure within North American forests arose more recently; since the divergence of North American lineages (A-C) from most Asian lineages around 9 - 14 kya. The population structure occurring on European trees arose since these lineages diverged from the Wine lineage approximately 5 - 8 kya.

The timing of these yeast migrations seems to roughly coincide with human migration into America (15 - 23 kya), sedentarization (~ 14 kya), widespread of settlements in the Americas (12.6 -13 kya), the origins of agriculture (~ 10 kya), fermentation practices in Asia (~ 9 kya), European agriculture (6 - 7 kya) and European wine-making (~ 4 kya; Figure 3A; Marsit et al., 2017; McGovern 2003; Nielsen et al., 2017). It is therefore possible that humans or their commensals carried yeast with their food as they moved around the world. The alternative, that yeast migrated naturally across the globe as the climate warmed, seems less likely for multiple reasons: (i) *S. cerevisiae* are only rarely isolated from trees in northern Europe, and upon further genetic investigation northern populations appear feral or domesticated (the Wine/European lineage or admixed; Robinson et al., 2016). Tree-associated genetic lineages of *S. cerevisiae* have not been reported from northern cool temperate regions such as Canada or eastern Russia, despite documentation of other *Saccharomyces* species (Charron et al., 2014; Naumov et al., 2000). Indeed tree-associated genetic lineages may have a mostly subtropical or tropical distribution (Robinson et al., 2016). A natural expansion through the cold climate of the Bering land bridge therefore seems unlikely. (ii) Such an expansion seems more likely for *S. paradoxus* which inhabits northern forests in Canada, northern Europe and Siberia, yet that species shows much greater genetic isolation between America, Asia and Europe (Liti et al., 2009) perhaps because it does not live in

association with humans. (iii) *S. cerevisiae* differs from *S. paradoxus* in that it is common on fruit. Humans might have inadvertently carried *S. cerevisiae* with their food, vessels or the insects that traveled with them. (iv) Human-associated migration can explain a relatively recent origin of European lineages that is roughly coincident with the expansion of agriculture into Europe (5 - 8 kya), the much later spread to New Zealand in the last 1,000 years and their concentration near New Zealand vineyards (Gayevskiy et al., 2016). A less likely possibility is that *S. cerevisiae* was naturally dispersed by insects capable of migrating from warm regions in Asia to North America. Monarch butterflies have dispersed from America to Australia via Pacific islands since the last glacial maximum (Zhan et al., 2014). Such long-distance dispersal is rare however even for known long-distance dispersers like locusts, which do not have cosmopolitan distributions (Lovejoy et al., 2005) like those of humans or other human commensal organisms.

Our time estimates are based on the well-studied mutation rate of *S. cerevisiae*. The estimate we use (1.67×10^{-10} per base per generation) is the most accurate; from hundreds of point mutations (867) observed in 145 genome sequences from diploid mutation accumulation lines (Zhu et al., 2014). Earlier mutation rate estimates did not differ greatly: 2.9×10^{-10} from mutations accumulated in a different diploid background (Nishant et al., 2010), 3.3×10^{-10} in haploids (Lynch et al., 2008), or 1.84×10^{-10} from reporter assays in haploid strains at individual loci (Drake 1991; Fay & Benavides 2005). The number of generations occurring in natural forest environments is more difficult to measure (Mozzachiodi et al., 2022). As in past analysis of wine yeast by Fay and Benavides (ca. 2,920 generations per year; 2005), we assume a lower growth rate than in

the laboratory and only 12 hours of growth per day. The rate we use for trees (ca. 1,460 generations per year) assumes slower growth because of fewer nutrients on trees and no growth for 6 months of the year because of low temperatures (see Methods). For some parts of the species range, such as Florida and Georgia, generation times could be underestimated because some of the maximum temperatures in the coldest 6 months (18 - 29°C) and minimum nighttime temperatures in the hottest months (16 - 23°C) are also high enough for yeast growth (Table S8; Sweeney et al., 2004). While the fossil record is excellent for estimating divergences among yeast families or genera (Douzery et al., 2004; Marcet-Houben & Gabaldon 2015; Shen et al., 2018), the most recent fossil for ascomycotans dates to 417 million years ago (Douzery et al., 2004), and therefore may not be accurate for intraspecies divergences. Our timings suggest an older divergence (17 - 23 kya) than that of Fay & Benavides (2005) for the split of wine and East Asian sake strains (11.9 kya), and are consistent with past estimates for divergence among wine strains (3.7 kya; Fay & Benavides 2005), the split between European oak and wine 10.3 - 1.3 kya (Almeida et al., 2015), and the arrival of *S. cerevisiae* in New Zealand less than 1 kya (Gayevskiy et al., 2016).

2.5.3 Ongoing migration between human and tree environments

Not all *S. cerevisiae* strains on trees are from tree-associated lineages. Strains from the European grape wine lineage and other human-associated lineages also live on trees (Gayevskiy et al., 2016; Hyma & Fay 2013; Robinson et al., 2016). Indeed, there were at least 10 migration events from Europe to New Zealand trees that happened in the last 1,000 years; since humans arrived in New Zealand (Gayevskiy et al., 2016). Here we

observe strains on trees from clades connected with grape and African wines, Asian fermentations, brewing, baking, and clinical strains (Table S1, Figure S2), which suggests transmission between humans and trees is ongoing.

2.5.4 Footprints of human activity in the genomes of European tree strains

Three strains from trees in Portugal, Spain, and Slovenia had genome sequences that were predominantly from North American forest clades. Chromosome painting of their genomes shows these strains must be descended from three different transatlantic migrants (Figure 4A). These migrations from North America must have been recent because analysis of large (90 and 120 kb) loci shows genomic tracts that look typical of current North American B, C and American wild lineages (Figure 4B). None of these American European strains resemble strains from the North American A lineage, which only occurred in Pennsylvania (Table S1). Interestingly, strains from Portugal and Slovenia (ZP530 and EXF6780) resemble the North American C lineage (Figure 4), which we only observe in the southern USA (Figure 1C).

A potential explanation for the presence of North American tree-associated lineages in Europe is the human response to the Great French Wine Blight. In the 1850s, humans accidentally introduced an insect pest, *Phylloxera*, from North America to Europe that destroyed most European vineyards. Native American vines are naturally resistant to *Phylloxera*. The European wine industry was rescued by the mass import of vines from the southern USA to Europe, and from the late 1800s to the present day most European vines are grafted onto resistant North American grapevine rootstock (Campbell, 2004). The imported American grapevines could have harbored North

American yeast. In support of an explanation, Portuguese and Slovenian American strains show admixture from the Wine/European lineage, suggesting recent association with vineyards whereas European oak and North American forest lineages rarely show admixture from the wine lineage (Figure 4). The puzzling occurrence of long stretches of Wine/European and North American DNA in yeast strains from orchards and secondary forest near Beijing in China, could then be explained by very recent human-assisted migration from Europe.

2.6 Conclusion

In summary, our analyses show forests harbor many isolated *S. cerevisiae* populations that are distinct from human-associated lineages. The phylogeographic structure of tree-associated lineages implies that migrants from humans rarely establish in forest niches. Yet even rare events can shape the distribution of a species. The postglacial spread of forest *S. cerevisiae* out of Asia and into North America and Europe suggests that this substantial impact was driven by people. Consistent with this, we also observe footprints of ongoing human-assisted movement of forest yeast. Fungal microbes introduced into forests can transform landscapes when they are mutualists or parasites (Hoeksema et al., 2020; Averill et al., 2022). For intimate human commensals and occasional pathogens, such as *S. cerevisiae* and *Candida sp.*, it seems important to consider their evolution in non-human environments - especially since environmental fungal microbes may adapt to fungicide use or rising temperatures (Garcia-Solache & Casadevall 2010; Leducq et al., 2014, Kang et al., 2022, Lockhart et al., 2023).

2.7 Acknowledgements

This work was funded by an NSF LSAMP Bridge to the Doctorate Fellowship Program award (grant number: 1702361) to JJP; by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program (grant no. GT13544) awarded to JJP and DB; and by a National Science Foundation grant (IOS, no. 1946046) awarded to DB. For some of the new genomes used, we would like to thank the Florida State Parks Service, the Duke Forest Teaching and Research Laboratory, Tyler Arboretum, Thompson Mills Forest, Skidaway Institute of Oceanography, and the University of Georgia for their support of fieldwork, and Diana Ambrocio, Domenic Won and Linda Habersham for strain isolation. Past and current members of the Bensasson lab provided helpful feedback. We would like to acknowledge that most of the research performed for this study was on Cherokee, Yuchi, and Muscogee indigenous lands.

2.8 References

- Averill, C., Anthony, M. A., Baldrian, P., Finkbeiner, F., van den Hoogen, J., Kiers, T., Kohout, P., Hirt, E., Smith, G. R., & Crowther, T. W. (2022). Defending Earth's terrestrial microbiome. *Nature Microbiology*, 7(11), 1717–1725.
<https://doi.org/10.1038/s41564-022-01228-3>
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655–1664.
<https://doi.org/10.1101/gr.094052.109>
- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J.-L., Serra, M., Dequin, S., Couloux, A., Guy, J., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular Ecology*, 24(21), 5412–5427.
<https://doi.org/10.1111/mec.13341>
- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of Natural Hybridization in Brazilian Wild Lineages of

Saccharomyces cerevisiae. *Genome Biology and Evolution*, 8(2), 317–329.
<https://doi.org/10.1093/gbe/evv263>

Bensasson, D. (2011) Evidence for a high mutation rate at rapidly evolving yeast centromeres, *BMC Evol Biol*, 11, 211. <https://doi.org/10.1186/1471-2148-11-211>

Bensasson, D., Dicks, J., Ludwig, J. M., Bond, C. J., Elliston, A., Roberts, I. N., & James, S. A. (2019). Diverse Lineages of *Candida albicans* Live on Old Oaks. *Genetics*, 211(1), 277–288. <https://doi.org/10.1534/genetics.118.301482>

Bergström, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. *Molecular Biology and Evolution*, 31(4), 872–888.
<https://doi.org/10.1093/molbev/msu037>

Boynton, P. J., & Greig, D. (2014). The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast*, 31(12), 449–462. <https://doi.org/10.1002/yea.3040>

Campbell, C. (2004). *Phylloxera: How wine was saved for the world*. London : HarperCollins Publishers. <http://archive.org/details/phylloxerahowwin0000camp>

Charron, G., Leducq, J., & Landry, C. R. (2014). Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Molecular Ecology*, 23(17), 4362–4372. <https://doi.org/10.1111/mec.12864>

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, J. X., Hostetler, S. W., & McCabe, A. M. (2009). The Last Glacial Maximum. *Science*, 325(5941), 710–714.
<https://doi.org/10.1126/science.1172873>

Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., Huang, A. A., Dudley, A. M., & Fay, J. C. (2013). Genomic Sequence Diversity and Population Structure of *Saccharomyces cerevisiae* Assessed by RAD-seq. *G3: Genes/Genomes/Genetics*, 3(12), 2163–2171.
<https://doi.org/10.1534/g3.113.007492>

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158.
<https://doi.org/10.1093/bioinformatics/btr330>

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.
<https://doi.org/10.1371/journal.pone.0005317>

- Douzery, E. J. P., Snell, E. A., Baptiste, E., Delsuc, F., & Philippe, H. (2004). The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proceedings of the National Academy of Sciences*, 101(43), 15386–15391. <https://doi.org/10.1073/pnas.0403984101>
- Drake, J. W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences*, 88(16), 7160–7164. <https://doi.org/10.1073/pnas.88.16.7160>
- Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., & Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nature Communications*, 9(1), 2690. <https://doi.org/10.1038/s41467-018-05106-7>
- Fay, J. C., & Benavides, J. A. (2005). Evidence for Domesticated and Wild Populations of *Saccharomyces cerevisiae*. *PLOS Genetics*, 1(1), e5. <https://doi.org/10.1371/journal.pgen.0010005>
- Fay, J. C., Liu, P., Ong, G. T., Dunham, M. J., Cromie, G. A., Jeffery, E. W., Ludlow, C. L., & Dudley, A. M. (2019). A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. *PLOS Biology*, 17(3), e3000147. <https://doi.org/10.1371/journal.pbio.3000147>
- Francis, R. M. (2017). pophelper: An R package and web app to analyse and visualize population structure. *Molecular Ecology Resources*, 17(1), 27–32. <https://doi.org/10.1111/1755-0998.12509>
- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., & Verstrepen, K. J. (2016). Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. *Cell*, 166(6), 1397-1410.e16. <https://doi.org/10.1016/j.cell.2016.08.020>
- Garcia-Solache, M. A., & Casadevall, A. (2010). Global Warming Will Bring New Fungal Diseases for Mammals. *mbio*, 1(1), 10.1128/mbio.00061-10. <https://doi.org/10.1128/mbio.00061-10>
- Gayevskiy, V., Lee, S., & Goddard, M. R. (2016). European derived *Saccharomyces cerevisiae* colonisation of New Zealand vineyards aided by humans. *FEMS Yeast Research*, 16(7), fow091. <https://doi.org/10.1093/femsyr/fow091>
- Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., & Sampaio, J. P. (2016). Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts. *Current Biology*, 26(20), 2750–2761. <https://doi.org/10.1016/j.cub.2016.08.040>

- Gouy, M., Guindon, S., Gascuel, O. (2010) SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building, *Molecular Biology and Evolution*, 27(2), 221–224.
- Han, D.-Y., Han, P.-J., Rumbold, K., Koricha, A. D., Duan, S.-F., Song, L., Shi, J.-Y., Li, K., Wang, Q.-M., & Bai, F.-Y. (2021). Adaptive Gene Content and Allele Distribution Variations in the Wild and Domesticated Populations of *Saccharomyces cerevisiae*. *Frontiers in Microbiology*, 12, 631250. <https://doi.org/10.3389/fmicb.2021.631250>
- Hénault, M., Eberlein, C., Charron, G., Durand, É., Nielly-Thibault, L., Martin, H., & Landry, C. R. (2017). *Population Genomics* (pp. 207–230). https://doi.org/10.1007/13836_2017_4
- Hoeksema, J. D., Averill, C., Bhatnagar, J. M., Brzostek, E., Buscardo, E., Chen, K.-H., Liao, H.-L., Nagy, L., Policelli, N., Ridgeway, J., Rojas, J. A., & Vilgalys, R. (2020). Ectomycorrhizal Plant-Fungal Co-invasions as Natural Experiments for Connecting Plant and Fungal Traits to Their Ecosystem Consequences. *Frontiers in Forests and Global Change*, 3. <https://doi.org/10.3389/ffgc.2020.00084>
- Hyma, K. E., & Fay, J. C. (2013). Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. *Molecular Ecology*, 22(11), 2917–2930. <https://doi.org/10.1111/mec.12155>
- Johnson, L. J., Koufopanou, V., Goddard, M. R., Hetherington, R., Schäfer, S. M., & Burt, A. (2004). Population Genetics of the Wild Yeast *Saccharomyces paradoxus*. *Genetics*, 166(1), 43–52. <https://doi.org/10.1534/genetics.166.1.43>
- Kang, S. E., Sumabat, L. G., Melie, T., Mangum, B., Momany, M., & Brewer, M. T. (2022). Evidence for the agricultural origin of resistance to multiple antimicrobials in *Aspergillus fumigatus*, a fungal pathogen of humans. *G3 Genes/Genomes/Genetics*, 12(2), jkab427. <https://doi.org/10.1093/g3journal/jkab427>
- Kaya, A., Phua, C. Z. J., Lee, M., Wang, L., Tyshkovskiy, A., Ma, S., Barre, B., Liu, W., Harrison, B. R., Zhao, X., Zhou, X., Wasko, B. M., Bammler, T. K., Promislow, D. E., Kaeberlein, M., & Gladyshev, V. N. (2021). Evolution of natural lifespan variation and molecular strategies of extended lifespan in yeast. *eLife*, 10, e64860. <https://doi.org/10.7554/eLife.64860>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. <https://doi.org/10.1111/1755-0998.12387>

- Kowallik, V., Miller, E., & Greig, D. (2015). The interaction of *Saccharomyces paradoxus* with its natural competitors on oak bark. *Molecular Ecology*, 24(7), 1596–1610. <https://doi.org/10.1111/mec.13120>
- Kumar, S., Stecher, G., Peterson, D., & Tamura, K. (2012). MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics*, 28(20), 2685–2686. <https://doi.org/10.1093/bioinformatics/bts507>
- Leducq, J.-B., Charron, G., Samani, P., Dubé, A. K., Sylvester, K., James, B., Almeida, P., Sampaio, J. P., Hittinger, C. T., Bell, G., & Landry, C. R. (2014). Local climatic adaptation in a widespread microorganism. *Proceedings of the Royal Society B: Biological Sciences*, 281(1777), 20132472. <https://doi.org/10.1098/rspb.2013.2472>
- Lee, T. J., Liu, Y.-C., Liu, W.-A., Lin, Y.-F., Lee, H.-H., Ke, H.-M., Huang, J.-P., Lu, M.-Y. J., Hsieh, C.-L., Chung, K.-F., Liti, G., & Tsai, I. J. (2022). Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages. *Genome Research*, 32(5), 864–877. <https://doi.org/10.1101/gr.276286.121>
- Legras, J., Merdinoglu, D., Cornuet, J., & Karst, F. (2007). Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular Ecology*, 16(10), 2091–2102. <https://doi.org/10.1111/j.1365-294x.2007.03266.x>
- Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldon, T., Schuller, D., Sampaio, J. P., & Dequin, S. (2018). Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. *Molecular Biology and Evolution*, 35(7), 1712–1727. <https://doi.org/10.1093/molbev/msy066>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O’Kelly, M. J. T., van Oudenaarden, A., Barton, D. B. H., Bailes, E., Nguyen Ba, A. N., Jones, M., ... Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, 458(7236), 337–341. <https://doi.org/10.1038/nature07743>

- Lockhart, S. R., Chowdhary, A., & Gold, J. A. W. (2023). The rapid emergence of antifungal-resistant human-pathogenic fungi. *Nature Reviews Microbiology*, 21(12), 818–832. <https://doi.org/10.1038/s41579-023-00960-9>
- Lovejoy, N. r et al. (2005) Ancient trans-Atlantic flight explains locust biogeography: molecular phylogenetics of *Schistocerca*. *Proceedings of the Royal Society B: Biological Sciences*, 273(1588), 767–774. <https://doi.org/10.1098/rspb.2005.3381>
- Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., Dickinson, W. J., Okamoto, K., Kulkarni, S., Hartl, D. L., & Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences*, 105(27), 9272–9277. <https://doi.org/10.1073/pnas.0803466105>
- Mable, B. K. (2001). Ploidy evolution in the yeast *Saccharomyces cerevisiae*: A test of the nutrient limitation hypothesis. *Journal of Evolutionary Biology*, 14(1), 157–170. <https://doi.org/10.1046/j.1420-9101.2001.00245.x>
- Magwene, P. M., Kayıkçı, Ö., Granek, J. A., Reininga, J. M., Scholl, Z., & Murray, D. (2011). Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 108(5), 1987–1992. <https://doi.org/10.1073/pnas.1012544108>
- Marcet-Houben, M., & Gabaldón, T. (2015). Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS Biology*, 13(8), e1002220. <https://doi.org/10.1371/journal.pbio.1002220>
- Marsit, S., Leducq, J.-B., Durand, É., Marchant, A., Filteau, M., & Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nature Reviews Genetics*, 18(10), 581–598. <https://doi.org/10.1038/nrg.2017.49>
- McGovern, P. E. (2003). *Ancient Wine: The Search for the Origins of Viniculture*. Princeton University Press. <https://doi.org/10.2307/j.ctvfjd0bk>
- McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., Nuñez, A., Butrym, E. D., Richards, M. P., Wang, C., Cheng, G., Zhao, Z., & Wang, C. (2004). Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences*, 101(51), 17593–17598. <https://doi.org/10.1073/pnas.0407921102>
- Michel, R. H., McGovern, P. E., & Badler, V. R. (1992). Chemical evidence for ancient beer. *Nature*, 360(6399), 24–24. <https://doi.org/10.1038/360024b0>

- Minh, B. Q., Nguyen, M. A. T., & Haeseler, A. von. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, 30(5), 1188–1195. <https://doi.org/10.1093/molbev/mst024>
- Mortimer, R. K. (2000). Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Research*, 10(4), 403–409. <https://doi.org/10.1101/gr.10.4.403>
- Mozzachiodi, S., Bai, F., Baldrian, P., Bell, G., Boundy-Mills, K., Buzzini, P., Čadež, N., Cubillos, F. A., Dashko, S., Dimitrov, R., Fisher, K. J., Gibson, B., Gouliamova, D., Greig, D., Heistinger, L., Hittinger, C. T., Jecmenica, M., Koufopanou, V., Landry, C. R., ... Boynton, P. (2022). Yeasts from temperate forests. *Yeast*. <https://doi.org/10.1002/yea.3699>
- Naumov, G.I. et al. (1992) Genetic homology between *Saccharomyces cerevisiae* and its sibling species *S. paradoxus* and *S. bayanus*: Electrophoretic karyotypes, *Yeast*, 8(8), 599–612. <https://doi.org/10.1002/yea.320080804>
- Naumov, G. et al. (2000) ‘Three new species in the *Saccharomyces* sensu stricto complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*’, *Int J Syst Evol Microbiol*, 50(5), 931–1942.
- Nguyen, L.-T., Schmidt, H. A., Haeseler, A. von, & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, 541(7637), Article 7637. <https://doi.org/10.1038/nature21347>
- Nishant, K. T., Wei, W., Mancera, E., Argueso, J. L., Schlattl, A., Delhomme, N., Ma, X., Bustamante, C. D., Korbel, J. O., Gu, Z., Steinmetz, L. M., & Alani, E. (2010). The Baker’s Yeast Diploid Genome Is Remarkably Stable in Vegetative Growth and Meiosis. *PLOS Genetics*, 6(9), e1001109. <https://doi.org/10.1371/journal.pgen.1001109>
- Opulente, D. A., Langdon, Q. K., Buh, K. V., Haase, M. A. B., Sylvester, K., Moriarty, R. V., Jarzyna, M., Considine, S. L., Schneider, R. M., & Hittinger, C. T. (2019). Pathogenic budding yeasts isolated outside of clinical settings. *FEMS Yeast Research*, 19(3). <https://doi.org/10.1093/femsyr/foz032>
- Osburn, K., Amaral, J., Metcalf, S. R., Nickens, D. M., Rogers, C. M., Sausen, C., Caputo, R., Miller, J., Li, H., Tennessen, J. M., & Bochman, M. L. (2018). Primary souring: A novel bacteria-free method for sour beer production. *Food Microbiology*, 70, 76–84. <https://doi.org/10.1016/j.fm.2017.09.007>

- Paradis, E., & Schliep, K. (2018). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Peter, J., Chiara, M. D., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istance, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), 339. <https://doi.org/10.1038/s41586-018-0030-5>
- Pontes, A., Hutzler, M., Brito, P. H., & Sampaio, J. P. (2020). Revisiting the Taxonomic Synonyms and Populations of *Saccharomyces cerevisiae*—Phylogeny, Phenotypes, Ecology and Domestication. *Microorganisms*, 8(6), Article 6. <https://doi.org/10.3390/microorganisms8060903>
- Potapov, P. *et al.* (2017) ‘The last frontiers of wilderness: Tracking loss of intact forest landscapes from 2000 to 2013’, *Science Advances*, 3(1), e1600821. <https://doi.org/10.1126/sciadv.1600821>
- Pretorius, I. S. (2000). Tailoring wine yeast for the new millennium: Novel approaches to the ancient art of winemaking. *Yeast*, 16(8), 675–729. [https://doi.org/10.1002/1097-0061\(20000615\)16:8<675::aid-yea585>3.0.co;2-b](https://doi.org/10.1002/1097-0061(20000615)16:8<675::aid-yea585>3.0.co;2-b)
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155(2), 945–959. <https://doi.org/10.1093/genetics/155.2.945>
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., Bakker, P. I. W. de, Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3), 559–575. <https://doi.org/10.1086/519795>
- Robinson, H. A., Pinharanda, A., & Bensasson, D. (2016). Summer temperature can predict the distribution of wild yeast populations. *Ecology and Evolution*, 6(4), 1236–1250. <https://doi.org/10.1002/ece3.1919>
- Schacherer, J., Shapiro, J. A., Ruderfer, D. M., & Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature*, 458(7236), 342–345. <https://doi.org/10.1038/nature07670>
- Scopel, E. F. C., Hose, J., Bensasson, D., & Gasch, A. P. (2021). Genetic variation in aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages. *Genetics*, iyab015-. <https://doi.org/10.1093/genetics/iyab015>
- Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T.,

- Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Čadež, N., Libkind, D., ... Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, 175(6), 1533-1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>
- Sherman, F. (2002). Getting started with yeast. *Methods in Enzymology*, 350, 3–41. [https://doi.org/10.1016/s0076-6879\(02\)50954-x](https://doi.org/10.1016/s0076-6879(02)50954-x)
- Skelly, D. A., Merrihew, G. E., Riffle, M., Connelly, C. F., Kerr, E. O., Johansson, M., Jaschob, D., Graczyk, B., Shulman, N. J., Wakefield, J., Cooper, S. J., Fields, S., Noble, W. S., Muller, E. G. D., Davis, T. N., Dunham, M. J., MacCoss, M. J., & Akey, J. M. (2013). Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Research*, 23(9), 1496–1504. <https://doi.org/10.1101/gr.155762.113>
- 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. <https://doi.org/10.1111/j.1567-1364.2002.tb00048.x>
- Song, G., Dickins, B. J. A., Demeter, J., Engel, S., Gallagher, J., Choe, K., Dunn, B., Snyder, M., & Cherry, J. M. (2015). AGAPE (Automated Genome Analysis PipelinE) for Pan-Genome Analysis of *Saccharomyces cerevisiae*. *PLOS ONE*, 10(3), e0120671. <https://doi.org/10.1371/journal.pone.0120671>
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., & McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Research*, 25(5), 762–774. <https://doi.org/10.1101/gr.185538.114>
- 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. [https://doi.org/10.1016/S1567-1356\(03\)00171-5](https://doi.org/10.1016/S1567-1356(03)00171-5)
- Tamura, K., & Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, 10(3), 512–526. <https://doi.org/10.1093/oxfordjournals.molbev.a040023>
- Tamura, K., Tao, Q., & Kumar, S. (2018). Theoretical Foundation of the RelTime Method for Estimating Divergence Times from Variable Evolutionary Rates. *Molecular Biology and Evolution*, 35(7), 1770–1782. <https://doi.org/10.1093/molbev/msy044>

- Tilakaratna, V., & Bensasson, D. (2017). Habitat Predicts Levels of Genetic Admixture in *Saccharomyces cerevisiae*. *G3: Genes, Genomes, Genetics*, 7(9), 2919–2929. <https://doi.org/10.1534/g3.117.041806>
- Tsai, I. J., Bensasson, D., Burt, A., & Koufopanou, V. (2008). Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proceedings of the National Academy of Sciences*, 105(12), 4957–4962. <https://doi.org/10.1073/pnas.0707314105>
- Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A., & Bai, F.-Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular Ecology*, 21(22), 5404–5417. <https://doi.org/10.1111/j.1365-294X.2012.05732.x>
- Willerslev, E., & Meltzer, D. J. (2021). Peopling of the Americas as inferred from ancient genomics. *Nature*, 594(7863), 356–364. <https://doi.org/10.1038/s41586-021-03499-y>
- Xu, S., Dai, Z., Guo, P., Fu, X., Liu, S., Zhou, L., Tang, W., Feng, T., Chen, M., Zhan, L., Wu, T., Hu, E., Jiang, Y., Bo, X., & Yu, G. (2021). ggtreeExtra: Compact Visualization of Richly Annotated Phylogenetic Data. *Molecular Biology and Evolution*, 38(9), 4039–4042. <https://doi.org/10.1093/molbev/msab166>
- Yu, G., Lam, T. T.-Y., Zhu, H., & Guan, Y. (2018). Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Molecular Biology and Evolution*, 35(12), 3041–3043. <https://doi.org/10.1093/molbev/msy194>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. <https://doi.org/10.1111/2041-210x.12628>
- Yue, J.-X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6), 913–924. <https://doi.org/10.1038/ng.3847>
- Zhan, S. *et al.* (2014) The genetics of monarch butterfly migration and warning colouration, *Nature*, 514(7522), 317–321. <https://doi.org/10.1038/nature13812>
- Zhu, Y. O., Siegal, M. L., Hall, D. W., & Petrov, D. A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proceedings of the National Academy of Sciences*, 111(22), E2310–E2318. <https://doi.org/10.1073/pnas.1323011111>

2.9 Data Accessibility and Benefit-Sharing

Short read genome data are available in the NCBI-SRA under project number PRJNA1090965. Consensus genome sequences generated by mapping to the sacCer3 reference, alignments and PLINK files used in ADMIXTURE analyses are available on Dryad under doi: 10.5061/dryad.pnvx0k6zq. Benefits from this research include this sharing of our data and results on public databases, and the sharing of the yeast strains isolated for this work which are available from our lab or public yeast collections.

2.10 Author Contributions

The research was conceptualized and designed by JJP and DB. AKW, JJP and EFCS performed DNA extractions for genome sequencing. JJP and EFCS developed bioinformatic pipelines, obtained and curated the data. JJP performed analyses and data visualization. JJP and DB wrote the paper with input from EFCS and AKW.

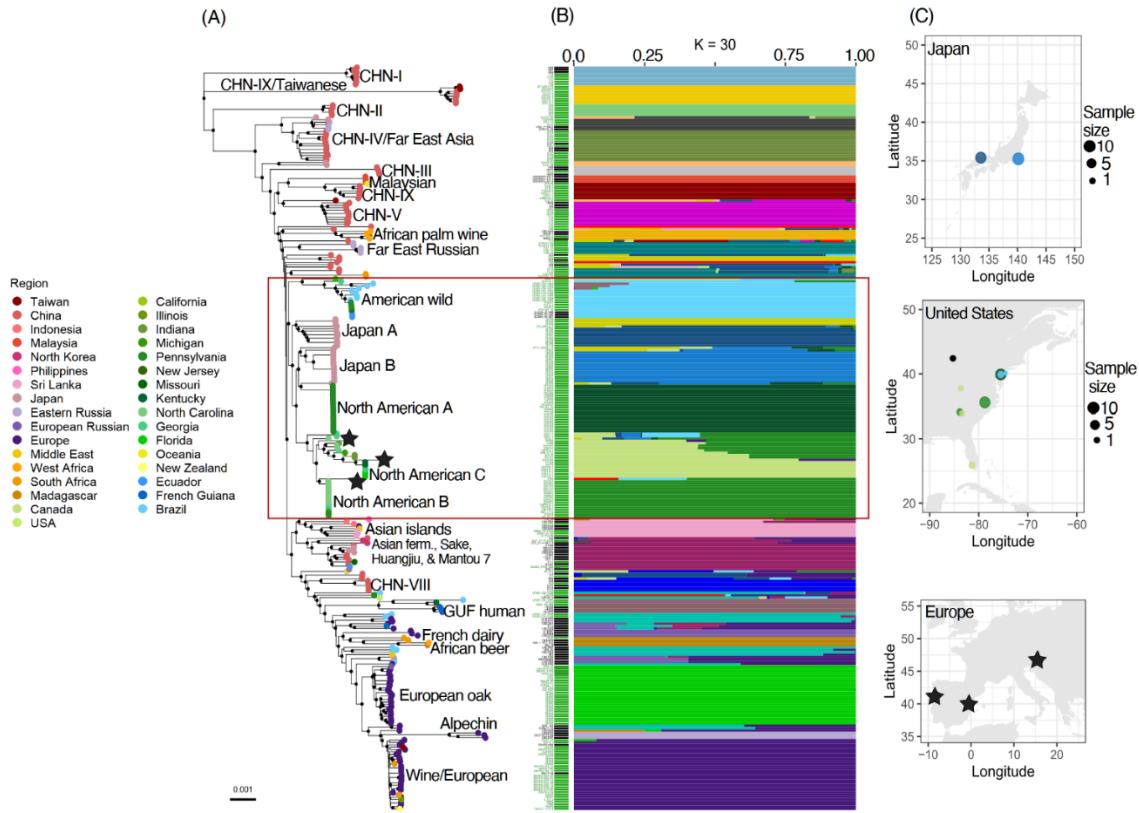


Figure 2.1 Trees harbor numerous genetically distinct *S. cerevisiae* lineages with population substructure in North America and Japan. (A) Whole-genome neighbor-joining tree of 313 strains after excluding heterozygous strains (Table S1). Strains isolated from trees are shown with green text and bar and reference panel strains with black. Black circles at nodes indicate bootstrap support > 95%. Colored circles at the tips of the tree show geographical origin. All strains within the red box were isolated from America or Japan except for 3 European strains (black stars). **(B)** ADMIXTURE plot with $K = 30$ showing the cluster ancestry proportion for each strain. **(C)** Maps showing the geographic source of Japanese, North American, and European (black stars in A) tree-sampled strains from the America and Japan lineages (red box in A). Circle sizes are based on square-root transformed sample sizes Japanese and North American strains are color coded by ancestry from ADMIXTURE plot. We removed North American and Japanese admixed strains (< 90% single-lineage ancestry) from maps for simplicity.

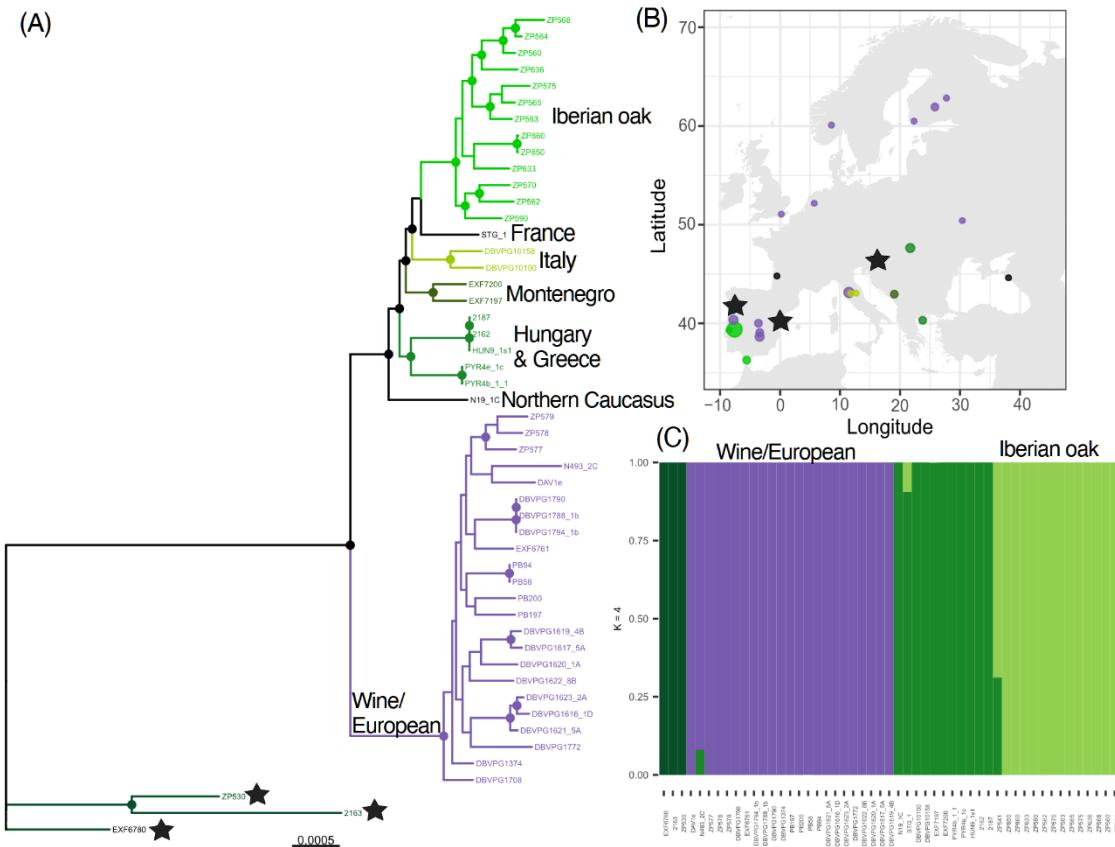


Figure 2.2 Fine-scale population structure of wild *S. cerevisiae* from Europe. (A)
 Whole-genome maximum likelihood phylogenetic tree of 50 strains after excluding one admixed strain, ZP541 from (C). Black circles at nodes indicate bootstrap support > 95%. Branches are color coded by geography or by ecology. Black stars at tree tips denote strains that genetically cluster with North American strains in Figure 1. **(B)** Map of Europe showing the geographic source of strains, circles are sized by the square-root transformed sample sizes and color coded by branch colors in the phylogenetic tree. Singleton strains from France and Northern Caucasus are colored in black. Black stars denote strains that genetically cluster with North American strains in Figure 1. **(C)** ADMIXTURE plot when $K = 4$ to examine percent ancestry per individual strain.

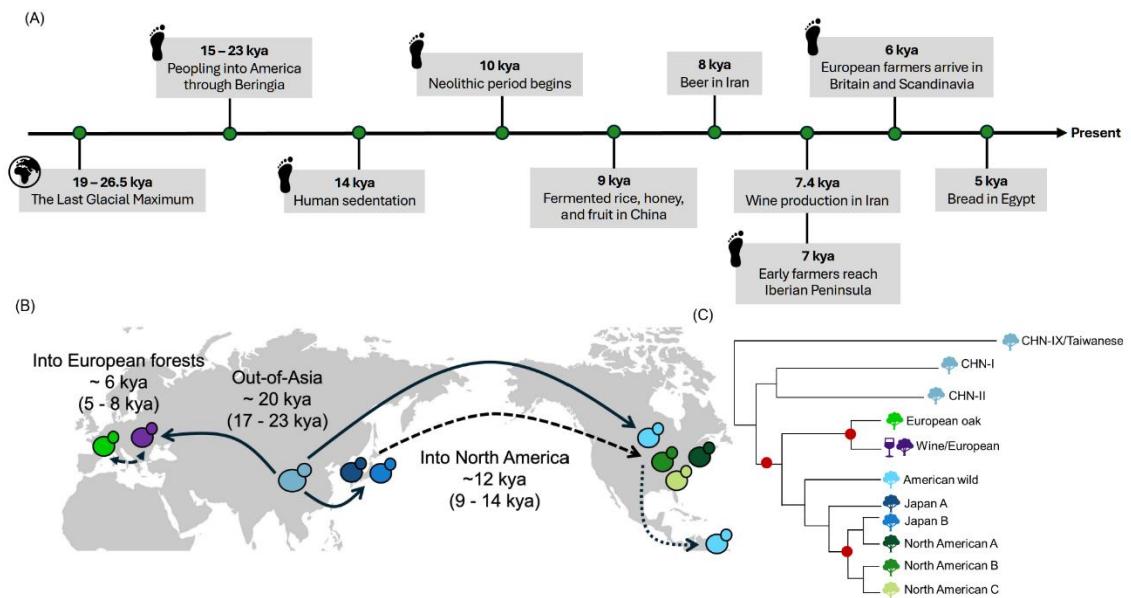


Figure 2.3 Out-of-Asia migration of forest yeast since the Last Ice Age. (A) Timeline showing early archaeological evidence of fermentation and human migration (dates from Clark et al., 2009; Marsit et al., 2017; Nielsen et al., 2017). **(B)** Map showing forest yeast migration events depicted with a red dot at nodes in (C) (i) out of Asia, (ii) into North America, and (iii) into European forests. Lineages are color coded by the lineages in the cladogram in (C). Dashed arrows indicate secondary migration events. **(C)** Cladogram showing phylogenetic relationships of lineages of interest for date estimation.

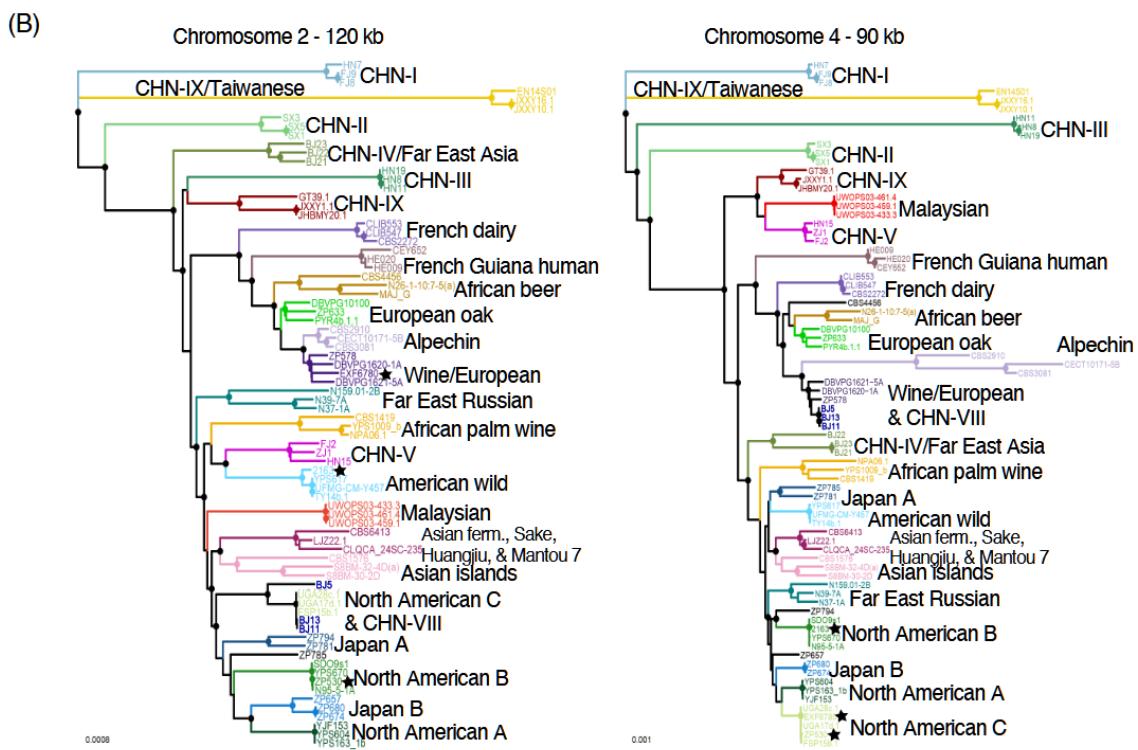


Figure 2.4 Occasional strains in Europe resemble present-day North American lineages. (A) Painted chromosomes of 2163, EXF6780, and ZP530 show admixture between multiple lineages. Genomic regions were “painted” based on the clade assignment of the most similar strain in 30 kb non-overlapping windows. Diverged regions were not colored (white) and were defined as regions that differed by 0.003 from all other strains in the backbone phylogeny. Black colored regions indicate low coverage. Colors are as in (B): North American B is forest green, American wild is light blue, North American C is light green, Wine/European is dark purple, and CHN-VIII is blue. Genomic regions (90-120 kb) were selected for phylogenetic analysis (red boxes with black stars). (B) Neighbor-joining phylogenetic trees for two loci. Solid circles at nodes indicate bootstrap support > 95%. Branches are color coded by clade. Phylogenetic analyses show that in the absence of admixture, 2163, EXF6780 and ZP530 (black stars) are very similar to strains from American wild, Wine/European, North American B, and C.

CHAPTER 3

CLIMATE CAN PREDICT THE SPECIES RANGES OF SYMPATRIC YEASTS FROM FORESTS

3.1 Introduction

In a rapidly changing climate, there is an urgent need to identify the conditions necessary for adaptation or maladaptation. Correlative ecological models can determine species distribution and abundance and identify which climate variables might limit species ranges (Araújo & Peterson, 2012; Pacifici et al., 2015). Furthermore, correlative models gain power if inferences are extended to sympatric species with contrasting ecological niches or if multiple populations from a single species are examined across a broad geographical range (Sexton et al., 2009). However, correlative models only give a snapshot of which climatic variables shape a species distribution, but in tandem with phylogeography can help us understand how postglacial refugia have shaped current natural populations (Gavin et al., 2014; Lafontaine et al., 2018). This could reveal unrelated populations or species that independently evolve to harsh environmental conditions in similar ways, thus providing natural experimental replicates (Lee & Coop, 2019).

Inferences made from correlative models and phylogeography are proving to be effective in understanding the vulnerability of many species of plants, birds or mammals to climate change (Pacifici et al., 2015). However, fungi have been studied less (Willis,

2018). Sympatric yeast with contrasting thermal niches isolated from the wild can be used to study which climate associations are important for natural populations. Even domesticated yeast species used to ferment food and beverages can be isolated from the wild and have a diverse phylogeographic pattern (Almeida et al., 2015; Langdon et al., 2020; Leducq et al., 2014; Nespolo et al., 2019; Peña et al., 2025) that can be used to study population biology. For instance, two members of the *Saccharomyces* genus, *Saccharomyces eubayanus*, and *S. uvarum*, distributions have been shaped by postglacial climate shifts (Langdon et al., 2020; Nespolo et al., 2019). Most species within the *Saccharomyces* genus, including the model organism, *S. cerevisiae* have been isolated from cold temperate regions (Charron et al., 2014; Leducq et al., 2014; Mozzachiodi et al., 2022; Naumov et al., 1998; Robinson et al., 2016; Sylvester et al., 2015).

Current climate envelope models using temperature suggest that hot climatic regions such as the subtropics and tropics are likely the natural ecological niche for wild *S. cerevisiae* (Robinson et al., 2016), but these regions are under sampled. In addition, these climate envelope models are based on a comparison of thermal growth optima under laboratory settings (Sweeney et al., 2004), which do not reflect wild environments. The biogeography of wild *Saccharomyces* yeasts has likely been shaped by temperature where sympatric species or populations have different thermal growth preferences (Gonçalves et al., 2011; Langdon et al., 2020; Leducq et al., 2014; Nespolo et al., 2020; Peris et al., 2023; Robinson et al., 2016; Salvadó et al., 2011; Sampaio & Gonçalves, 2008; Sniegowski et al., 2002; Spurley et al., 2022; Sweeney et al., 2004; Sylvester et al., 2015). Based on previous models generated from European forest data (Robinson et al., 2016) and laboratory observations of thermal growth optima (Sweeney et al., 2004), we

hypothesize that climatic region and tree girth will predict the prevalence of yeasts. Here, we sampled bark from oak trees (*Quercus*) that were over 100 years old across a natural temperature gradient along the eastern United States to determine what climatic conditions could limit the species ranges of *S. cerevisiae*. Specifically, we targeted *Quercus* spp. because *S. cerevisiae* and other *Saccharomyces* spp. are relatively abundant on *Quercus* spp. (Alsammar & Delneri, 2020; Boynton & Greig, 2014; Mozzachiodi et al., 2022) and yeasts are abundant on older trees (Robinson et al., 2016). We show that (i) a broad range of wild yeast species can be isolated from forests and marshes, (ii) *S. cerevisiae*, *S. paradoxus*, *Lachancea fermentati*, and *Lachancea thermotolerans* were the most common species isolated from forests, which allowed us to extend our analyses to species other than *S. cerevisiae* and lastly, (ii) *S. cerevisiae*, *S. paradoxus*, and *L. fermentati* isolation frequency varies by temperature and precipitation across continents.

3.2 Materials and methods

3.2.1 Collecting environmental samples in the eastern United States

Environmental samples from angiosperms (tree bark N = 825, flowers N = 23, and fig N = 12), gymnosperms (tree bark N = 74 and exudate N = 4), marsh sediment (N = 15), arthropods (N= 2), and mushrooms (N = 3) were collected from the eastern United States between September 2019 to October 2022 (Figure 1 and Table S1). Samples were collected into 7 mL bijou or 30 mL Falcon tubes, as Robinson et al. (2016) described. We used sterile techniques by wearing gloves and spraying gloved hands with 70% ethanol. For each substrate sampled, we took photographs and recorded latitude and longitude. We recorded the collection date to control for resampling in different years. For tree

hosts, we obtained the circumference of the tree at chest height and four bark samples per tree. At the Skidaway Institute of Oceanography (Savannah, Georgia) we sampled marsh sediment in October 2021. Sediment was scooped into a sterile 30 mL Falcon tube, leaving room for enrichment medium. All hosts and sediment sampling points were associated with a separate negative control: gloved fingers used for sampling were wiped into an empty tube to account for potential contamination. Three out of 188 negative controls (FSP14-ve, TMF27-ve, and UGA29-ve) were contaminated with yeast growth at the initial stage of streaking 30-50 µL aliquots onto a solid medium. Environmental samples (four per host) that were associated with these contaminated negative controls were excluded from all analyses (N = 9 environmental samples). Each contaminant was cultured and identified as *Lachancea fermentati* (FSP14-ve), *Meyerozyma guilliermondii* (TMF27-ve), and *L. thermotolerans* (UGA29-ve).

3.2.2 Microclimate and longitudinal sampling from Georgia

Bark samples were collected from the tree's north and south cardinal directions (two bark samples each) across all field sites to see if there are microclimate effects on yeast isolation frequency. To determine year-long microclimate effects, temperature (°C) from five *Quercus alba* (white oak) trees from Thompson Mills Forest were monitored on the north side and south side of the tree between November 2021 and November 2022 using iButtons data loggers to record daily temperature every four hours from 8:00 a.m. (Table S2). Additionally, we collected bark samples from *Q. alba* and *Q. virginiana* (live oak) trees across multiple years (2019-2022) to see if yeast frequency varies annually across all field sites within Georgia. Lastly, we sampled different angiosperm trees

(*Quercus* spp., *Liriodendron tulipifera*, *Acer saccharum*, *Magnolia virginiana*, and *Ulmus americana*) and gymnosperm trees (*Cedrus deodara*, *Pinus echinata*, and *P. taeda*) to determine if yeast species and isolation rates differ.

3.2.3 Culturing and identification of wild yeast species

Each environmental sample was weighed, and the sample weight of the collection tube was recorded before and after the sample was collected. Then, each tube was filled with a liquid medium that enriches for *Saccharomyces* or yeast species that have ethanol tolerance and that inhibited DNA replication of prokaryotic DNA using chloramphenicol as an antibiotic (Robinson et al., 2016; Sniegowski et al., 2002). Environmental samples were incubated with the enrichment medium for at least two weeks at 30°C. Then, 30-50 µL aliquots were streaked onto a solid medium that used methyl-a-D-glucopyranoside as the only carbon source (Robinson et al., 2016; Sniegowski et al., 2002). For each environmental sample that yielded-like colonies on solid plates, we selected one single isolated colony for DNA extraction using the Promega Wizard® Genomic DNA purification kit using the manufacturer's protocol for yeast, with only 75 units of lyticase (Sigma) as described by Robinson et al. (2016). DNA was amplified by PCR with ITS primers specific to *Saccharomycetales* (ITSf1 and ITSt; Robinson et al., 2016), and PCR products were cleaned using ExoSAP-IT™. Cleaned PCR products were sent for Sanger sequencing at GENEWIZ (Azenta Life Sciences) to generate rDNA sequences for yeast species identification.

For each strain, trace files were downloaded and converted to fastq files using a Python script cli.py from the abi2fastq Python Package Index (Biopython version 1.81).

The forward and reverse fastq file quality were checked for base sequence quality using FastQC (version 0.11.9 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

High-quality forward and reverse sequences with a quality score over 30 were merged into a single fastq file per strain. Fastq files were converted to high-quality fasta files with a phred score over 40 using seqtk seq -q 40 (version 1.3) (Li et al., 2009), and low-quality scores were converted to lowercase n. We used NCBI BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to confirm species identity based on sequence similarity to known yeast species. Each DNA sequence was queried against the nucleotide database restricted to Ascomycota (txid4890) using the blastn algorithm (BLAST+ version 2.12.0). BLAST searches were restricted to ten descriptions and alignments, and low-complexity sequence data was not masked. Yeast isolates had over 98% BLAST hits to a single species and/or to multiple synonyms, and we identified these yeast species as such (N = 485 strains; Table S3). Instances when species matched to multiple descriptions, were cases for species complexes with multiple synonyms, such as the *Meyerozyma guillermondii* complex that tend to have matches with *Candida carpophilia* and *M. caribbica*, which have been identified to be the same species (Kurtzman & Suzuki, 2010). Other common species synonyms were for *S. paradoxus*, where *S. cariocanus* matched multiple queries and *S. cariocanus* is the same species as *S. paradoxus* (Alsammar & Delneri, 2020). There are 14 out of 502 isolates (3%) that have pending species identification that are expected to be completed in April 2025. The isolates that are still pending include: Thompson Mills Forest 2020: 3 isolates; University of Georgia 2021: 1 isolate; Skidaway Institute of Oceanography 2020: 2 isolates; Skidaway Institute of Oceanography 2021: 6 isolates; and Skidaway Institute of

Oceanography 2022: 2 isolates. The bark samples associated with the 14 unidentified yeast isolates were excluded from all analyses.

3.2.4 Testing for associations between yeast prevalence and ecological or climatic factors

All statistical and graphical analyses were performed in R (version 4.2.2), and geographical maps were drawn using ggplot2 (version 3.4.4), maps (version 3.4.0), and raster packages (version 3.6-20). For each environmental sample, the temperature of the warmest month ($^{\circ}\text{C}$) and precipitation of the warmest quarter (mm) were included from WorldClim 2.1 (<https://www.worldclim.org/>) at 30 arc-second resolutions. For environmental samples from trees, we estimated the tree age if there were a known growth factor to account for different growth rates among tree species (<https://purduelandscapereport.org/article/how-old-is-my-tree/> and <https://www.datawhistory.org/dataw-island-history/dathas-grande-dames/>). Tree girth (m) was converted to inches (0.0254 inches per meter) to calculate the diameter (tree circumference * π) and then multiplied by the tree growth factor to obtain a crude estimate of tree age. Most of the environmental samples were collected from the bark of *Quercus alba* and *Q. virginiana* in 2021 across all eight field sites ($N = 418$ bark samples). Therefore, all downstream statistical analyses were performed from these tree hosts to model yeast frequency. We used a Bonferroni correction ($\alpha = 0.0167$) because the analysis was repeated on the three most common yeast species in the USA: *Saccharomyces cerevisiae*, *S. paradoxus*, and *Lachancea fermentati*. Each statistical analysis started with the full maximal model and then simplified by subtracting terms

starting with interactions down to the main effects stepwise to obtain the minimal adequate model. We tested whether this resulted in a worse model using chi-squared tests at each stepwise term subtraction and if generalized linear models were overdispersed we corrected using quasibinomial and used F-tests for stepwise model simplification tests recommended by Crawley (2014). Statistical models were checked for overdispersion using a simulation-based approach using the R package, DHARMA (version 0.4.6) (Hartig et al., 2024), which makes residuals for fitted generalized linear models readily interpretable.

For each species, we tested the probability of yeast isolation differences among tree hosts to see if there are effects of bark sample position (from the north or south side of the tree) or sample weight (grams), using a binary logistic regression with binomial errors to model the presence or absence of yeast for each bark sample. The full model included three explanatory variables: the effect of each tree host (102 levels), the cardinal direction of the sampled bark piece (2 levels: north and south), bark sample weight as a continuous variable, and the two-way interaction between the tree host and sample position.

We found no differences in yeast isolation frequency of bark samples from cardinal directions (north and south) of trees (see results). For *S. cerevisiae* and *S. paradoxus*, we combined this with a European dataset from Robinson et al. (2016). This data includes bark samples from *Quercus* spp. from the United Kingdom, France, and Greece. We excluded two field sites from the United Kingdom (Davenport Vineyard and Plumpton College) because these oak trees were sampled within vineyards. Yeast strains

from these sites are likely feral because they are associated with human activity (Robinson et al., 2016). We tested the probability of yeast isolation differences among tree hosts and bark sample weight (grams) using a binary logistic regression with binomial errors to model the presence or absence of yeast for each sample. The full maximal model included the two explanatory variables: sample weight as a continuous variable and tree host (194 levels) and the two-way interaction between sample weight and tree host.

Logistic regressions for the effect of bark sample weight and cardinal direction where the bark sample was collected were not good predictors for the presence or absence of yeast (see results). Therefore, we pooled multiple bark samples collected for each tree host to use the proportion of yeast isolates as a response variable for each host (Table S4). Generalized linear models (GLM) with binomial errors were modeled using a two-vector proportion response variable: the number of bark samples with the study species and the number of bark samples without that species. The full maximal model included four explanatory variables and all their interactions: maximum temperature ($^{\circ}\text{C}$) of the warmest month (Tmax) as a continuous variable, precipitation of the warmest quarter (mm) as a continuous variable, tree age (year) as a continuous variable, and region as a categorical variable with four levels (northern Europe, southern Europe, northern United States, and southern United States). We used the same north-south split of field sites from Robinson et al. (2016), where UK sites were in northern Europe, and French and Greek sites were in southern Europe. For field sites from the United States, the north-south split for the region explanatory variable was determined based on Tmax and latitude, where sites with a Tmax under $32\text{ }^{\circ}\text{C}$ were classified as the ‘northern United

States' (field sites: TY, DF, BNF, and TMF) and sites with a Tmax over 32 °C were classified as the 'southern United States' (field sites: UGA, SIO, HH, and FSP). Depending on yeast species, the region was further simplified to three levels (northern Europe, southern Europe, and the United States). Nested models were similarly compared using chi-squared and F tests (overdispersed models only).

3.2.5 Worldwide presence and absence data for wild yeast species

In cases where model predictions suggested a thermal optimum, we used that predictive optimal to draw a climate envelope map to compare this species range to observations of previously reported isolates from published studies. Wild yeast strains isolated from vineyards or from fruit, flowers, and insects were not included because these habitats are either associated with human activity or are transient habitats (Günther & Goddard, 2019; Hyma & Fay, 2013; Tilakaratna & Bensasson, 2017) and could be sink populations (Robinson et al., 2016), which make correlative models less accurate (Araújo & Peterson, 2012). I used the same metadata compiled by Robinson et al. (2016) and updated this resource by adding large-scale field surveys published since 2016 (Table S5) using the same data compilation methods and criteria. I added worldwide information for *L. fermentati*, a species not included in Robinson et al. (2016). For studies that did not have latitude and longitude information, I used the same search term criteria in Google Maps described by Robinson et al. (2016). In addition, I did not include data for *S. cerevisiae*, *S. paradoxus*, and *L. fermentati* strains from this study to test predicted Tmax optimum ranges. Furthermore, I visualized the geographic regions showing optimum

temperatures together with information on wild forest *S. cerevisiae* lineages from Europe and North America (Peña et al., 2025).

3.3 Results

3.3.1 A broad range of yeast species from North American forests and marsh

Our extensive field survey included 860 samples from angiosperms (tree bark, flowers, and fig), 79 samples from gymnosperms (tree bark and exudate), 15 samples from salt marsh sediment, three samples from Basidiomycota fruiting bodies, and two samples from arthropods (Figure 1 and Table S1). This yielded over 450 yeast strains, with most isolates belonging to the Saccharomycotina subphylum from seven different families and clades: Dipodascaceae/Trichomonascaceae, Pichiaceae, CUG-Scer1 clade, Phaffomycetaceae, Saccharomycodaceae, and Saaccharomycetaceae isolated from marsh sediment, a fiddler crab (Arthropoda), *Rhododendron* sp., *Echinacea purpurea*, *Acer rubrum*, *Magnoliaceae* spp., *Pinus* spp., and *Quercus* spp. (Table 3.1 and Table S3.3). Mushrooms, gymnosperm exudate, flowers from *Elephantopus tomentosus*, and one insect environmental sample did not yield any yeast isolates (Table S3.3). Across forest environments, *S. cerevisiae*, *S. paradoxus*, *L. fermentati*, and *L. thermotolerans* had the highest frequency (Figure 3.2) and there were no obvious differences among different tree species (Table 3.1).

Interestingly, this is the first field survey to recover *S. paradoxus*, a presumed cold temperate forest yeast (Mozzachiodi et al., 2022), from hot and wet forests; from cypress swamps in the Everglades (field site FSP) and central Florida (field site HH; Figure 3.2). Similar to a previous study that used the same culturing methods in Europe

(Robinson et al., 2016), we have isolated the following species that were also isolated: *L. thermotolerans*, *Wickerhamomyces anomalus*, *Pichia manshurica*, *Kluyveromyces lactis*, and *Zygosaccharomyces bailii*, but also recovered different yeast species. For instance, we have isolated over a hundred *L. fermentati* isolates with the highest frequency from Florida (field sites HH and FSP) (Figure 3.2) whereas European forests did not yield any *L. fermentati* isolates (Robinson et al., 2016). Furthermore, as in Robinson et al. (2016), we isolated human pathogenic yeast, but we did not isolate *Candida albicans*, but instead isolated *Nakaseomyces glabrata* (*Candida glabrata*) and *Candida parapsilosis* both from *Quercus* spp. (Table 3.1 and Table S3.3) and both listed as high priority pathogens by the World Health Organization (*WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action*, 2022). Other notable yeast species include two different fission yeast species that are not part of the Saccharomycotina subphylum: *Schizosaccharomyces japonicus* from *Pinus taeda* (Loblolly pine) and *Schizosaccharomyces pombe* from *Quercus virginiana* (live oak), that are commonly used in biotechnology. Lastly, in comparison between plant habitats and salt marsh sediment, *S. cerevisiae* and *S. paradoxus* were less frequent on marsh sediment (1 isolate each) whereas *Meyerozyma guillermondii* was relatively more abundant on marsh sediment compared to bark from *Quercus* spp. (5 isolates out 7 isolates from marsh sediment and 19 isolates out of 433 isolates from bark, Fisher's exact test, P-value = 6.17×10^{-6} ; Table 3.1).

3.3.2 Yeast prevalence on bark samples and sampling strategy

For *S. cerevisiae*, *L. fermentati*, and *S. paradoxus*, the three most common species in the USA, I used statistical modeling to test for associations between yeast presence on bark from the USA survey and bark position, weight or tree host. The effect of bark position (north vs. south side of a tree) was not a good predictor for the absence or presence of yeast isolation for any yeast species (GLM: d.f. = 1, *P*-value > 0.05). This is unsurprising because, when we used iButtons to record temperatures at Thompson Mills Forest for five trees over one year, we did not observe consistent differences between the north and south sides of trees. (Figure S3.1).

Field site was the only predictor for the absence or presence of *S. cerevisiae* (GLM, 13% deviance, d.f. = 7, *P*-value = 1.5×10^{-9}); there were no significant differences among trees within field sites (GLM, deviance = 24%, d.f. = 94, *P*-value = 0.3). The field site in southern Georgia (SIO) had the lowest isolation frequency (Figure 3.2). For *L. fermentati*, the tree host was the best predictor for the absence or presence of yeast (GLM, 43% deviance, d.f. = 101, *P*-value = 7.1×10^{-5}). Although field sites are associated with *L. fermentati* isolation rates (GLM, 9% deviance, d.f. = 7, *P*-value = 1.3×10^{-5}), there were significant differences among tree hosts within sites (deviance = 129.6, d.f. = 94, *P*-value = 0.009) that are probably partially explained by tree age at some sites (treeAge:fieldSite, deviance = 16.9, d.f. = 7, *P*-value = 0.018). Tree age was the only predictor for the absence or presence of *S. paradoxus* (GLM, 4% deviance, d.f. = 1, *P*-value = 0.004).

To determine whether the presence or absence of yeast varies by year, we collected bark samples from *Q. alba* (white oak) and *Q. virginiana* (live oak) across multiple years (2019-2022) within Georgia (TMF, UGA, and SIO) (Figure 3.3). For *S. cerevisiae* and *S. paradoxus*, sampling year was not a good predictor for the absence or presence of these species after simplifying the maximal model that included all explanatory variables: year, field site, and the two-way interaction between year and field site. Field site was the only predictor for the absence or presence of *S. cerevisiae* and *S. paradoxus* (*S. cerevisiae*: GLM, 25% deviance, d.f. = 2, P-value < 2×10^{-16} and *S. paradoxus*: GLM, 4% deviance, d.f. = 2, P-value = 0.022). To properly test for an effect of sampling year in *S. paradoxus* however, a future model will compare isolation frequency while accounting for all differences among trees including tree age. In contrast, there were significant differences among sampling years for the absence or presence of *L. fermentati* (GLM, 4% deviance, d.f. = 2, P-value = 0.0032). When we visualized the proportion of yeast isolates across all sampling years within each site (Figure 3.3), *S. cerevisiae* and *S. paradoxus* isolation rates do appear similar among sampling years, consistent with the final GLMs. Also consistent with the models, *L. fermentati* did appear more abundant in southern Georgia in 2021 than in other years; weather could affect the relative abundance of this species. These models are not final, however, because at the time of writing there were a total of 17 isolates with pending species identification.

Initially, we did not combine the results of the USA with published results because the European data (Robinson et al., 2016) did not include bark position. Since bark position did not have significant differences for the presence or absence of *S. cerevisiae* and *S. paradoxus* isolation, we combined this with European forest data

bringing the total to 196 tree hosts. The maximal model included all explanatory variables: tree host, bark sample weight, and a two-way interaction between tree host and bark sample weight to test if sample weight affects the absence or presence of yeast. The tree host was further simplified down to field site, tree age (years), and the two-way interaction between tree age and field site. Field site was the only predictor for *S. cerevisiae* isolation frequency in the final GLM (34% deviance, d.f. = 18, P-value < 2.2×10^{-16}). For *S. paradoxus*, the final GLM had the two-way interaction between tree age and field site (GLM, 20% deviance), and dropping the interaction resulted in a worse model (GLM, 4% deviance, d.f. = 16, P-value = 0.0082).

3.3.3 *Saccharomyces cerevisiae* prevalence is associated with summer temperature

Climate envelope models of *S. cerevisiae* (Robinson et al., 2016) and laboratory thermal growth profiles suggest that *S. cerevisiae* could be prevalent in environments at temperatures of 30°C or greater. We hypothesize that temperature (°C) of the warmest month (summer temperature; Tmax) in conjunction with precipitation (mm) of the warmest quarter, tree age (years), and region (northern and southern Europe, and northern and southern United States) could predict the species distribution of wild *S. cerevisiae*. The maximal model included these explanatory variables and all their interactions. The final GLM explained 49% deviance among trees where the additive effects of Tmax (d.f. = 1, P-value = 3.3×10^{-4}) and region (d.f. = 3, P-value = 2.5×10^{-5}) were the only important predictors for the isolation frequency of *S. cerevisiae* (Figure 3.4), but a precipitation effect might be detected with more statistical power (d.f. = 1, P-value = 0.051). For the European sample, model predictions and the lack of isolates suggest that it was too cold

for *S. cerevisiae* isolation. In the northern United States, there appears to be a positive correlation; *S. cerevisiae* isolation frequency increases with increasing Tmax, but isolation rates appear slightly reduced in the hotter south (Figure 3.4).

This led me to hypothesize that the optimal thermal range for wild *S. cerevisiae* is between 29°C and 33°C with a lower limit at 26°C and an upper limit at 36°C based on Tmax for the United States. To test this predicted Tmax optimum range, we used published strains from extensive field surveys of wild *S. cerevisiae* isolates associated with plants (e.g., tree bark, leaves, litter under the tree) (Almeida et al., 2015; Barbosa et al., 2016; Cromie et al., 2013; Diezmann & Dietrich, 2009; Duan et al., 2018; Fay et al., 2019; Han et al., 2021; Kuehne et al., 2007; T. J. Lee et al., 2022; Naumov et al., 1997; Peter et al., 2018; Pontes et al., 2020; Robinson et al., 2016; Sniegowski et al., 2002; Song et al., 2015; Spurley et al., 2022; Sylvester et al., 2015; Wang et al., 2012; Zhang et al., 2010) to map reported isolates onto a climate envelope map of the predicted Tmax (Figure 3.5A). We included absence data from extensive field surveys that collected over 100 tree bark samples that did not yield any *S. cerevisiae* isolates (Charron et al., 2014; Johnson et al., 2004; Kowallik et al., 2015; Robinson et al., 2016). Globally, most strains fall within the predicted range between 29°C and 33°C, with a few strains that appear to be occurring at the edges of the species range limit near Russia and North America. To further test the predicted Tmax optimum range, I mapped the distribution of wild forest lineages from Europe and North America (Peña et al., 2025) and found that forest lineages fall within the predicted Tmax range (Figure 3.5B and C). In North America, the American wild and North American A occur at cooler temperatures (29°C) in eastern North America. Lineages North American B and C are more broadly distributed across

the optimum Tmax range, with a single North American B isolate from Michigan at its edge (28.2°C) and a few North American C strains from Florida at the upper end (33°C). In Europe, most strains from all lineages fall within the predicted optimum Tmax range, except for a few strains from France, Slovenia, and Hungary that occur at the lower limit of the range (26°C). Together, these results suggest that optimal summer temperature can predict the prevalence of wild *S. cerevisiae* and that the optimal thermal niche is between 29°C and 33°C.

3.3.4 *Lachancea fermentati* prevalence is associated with summer temperature and precipitation

Once again, we used a maximal model of all the explanatory variables and all their interactions and stepwise model simplification previously described (see above) to determine which ecological and climatic factors could predict the thermal optimum of wild *L. fermentati*. The forest European data did not yield any *L. fermentati* isolates (Robinson et al., 2016), and subsequent analyses were performed only in the United States. The final GLM explained 19% deviance among trees and only the additive effects of Tmax (d.f. = 1, P-value = 0.0016), region (d.f. = 1, P-value = 7.3x10⁻⁵), and precipitation (d.f. = 1, P-value = 0.0003) were important predictors for the isolation frequency of *L. fermentati*. The model showed that *L. fermentati* isolation frequency was higher in the northern USA and increases with increasing Tmax and precipitation (Figure 3.6).

Based on these results, we could not predict the optimal summer range for Tmax and precipitation because even at lower quartiles, the model predicts a positive

correlation (Figure 3.6). Instead, we visualized the range of Tmax (29-34°C) and precipitation (312-590 mm) across our sampling range in the United States (Figure 3.7) and mapped this climate range onto Europe to see if there could be suitable areas for *L. fermentati*. We mapped published strains from extensive field surveys that have isolated wild *L. fermentati* associated with plants to see if these reported isolates fall within our sampling Tmax and precipitation range (Spurley et al., 2022; Sylvester et al., 2015). We included absence data for extensive field surveys that collected over 100 tree bark samples that did not yield any *L. fermentati* isolates (Charron et al., 2014; Johnson et al., 2004; Kowallik et al., 2015; Robinson et al., 2016). Only a few isolates from the upper Midwest fall outside our sampling range. Past sampling of Europe did not appear to have overlapping Tmax and precipitation ranges suitable for *L. fermentati* because summer temperatures were either too cool (United Kingdom and northern Germany) or too dry (Greece). Indeed, there are no parts of Europe with similar climatic conditions to those we sampled in the USA. (Figure 3.7).

3.3.5 *Saccharomyces paradoxus* prevalence is associated with temperature, precipitation, and tree age

Laboratory thermal growth profiles suggest that *S. paradoxus* can grow at a broad range of temperatures, between 10-37°C (Leducq et al., 2014; Sweeney et al., 2004), but has a thermal optimum range between 22-28°C according to models of *S. paradoxus* isolation frequency from European forests (Robinson et al., 2016). Given the optimum temperature range, we did not expect to find *S. paradoxus* from the southern United States; especially not from Florida, where the Tmax was at 33.3°C. Using these new data

together with European data, we still proceeded with our analyses to determine if climatic and ecological variables could explain the broad thermal growth range for *S. paradoxus*. The final GLM explained 35% deviance among trees where the two-way interactions between (i) temperature and precipitation (d.f. = 1, P-value = 1.1×10^{-6}), (ii) precipitation and continent (d.f. = 2, P-value = 1.7×10^{-6}), and (iii) continent and tree age (d.f. = 2, P-value = 1.9×10^{-5}) and their additive effects were important predictors for the isolation frequency of *S. paradoxus* after simplifying northern and southern United States into a single factor. Tree age was an important predictor across all continents, where the isolation frequency of *S. paradoxus* was prevalent among older trees, but this effect was more substantial in Europe (Figure 3.8).

When we examined the predicted probability of isolating *S. paradoxus* separated by region with the combined effect of Tmax and precipitation, we used the overall tree age median (102.5 years) and overall quartiles for Tmax and precipitation (Figure 3.9). In northern Europe, most trees were sampled in regions with intermediate precipitation, and the model does not predict a correlation with Tmax at the median or the upper quartile for precipitation (Figure 3.9A green and grey lines). In the United States, trees sampled were only at the median and upper quartiles for precipitation and there was not a clear correlation at these quartiles between isolation frequency and Tmax (Figure 3.9A). In southern Europe, most trees sampled were in the driest quartile, and the model predicts a negative correlation between *S. paradoxus* frequency and Tmax under those conditions. When examining the effect of precipitation on *S. paradoxus* isolation frequency, in Europe and the United States there is no obvious correlation (Figure 3.9B). In southern Europe, where most trees were at intermediate temperatures (grey), the predicted

isolation frequency increased with increasing precipitation (Figure 3.9B grey line) and there were not any strains isolated at the lower and upper Tmax quartiles.

For *S. paradoxus*, I did not predict an optimal summer range for Tmax and precipitation because the associations appeared different between Europe and the USA. During the model simplification in preliminary analyses, there was a significant three-way interaction between tree age, Tmax, and precipitation (d.f. = 1, P-value = 0.0007), which could further explain differences among regions because this three-way interaction implies that the two-way interaction between Tmax and precipitation is correlated with tree age, suggesting that the relationships presented here may be oversimplified.

3.4 Discussion

Here, we have isolated a broad range of wild yeast species, mainly from oaks, with the aim of testing climate associations in the most prevalent species. Using our methods, the most common species in the United States were *S. cerevisiae* *L. fermentati* and *S. paradoxus* and all showed climate associations. For *S. cerevisiae* and *S. paradoxus*, we combined our data with those from European forests to examine continental differences between North America and Europe. We discovered that summer temperature is a predictor of the prevalence of all species examined, and precipitation was also associated with *S. paradoxus* and *L. fermentati* prevalence. Interestingly, tree age was only correlated with *S. paradoxus* prevalence and showed a more substantial effect in Europe. Furthermore, model predictions for *S. cerevisiae* from the United States suggested a thermal optimum between 29 and 33°C. This optimum falls within 30 and 37°C estimates from a comparison of laboratory thermal growth profiles for *S. cerevisiae* and *S.*

paradoxus (Robinson et al., 2016; Sweeney et al., 2004). Using previously reported wild *S. cerevisiae* associated with plants and the location of wild genomic forest lineages, we found that most isolates and the distribution of wild lineages fall within the predicted range. This suggests that climate affects the distribution of budding yeast in the wild and probably that of other fungal microbes. Climate associations such as the ones reported here could help us predict how future climate change may affect species ranges.

3.4.1 Salt marshes have a different species composition compared to forests

Within forests, *S. cerevisiae* and *S. paradoxus* were prevalent on both angiosperms (e.g. oaks) and gymnosperms (loblolly pines; Table 3.1), which is consistent with other studies that isolated a broad range of yeast species from a variety of plant families (Lee et al., 2022; Spurley et al., 2022; Sylvester et al., 2015). However, salt marsh sediment had a different yeast species composition. *Meyerozyma guillermondii* was more abundant in sediments compared with other yeasts (Table 3.1). Furthermore, oak trees sampled near salt marshes (field site SIO) had the lowest yeast isolation frequency (Figure 3.2), suggesting that some yeast species may not tolerate salty environments. This could be especially true for wild *S. cerevisiae*, which had the lowest isolation frequency on the Georgia coast compared to all other sites where trees were not within a few meters of salt water (Figure 3.2). Even when sampling SIO trees across multiple years (2020-2022), the relative abundance of wild *S. cerevisiae* was low (Figure 3.3), suggesting that this effect was not due to weather fluctuations.

The accumulation of salt water is toxic to many terrestrial plants, but not to salt tolerant grasses and seepweed (Gurevitch et al., 2020) and, therefore, could be toxic for

some yeast species living on trees near marshes. It has been hypothesized that salt marshes could be the wild ecological niche for human pathogenic yeasts such as *Candida auris* (Arora et al., 2021; Casadevall et al., 2019). In this study, we have isolated the human pathogenic yeast *N. glabrata* from a live oak tree near salt marshes and *M. guillermondii* is often a commensal on human skin. Therefore, the findings of this study may lend support to the idea that the salt marsh environment may contain more yeast species able to colonize humans.

3.4.2 Climate and the species ranges of budding yeast from forests

There is mounting evidence that the species ranges of *Saccharomyces* yeasts differ among species and are limited by climate (Leducq et al., 2014; Nespolo et al., 2020; Robinson et al., 2016). We also see differing associations between *S. cerevisiae*, *S. paradoxus* and *L. fermentati* among climate variables (Figures 3.4, 3.6 and 3.9). For instance, our statistical models suggested that for *S. cerevisiae* there could be a thermal optimum between the northernmost and southernmost ends of our sampling range in the United States (Figure 3.4) and a possible association with precipitation. While *L. fermentati* occurred more frequently at wet sampling sites and showed no clear thermal optimum whereas *S. cerevisiae* overlapped in between with the only species that had a predicted thermal optimum. The climate model for *S. paradoxus* showed associations with temperature, precipitation and tree age, but it was difficult to interpret in part because the associations were significantly different between North America and Europe. There is evidence that one lineage of *S. paradoxus* adapted to high temperatures in the eastern United States and Canada; this southern lineage is locally adapted to warm

temperatures and freeze-thaw cycles (Leducq et al., 2014). For *S. paradoxus*, differences in thermal optima may contribute to the complexity of the associations we observe. As in other studies we did not find significant differences among sampling years (Charron et al., 2014; Robinson et al., 2016). There are many variables that were not included in this study that might also affect yeast isolation frequencies, such as humidity and temperature seasonality.

Wild *L. fermentati* might be endemic to the United States or it could be restricted to hot and wet climates not seen in Europe (Figure 3.6). This is consistent with field surveys from European forests that yielded no *L. fermentati* isolates (Robinson et al., 2016). Wild yeast endemism is not necessarily new and may be common in *Saccharomyces* yeasts (Alsammar & Delneri, 2020; Boynton & Greig, 2014; Mozzachiodi et al., 2022). *Lachancea fermentati* is similar to *S. cerevisiae* in that it can be isolated from both wild (e.g., tree bark) and fermented (e.g., wine) environments. However, the ecological niche in both environments is still unclear (Mozzachiodi et al., 2022). Precipitation may be a vital climate association structuring wild *L. fermentati* populations and could explain why it has not been isolated from European forests. Our results suggest that European forests are either too cold in the summer or too dry (Figure 3.6), although this could differ for strains associated with fermented food and beverages. The species distribution of wild *L. fermentati* needs further investigation by sampling in tropical rain forests (e.g., Central or South America, or southeast Asia) and in cooler regions past our latitudinal sampling range in North America (e.g., southern Canada) to determine whether *L. fermentati* is endemic to the United States and to provide a better understanding of ecological niches for non-model yeasts.

3.4.3 The ecology of *S. paradoxus* is complicated

In the case of *S. paradoxus*, this species produced the most complex statistical model. Our analysis showed that tree age was important for predicting the prevalence of *S. paradoxus* but not for *S. cerevisiae* and *L. fermentati*. A possible explanation is microbial succession, where older trees have had more time to be colonized by yeast or other microbes. This was first proposed by Robinson et al. (2016); they hypothesized that this process is similar to belowground microbial processes. A positive tree age correlation with increased isolation frequency of *S. paradoxus* was stronger for European forests. Sampled European trees had a broader range of tree ages (9 to 530 years old) compared to those sampled in the United States (most trees over 100 years old), perhaps explaining the stronger positive correlation in Europe. However, does this positive relationship with tree age still hold true for sampled trees that are over 100 years old? When we removed European trees from analyses, tree age was still an important predictor for the prevalence of *S. paradoxus* in the United States. *Saccharomyces paradoxus* could be more sensitive to forest succession and climate because our analyses showed that tree age could alter associations with temperature and precipitation (significant three-way interaction) that needs to be further teased apart. These climate associations are likely correlated, and a species distribution model that considers all climatic factors (bioclimatic variables from WorldClim2.1) using MaxEnt or other methods could identify other climate variables that could be associated with isolation frequency.

3.5 References

- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J.-L., Serra, M., Dequin, S., Couloux, A., Guy, J., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular Ecology*, 24(21), 5412–5427. <https://doi.org/10.1111/mec.13341>
- Alsammar, H., & Delneri, D. (2020). An update on the diversity, ecology and biogeography of the *Saccharomyces* genus. *FEMS Yeast Research*, 20(3). <https://doi.org/10.1093/femsyr/foaa013>
- Araújo, M. B., Anderson, R. P., Barbosa, A. M., Beale, C. M., Dormann, C. F., Early, R., Garcia, R. A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R. B., Zimmermann, N. E., & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858. <https://doi.org/10.1126/sciadv.aat4858>
- Araújo, M. B., & Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology*, 93(7), 1527–1539. <https://doi.org/10.1890/11-1930.1>
- Arora, P., Singh, P., Wang, Y., Yadav, A., Pawar, K., Singh, A., Padmavati, G., Xu, J., & Chowdhary, A. (2021). Environmental Isolation of *Candida auris* from the Coastal Wetlands of Andaman Islands, India. *mBio*, 12(2), e03181-20. <https://doi.org/10.1128/mbio.03181-20>
- Bahram, M., & Netherway, T. (2021). Fungi as mediators linking organisms and ecosystems. *FEMS Microbiology Reviews*, 46(2), fuab058-. <https://doi.org/10.1093/femsre/fuab058>
- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of Natural Hybridization in Brazilian Wild Lineages of *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, 8(2), 317–329. <https://doi.org/10.1093/gbe/evv263>
- Bensasson, D., Dicks, J., Ludwig, J. M., Bond, C. J., Elliston, A., Roberts, I. N., & James, S. A. (2019). Diverse Lineages of *Candida albicans* Live on Old Oaks. *Genetics*, 211(1), 277–288. <https://doi.org/10.1534/genetics.118.301482>
- Boekhout, T., Amend, A. S., Baidouri, F. E., Gabaldón, T., Geml, J., Mittelbach, M., Robert, V., Tan, C. S., Turchetti, B., Vu, D., Wang, Q.-M., & Yurkov, A. (2022). Trends in yeast diversity discovery. *Fungal Diversity*, 114(1), 491–537. <https://doi.org/10.1007/s13225-021-00494-6>

- Boult, V. L., & Evans, L. C. (2021). Mechanisms matter: Predicting the ecological impacts of global change. *Global Change Biology*.
<https://doi.org/10.1111/gcb.15527>
- Boynton, P. J., & Greig, D. (2014). The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast*, 31(12), 449–462. <https://doi.org/10.1002/yea.3040>
- Buckley, L. B., & Kingsolver, J. G. (2019). Environmental variability shapes evolution, plasticity and biogeographic responses to climate change. *Global Ecology and Biogeography*, 28(10), 1456–1468. <https://doi.org/10.1111/geb.12953>
- Buzzini, P., Lachance, M. A., & Yurkov, A. (Eds.). (2017). *Yeasts in Natural Ecosystems: Diversity*. Springer International Publishing.
<https://doi.org/10.1007/978-3-319-62683-3>
- Carroll, S. P., Hendry, A. P., Reznick, D. N., & Fox, C. W. (2007). Evolution on ecological time-scales. *Functional Ecology*, 21(3), 387–393.
<https://doi.org/10.1111/j.1365-2435.2007.01289.x>
- Casadevall, A., Kontoyiannis, D. P., & Robert, V. (2019). On the Emergence of *Candida auris*: Climate Change, Azoles, Swamps, and Birds. *mBio*, 10(4), 10.1128/mbio.01397-19. <https://doi.org/10.1128/mbio.01397-19>
- Case, N. T., Berman, J., Blehert, D. S., Cramer, R. A., Cuomo, C., Currie, C. R., Ene, I. V., Fisher, M. C., Fritz-Laylin, L. K., Gerstein, A. C., Glass, N. L., Gow, N. A. R., Gurr, S. J., Hittinger, C. T., Hohl, T. M., Iliev, I. D., James, T. Y., Jin, H., Klein, B. S., ... Cowen, L. E. (2022). The future of fungi: Threats and opportunities. *G3*, 12. <https://doi.org/10.1093/g3journal/jkac224>
- Cavalieri, D., Valentini, B., & Stefanini, I. (2022). Going wild: Ecology and genomics are crucial to understand yeast evolution. *Current Opinion in Genetics & Development*, 75, 101922. <https://doi.org/10.1016/j.gde.2022.101922>
- Charron, G., Leducq, J.-B., Bertin, C., Dubé, A. K., & Landry, C. R. (2014). Exploring the northern limit of the distribution of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* in North America. *FEMS Yeast Research*, 14(2), 281–288. <https://doi.org/10.1111/1567-1364.12100>
- Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., Huang, A. A., Dudley, A. M., & Fay, J. C. (2013). Genomic Sequence Diversity and Population Structure of *Saccharomyces cerevisiae* Assessed by RAD-seq. *G3: Genes/Genomes/Genetics*, 3(12), 2163–2171.
<https://doi.org/10.1534/g3.113.007492>

- 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. <https://doi.org/10.1371/journal.pone.0005317>
- Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., & Bai, F.-Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nature Communications*, 9(1), 2690. <https://doi.org/10.1038/s41467-018-05106-7>
- Dujon, B. A., & Louis, E. J. (2017). Genome Diversity and Evolution in the Budding Yeasts (Saccharomycotina). *Genetics*, 206(2), 717–750. <https://doi.org/10.1534/genetics.116.199216>
- Fay, J. C., Liu, P., Ong, G. T., Dunham, M. J., Cromie, G. A., Jeffery, E. W., Ludlow, C. L., & Dudley, A. M. (2019). A polyploid admixed origin of beer yeasts derived from European and Asian wine populations. *PLOS Biology*, 17(3), e3000147. <https://doi.org/10.1371/journal.pbio.3000147>
- Fussmann, G. F., Loreau, M., & Abrams, P. A. (2007). Eco-evolutionary dynamics of communities and ecosystems. *Functional Ecology*, 21(3), 465–477. <https://doi.org/10.1111/j.1365-2435.2007.01275.x>
- Gavin, D. G., Fitzpatrick, M. C., Gugger, P. F., Heath, K. D., Rodríguez-Sánchez, F., Dobrowski, S. Z., Hampe, A., Hu, F. S., Ashcroft, M. B., Bartlein, P. J., Blois, J. L., Carstens, B. C., Davis, E. B., Lafontaine, G. de, Edwards, M. E., Fernandez, M., Henne, P. D., Herring, E. M., Holden, Z. A., ... Williams, J. W. (2014). Climate refugia: Joint inference from fossil records, species distribution models and phylogeography. *New Phytologist*, 204(1), 37–54. <https://doi.org/10.1111/nph.12929>
- Gonçalves, P., Valério, E., Correia, C., Almeida, J. M. G. C. F. de, & Sampaio, J. P. (2011). Evidence for Divergent Evolution of Growth Temperature Preference in Sympatric *Saccharomyces* Species. *PLoS ONE*, 6(6), e20739. <https://doi.org/10.1371/journal.pone.0020739>
- Günther, C. S., & Goddard, M. R. (2019). Do yeasts and *Drosophila* interact just by chance? *Fungal Ecology*, 38, 37–43. <https://doi.org/10.1016/j.funeco.2018.04.005>
- Han, D.-Y., Han, P.-J., Rumbold, K., Koricha, A. D., Duan, S.-F., Song, L., Shi, J.-Y., Li, K., Wang, Q.-M., & Bai, F.-Y. (2021). Adaptive Gene Content and Allele Distribution Variations in the Wild and Domesticated Populations of *Saccharomyces cerevisiae*. *Frontiers in Microbiology*, 12, 631250. <https://doi.org/10.3389/fmicb.2021.631250>
- Harrison, M., Opulente, D. A., Wolters, J. F., Shen, X., Zhou, X., Groenewald, M., Hittinger, C. T., Rokas, A., & LaBella, A. L. (2024). Exploring Saccharomycotina

Yeast Ecology Through an Ecological Ontology Framework. *Yeast*.
<https://doi.org/10.1002/yea.3981>

Hartig, F., Lohse, L., & leite, M. de S. (2024). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (Version 0.4.7) [Computer software]. <https://cran.r-project.org/web/packages/DHARMa/index.html>

Hendry, A. P. (2016). *Eco-evolutionary Dynamics*. Princeton University Press.

Hyma, K. E., & Fay, J. C. (2013). Mixing of vineyard and oak-tree ecotypes of *Saccharomyces cerevisiae* in North American vineyards. *Molecular Ecology*, 22(11), 2917–2930. <https://doi.org/10.1111/mec.12155>

Jessica Gurevitch; Samuel M. Scheiner; Gordon A. Fox. (2020). *The Ecology of Plants* (3rd ed.). Oxford University Press. <https://global.oup.com/ushe/product/the-ecology-of-plants-9781605358291>

Johnson, L. J., Koufopanou, V., Goddard, M. R., Hetherington, R., Schäfer, S. M., & Burt, A. (2004). Population Genetics of the Wild Yeast *Saccharomyces paradoxus*. *Genetics*, 166(1), 43–52. <https://doi.org/10.1534/genetics.166.1.43>

Kleinhesselink, A. R., & Adler, P. B. (2018). The response of big sagebrush (*Artemisia tridentata*) to interannual climate variation changes across its range. *Ecology*, 99(5), 1139–1149. <https://doi.org/10.1002/ecy.2191>

Kowalllik, V., Miller, E., & Greig, D. (2015). The interaction of *Saccharomyces paradoxus* with its natural competitors on oak bark. *Molecular Ecology*, 24(7), 1596–1610. <https://doi.org/10.1111/mec.13120>

Kuehne, H. A., Murphy, H. A., Francis, C. A., & Sniegowski, P. D. (2007). Allopatric Divergence, Secondary Contact, and Genetic Isolation in Wild Yeast Populations. *Current Biology*, 17(5), 407–411. <https://doi.org/10.1016/j.cub.2006.12.047>

Kurtzman, C. P., & Suzuki, M. (2010). Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera *Babjeviella*, *Meyerozyma*, *Millerozyma*, *Priceomyces*, and *Scheffersomyces*. *Mycoscience*, 51(1), 2–14. <https://doi.org/10.1007/S10267-009-0011-5>

Lachance, M.-A. (2003). The Phaff school of yeast ecology. *International Microbiology*, 6(3), 163–167. <https://doi.org/10.1007/s10123-003-0129-9>

Lafontaine, G. de, Napier, J. D., Petit, R. J., & Hu, F. S. (2018). Invoking adaptation to decipher the genetic legacy of past climate change. *Ecology*, 99(7), 1530–1546. <https://doi.org/10.1002/ecy.2382>

- Langdon, Q. K., Peris, D., Eizaguirre, J. I., Opulente, D. A., Buh, K. V., Sylvester, K., Jarzyna, M., Rodríguez, M. E., Lopes, C. A., Libkind, D., & Hittinger, C. T. (2020). Postglacial migration shaped the genomic diversity and global distribution of the wild ancestor of lager-brewing hybrids. *PLoS Genetics*, 16(4), e1008680. <https://doi.org/10.1371/journal.pgen.1008680>
- Leducq, J.-B., Charron, G., Samani, P., Dubé, A. K., Sylvester, K., James, B., Almeida, P., Sampaio, J. P., Hittinger, C. T., Bell, G., & Landry, C. R. (2014). Local climatic adaptation in a widespread microorganism. *Proceedings of the Royal Society B: Biological Sciences*, 281(1777), 20132472. <https://doi.org/10.1098/rspb.2013.2472>
- Lee, K. M., & Coop, G. (2019). Population genomics perspectives on convergent adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1777), 20180236. <https://doi.org/10.1098/rstb.2018.0236>
- Lee, T. J., Liu, Y.-C., Liu, W.-A., Lin, Y.-F., Lee, H.-H., Ke, H.-M., Huang, J.-P., Lu, M.-Y. J., Hsieh, C.-L., Chung, K.-F., Liti, G., & Tsai, I. J. (2022). Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages. *Genome Research*, 32(5), 864–877. <https://doi.org/10.1101/gr.276286.121>
- Legras, J.-L., Galeote, V., Bigey, F., Camarasa, C., Marsit, S., Nidelet, T., Sanchez, I., Couloux, A., Guy, J., Franco-Duarte, R., Marcet-Houben, M., Gabaldon, T., Schuller, D., Sampaio, J. P., & Dequin, S. (2018). Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. *Molecular Biology and Evolution*, 35(7), 1712–1727. <https://doi.org/10.1093/molbev/msy066>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lofgren, L. A., & Stajich, J. E. (2021). Fungal biodiversity and conservation mycology in light of new technology, big data, and changing attitudes. *Current Biology*, 31(19), R1312–R1325. <https://doi.org/10.1016/j.cub.2021.06.083>
- McGaughran, A., Laver, R., & Fraser, C. (2021). Evolutionary Responses to Warming. *Trends in Ecology & Evolution*. <https://doi.org/10.1016/j.tree.2021.02.014>
- Micheal J Crawley. (2014). *Statistics: An Introduction Using R, 2nd Edition* (Second). John Wiley & Sons.

- Moritz, C., & Agudo, R. (2013). The Future of Species Under Climate Change: Resilience or Decline? *Science*, 341(6145), 504–508.
<https://doi.org/10.1126/science.1237190>
- Mozzachiodi, S., Bai, F., Baldrian, P., Bell, G., Boundy-Mills, K., Buzzini, P., Čadež, N., Cubillos, F. A., Dashko, S., Dimitrov, R., Fisher, K. J., Gibson, B., Gouliamova, D., Greig, D., Heistinger, L., Hittinger, C. T., Jecmenica, M., Koufopanou, V., Landry, C. R., ... Boynton, P. (2022). Yeasts from temperate forests. *Yeast*.
<https://doi.org/10.1002/yea.3699>
- Naumov, G. I., Naumova, E. S., & Sniegowski, P. D. (1997). Differentiation of European and Far East Asian populations of *Saccharomyces paradoxus* by allozyme analysis. *International Journal of Systematic Bacteriology*, 47(2), 341–344.
<https://doi.org/10.1099/00207713-47-2-341>
- 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.
- Nespolo, R. F., Villarroel, C. A., Oporto, C. I., Tapia, S. M., Vega-Macaya, F., Urbina, K., Chiara, M. D., Mozzachiodi, S., Mikhalev, E., Thompson, D., Larrondo, L. F., Saenz-Agudelo, P., Liti, G., & Cubillos, F. A. (2020). An Out-of-Patagonia migration explains the worldwide diversity and distribution of *Saccharomyces eubayanus* lineages. *PLOS Genetics*, 16(5), e1008777.
<https://doi.org/10.1371/journal.pgen.1008777>
- Nnadi, N. E., & Carter, D. A. (2021). Climate change and the emergence of fungal pathogens. *PLoS Pathogens*, 17(4), e1009503.
<https://doi.org/10.1371/journal.ppat.1009503>
- Opulente, D. A., LaBella, A. L., Harrison, M.-C., Wolters, J. F., Liu, C., Li, Y., Kominek, J., Steenwyk, J. L., Stoneman, H. R., VanDenAvond, J., Miller, C. R., Langdon, Q. K., Silva, M., Gonçalves, C., Ubbelohde, E. J., Li, Y., Buh, K. V., Jarzyna, M., Haase, M. A. B., ... Hittinger, C. T. (2024). Genomic factors shape carbon and nitrogen metabolic niche breadth across Saccharomycotina yeasts. *Science*, 384(6694), eadj4503. <https://doi.org/10.1126/science.adj4503>
- Opulente, D. A., Langdon, Q. K., Buh, K. V., Haase, M. A. B., Sylvester, K., Moriarty, R. V., Jarzyna, M., Considine, S. L., Schneider, R. M., & Hittinger, C. T. (2019). Pathogenic budding yeasts isolated outside of clinical settings. *FEMS Yeast Research*, 19(3). <https://doi.org/10.1093/femsyr/foz032>
- Osland, M. J., Stevens, P. W., Lamont, M. M., Brusca, R. C., Hart, K. M., Waddle, J. H., Langtimm, C. A., Williams, C. M., Keim, B. D., Terando, A. J., Reyier, E. A., Marshall, K. E., Loik, M. E., Boucek, R. E., Lewis, A. B., & Seminoff, J. A. (2021). Tropicalization of temperate ecosystems in North America: The

northward range expansion of tropical organisms in response to warming winter temperatures. *Global Change Biology*. <https://doi.org/10.1111/gcb.15563>

Pacifci, M., Foden, W. B., Visconti, P., Watson, J. E. M., Butchart, S. H. M., Kovacs, K. M., Scheffers, B. R., Hole, D. G., Martin, T. G., Akçakaya, H. R., Corlett, R. T., Huntley, B., Bickford, D., Carr, J. A., Hoffmann, A. A., Midgley, G. F., Pearce-Kelly, P., Pearson, R. G., Williams, S. E., ... Rondinini, C. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, 5(3), Article 3. <https://doi.org/10.1038/nclimate2448>

Peay, K. G., Kennedy, P. G., & Talbot, J. M. (2016). Dimensions of biodiversity in the Earth mycobiome. *Nature Reviews Microbiology*, 14(7), Article 7. <https://doi.org/10.1038/nrmicro.2016.59>

Pelletier, F., Garant, D., & Hendry, A. P. (2009). Eco-evolutionary dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1523), 1483–1489. <https://doi.org/10.1098/rstb.2009.0027>

Peña, J. J., Scopel, E. F. C., Ward, A. K., & Bensasson, D. (2025). Footprints of Human Migration in the Population Structure of Wild Baker's Yeast. *Molecular Ecology*, e17669(n/a), e17669. <https://doi.org/10.1111/mec.17669>

Peris, D., Ubbelohde, E. J., Kuang, M. C., Kominek, J., Langdon, Q. K., Adams, M., Koshalek, J. A., Hulfachor, A. B., Opulente, D. A., Hall, D. J., Hyma, K., Fay, J. C., Leducq, J.-B., Charron, G., Landry, C. R., Libkind, D., Gonçalves, C., Gonçalves, P., Sampaio, J. P., ... Hittinger, C. T. (2023). Macroevolutionary diversity of traits and genomes in the model yeast genus *Saccharomyces*. *Nature Communications*, 14(1), 690. <https://doi.org/10.1038/s41467-023-36139-2>

Peter, J., Chiara, M. D., Friedrich, A., Yue, J.-X., Pfleiger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), 339. <https://doi.org/10.1038/s41586-018-0030-5>

Pontes, A., Hutzler, M., Brito, P. H., & Sampaio, J. P. (2020). Revisiting the Taxonomic Synonyms and Populations of *Saccharomyces cerevisiae*—Phylogeny, Phenotypes, Ecology and Domestication. *Microorganisms*, 8(6), Article 6. <https://doi.org/10.3390/microorganisms8060903>

Robinson, H. A., Pinharanda, A., & Bensasson, D. (2016). Summer temperature can predict the distribution of wild yeast populations. *Ecology and Evolution*, 6(4), 1236–1250. <https://doi.org/10.1002/ece3.1919>

Rodríguez-Verdugo, A., Buckley, J., & Stapley, J. (2017). The genomic basis of eco-evolutionary dynamics. *Molecular Ecology*. <https://doi.org/10.1111/mec.14045>

- Sage, R. F. (2019). Global change biology: A primer. *Global Change Biology*, 26(1), 3–30. <https://doi.org/10.1111/gcb.14893>
- Salvadó, Z., Arroyo-López, F. N., Guillamón, J. M., Salazar, G., Querol, A., & Barrio, E. (2011b). Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. *Applied and Environmental Microbiology*, 77(7), 2292–2302. <https://doi.org/10.1128/aem.01861-10>
- 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. <https://doi.org/10.1128/AEM.02396-07>
- Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and Ecology of Species Range Limits. *Annual Review of Ecology, Evolution, and Systematics*, 40, 415–436.
- Shen, X.-X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Čadež, N., Libkind, D., ... Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, 175(6), 1533–1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>
- 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. <https://doi.org/10.1111/j.1567-1364.2002.tb00048.x>
- Song, G., Dickins, B. J. A., Demeter, J., Engel, S., Dunn, B., & Cherry, J. M. (2015). AGAPE (Automated Genome Analysis PipelinE) for Pan-Genome Analysis of *Saccharomyces cerevisiae*. *PLOS ONE*, 10(3), e0120671. <https://doi.org/10.1371/journal.pone.0120671>
- Spurley, W. J., Fisher, K. J., Langdon, Q. K., Buh, K. V., Jarzyna, M., Haase, M. A. B., Sylvester, K., Moriarty, R. V., Rodriguez, D., Sheddan, A., Wright, S., Sorlie, L., Hulfachor, A. B., Opulente, D. A., & Hittinger, C. T. (2022). Substrate, temperature, and geographical patterns among nearly 2000 natural yeast isolates. *Yeast*, 39(1–2), 55–68. <https://doi.org/10.1002/yea.3679>
- Stajich, J. E. (2017). Fungal Genomes and Insights into the Evolution of the Kingdom. *Microbiology Spectrum*, 5(4). <https://doi.org/10.1128/microbiolspec.funk-0055-2016>

- 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. [https://doi.org/10.1016/S1567-1356\(03\)00171-5](https://doi.org/10.1016/S1567-1356(03)00171-5)
- Sylvester, K., Wang, Q.-M., James, B., Mendez, R., Hulfachor, A. B., & Hittinger, C. T. (2015). Temperature and host preferences drive the diversification of *Saccharomyces* and other yeasts: A survey and the discovery of eight new yeast species. *FEMS Yeast Research*, 15(3). <https://doi.org/10.1093/femsyr/fov002>
- Taylor, J. W., Turner, E., Townsend, J. P., Dettman, J. R., & Jacobson, D. (2006). Eukaryotic microbes, species recognition and the geographic limits of species: Examples from the kingdom Fungi. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1475), 1947–1963. <https://doi.org/10.1098/rstb.2006.1923>
- Tilakaratna, V., & Bensasson, D. (2017). Habitat Predicts Levels of Genetic Admixture in *Saccharomyces cerevisiae*. *G3: Genes, Genomes, Genetics*, 7(9), 2919–2929. <https://doi.org/10.1534/g3.117.041806>
- Tsai, I. J. (2022). The teenage years of yeast population genomics trace history, admixing and getting wilder. *Current Opinion in Genetics & Development*, 75, 101942. <https://doi.org/10.1016/j.gde.2022.101942>
- Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A., & Bai, F.-Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular Ecology*, 21(22), 5404–5417. <https://doi.org/10.1111/j.1365-294x.2012.05732.x>
- WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action* (1st ed). (2022). World Health Organization.
- Willis, K. J. (2018). *State of the World's Fungi*. <https://stateoftheworldsfungi.org/>
- Zhang, H., Skelton, A., Gardner, R. C., & Goddard, M. R. (2010). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: Evidence for migration from Europe and interspecies hybrids. *FEMS Yeast Research*, 10(7), 941–947. <https://doi.org/10.1111/j.1567-1364.2010.00681.x>

3.6 Author contributions

The research was conceptualized and designed by Jacqueline J. Peña and Douda Bensasson. Jacqueline J. Peña, Douda Bensasson, Audrey Ward, Prince Duepa, Eduardo Scopel, Bran Celia-Sanchez, Diana Ambrocio, Linda Habershaw, Domenic Won, and

Rosemary Willis conducted fieldwork. Jacqueline J. Peña, Audrey Ward, Eduardo Scopel, Bran Celia-Sanchez, Diana Ambrocio, Linda Habersharn, Domenic Won, Miranda McKibben, Frema Owusu-Ansah, Oliver Nemeth, and Grady Waple performed yeast culturing and isolation, and DNA extractions. Miranda McKibben developed the bioinformatics pipeline for yeast identification. Audrey Ward, Diana Ambrocio, Linda Habersharn, Domenic Won, Miranda McKibben, Frema Owusu-Ansah, Oliver Nemeth, and Grady Waple conducted PCR for Sanger sequencing and yeast species identification using our bioinformatic pipelines.

Table 3.1 Yeast species isolated from different habitats.

Host	Strain(s)	Species
Marsh sediment	5 1	<i>Meyerozyma guillermondii</i> <i>Saccharomyces cerevisiae</i> and <i>Saccharomyces paradoxus</i>
<i>Arthropoda</i>	1	<i>Lachancea fermentati</i>
<i>Rhododendron sp.</i>	1	<i>Saccharomyces paradoxus</i>
<i>Echinacea purpurea</i>	1	<i>Zygosaccharomyces rouxii</i>
<i>Pinus spp.</i>	2 1	<i>Saccharomyces cerevisiae</i> <i>Kazachstania martiniae</i> , <i>Ogataea zsolti</i> , <i>Pichia manshurica</i> , <i>Saccharomyces paradoxus</i> , and <i>Schizosaccharomyces japonicus</i>
<i>Acer rubrum</i>	3 3 2 1	<i>Debaryomyces hansenii</i> <i>Saccharomyces cerevisiae</i> <i>Hanseniaspora vineae</i> <i>Candida insectorum</i> and <i>Lachancea fermentati</i>
<i>Magnoliaceae</i>	11 4 2 1	<i>Saccharomyces cerevisiae</i> <i>Lachancea fermentati</i> <i>Saccharomyces paradoxus</i> <i>Candida californica</i> , <i>Candida railensis</i> , <i>Meyerozyma guillermondii</i> , and <i>Pichia garciniae</i>
<i>Quercus spp.</i>	140 112 56 40 25 19 5 4 4 3 3 2 2 2 2 1	<i>Saccharomyces cerevisiae</i> <i>Lachancea fermentati</i> <i>Saccharomyces paradoxus</i> <i>Lachancea thermotolerans</i> <i>Pichia manshurica</i> <i>Meyerozyma guillermondii</i> <i>Kazachstania martiniae</i> <i>Wicherhamomyces anomalus</i> <i>Yarrowia lipolytica</i> <i>Debaryomyces hansenii</i> <i>Kluyveromyces lactis</i> <i>Torulaspora delbrueckii</i> <i>Candida glabrata</i> <i>Candida railensis</i> <i>Hyphopichia burtonii</i> <i>Zygosaccharomyces bailii</i> <i>Candida californica</i> , <i>Candida orthopsis</i> , <i>Candida parapsilosis</i> , <i>Candida silvanorum</i> , <i>Clavispora lusitaniae</i> , <i>Cyberlindnera fabianii</i> , <i>Lachancea quebecensis</i> , <i>Meyerozyma smithsonii</i> , <i>Pichia kudriavzevii</i> , <i>Pichia membranifaciens</i> , <i>Schizosaccharomyces pombe</i>

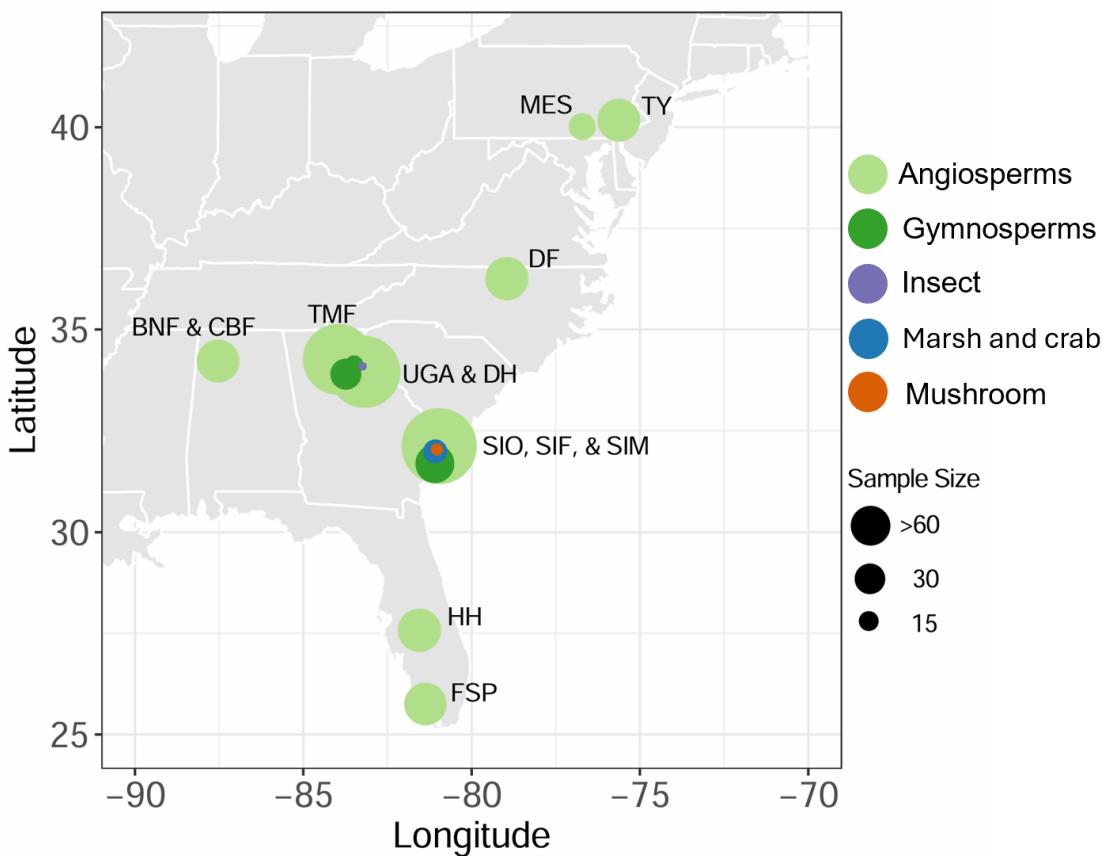


Figure 3.1 Environmental sample locations across the eastern United States from September 2019 to October 2022. Sample sizes are scaled by the square root, and each field site is abbreviated. Field site abbreviations include: TY (Tyler Arboretum), MES (Glen Rock), DF (Duke Forest), BNF (Bankhead National Forest), CBF (Athens, Alabama), TMF (Thompson Mills Forest), UGA (University of Georgia), DH (Athens, Georgia), SIO (Skidaway Institute of Oceanography), SIF (Skidaway Institute of Oceanography Flower), SIM (Skidaway Institute of Oceanography Marsh), HH (Highlands Hammock), and FSP (Fakahatchee Strand Preserve).

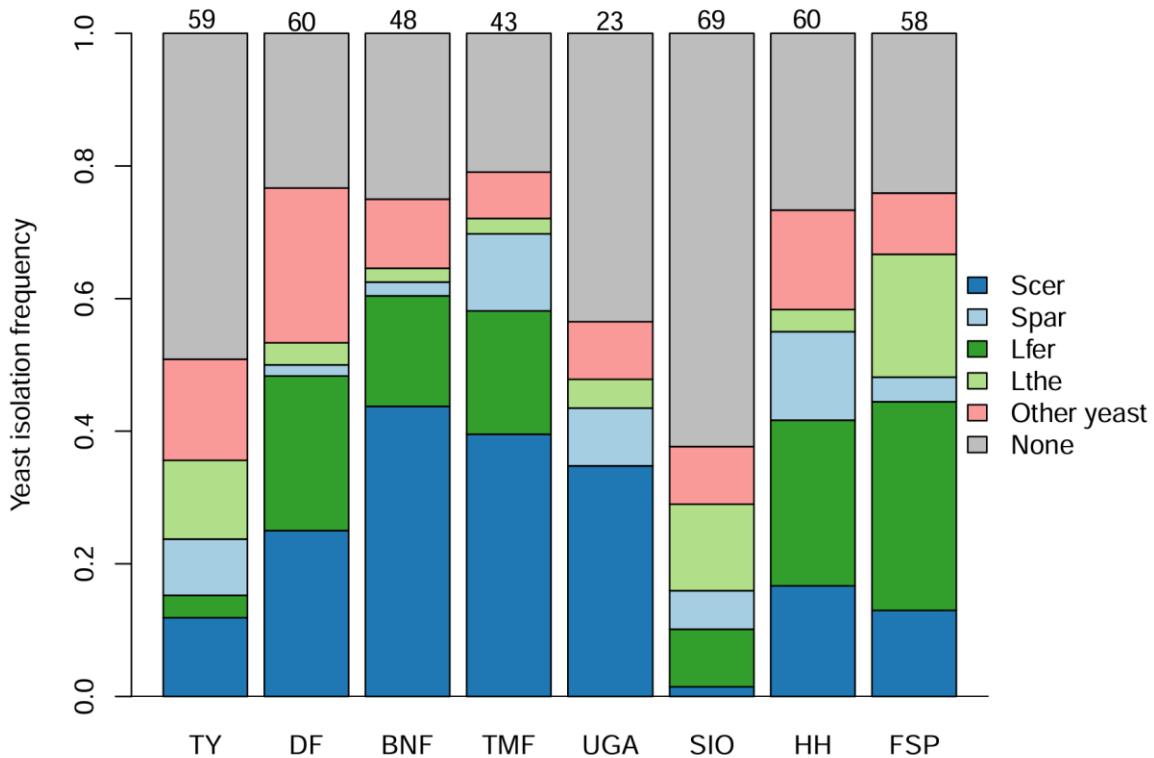


Figure 3.2 Yeast isolation frequency from *Quercus alba* and *Q. virginiana* across all field sites visited in 2021. Field sites are ordered by decreasing latitude (from left to right). The numbers above bars are the total number of bark samples collected. The most abundant yeast species isolated are color-coded and abbreviated as follows: Scer (*Saccharomyces cerevisiae*), Spar (*S. paradoxus*), Lfer (*Lachancea fermentati*), Lthe (*L. thermotolerans*), other yeast (least abundant yeast species amplified with ITS primers specific to *Saccharomycetales*), and bark samples not yielding any yeast isolates are classified as none. Field site abbreviations include: TY (Tyler Arboretum), MES (Glen Rock), DF (Duke Forest), BNF (Bankhead National Forest), CBF (Athens, Alabama), TMF (Thompson Mills Forest), UGA (University of Georgia), DH (Athens, Georgia), SIO (Skidaway Institute of Oceanography), SIF (Skidaway Institute of Oceanography Flower), SIM (Skidaway Institute of Oceanography Marsh), HH (Highlands Hammock), and FSP (Fakahatchee Strand Preserve).

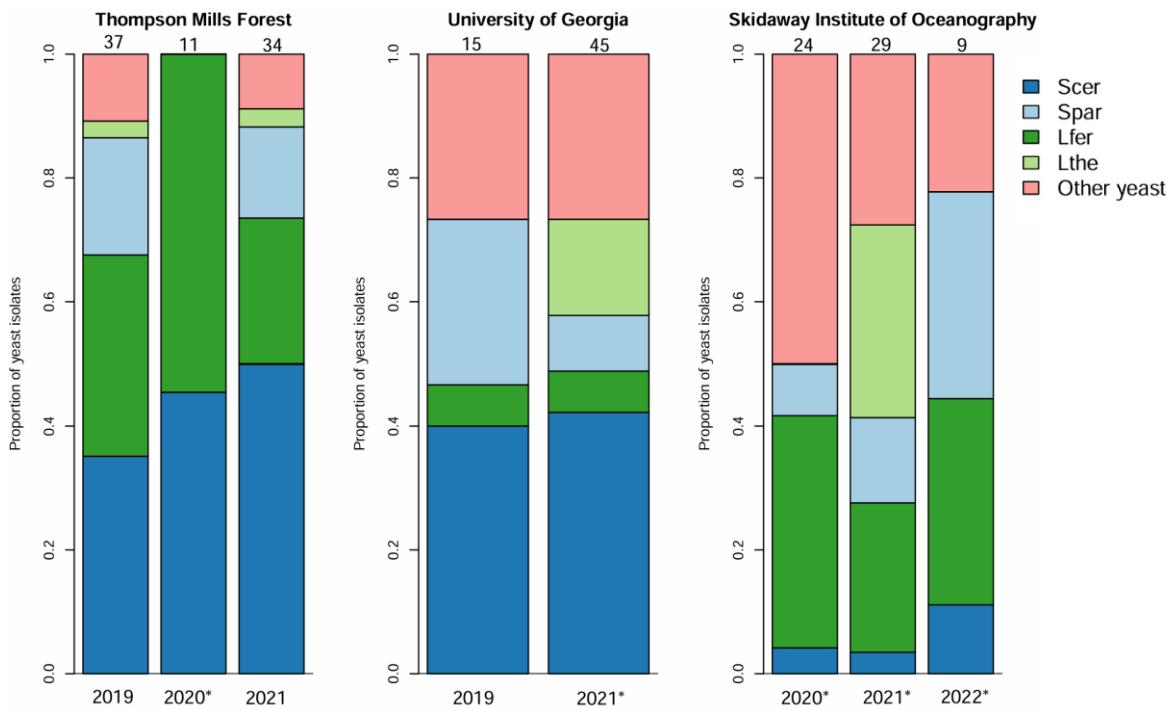


Figure 3.3 The proportion of yeast isolates from *Quercus alba* and *Q. virginiana* across multiple years in Georgia. Bar plots are ordered by decreasing latitude (from left to right). The numbers above bars are the total number of yeast strains isolated per year. Asterisks indicate isolates still pending for species identification: Thompson Mills Forest 2020: 3 isolates; University of Georgia 2021: 1 isolate; Skidaway Institute of Oceanography 2020: 2 isolates; Skidaway Institute of Oceanography 2021: 6 isolates; and Skidaway Institute of Oceanography 2022: 2 isolates. The most abundant yeast species isolated are color-coded and abbreviated as follows: Scer (*Saccharomyces cerevisiae*), Spar (*S. paradoxus*), Lfer (*Lachancea fermentati*), Lthe (*L. thermotolerans*), and other yeast (least abundant yeast species amplified with ITS primers specific to *Saccharomycetales*).

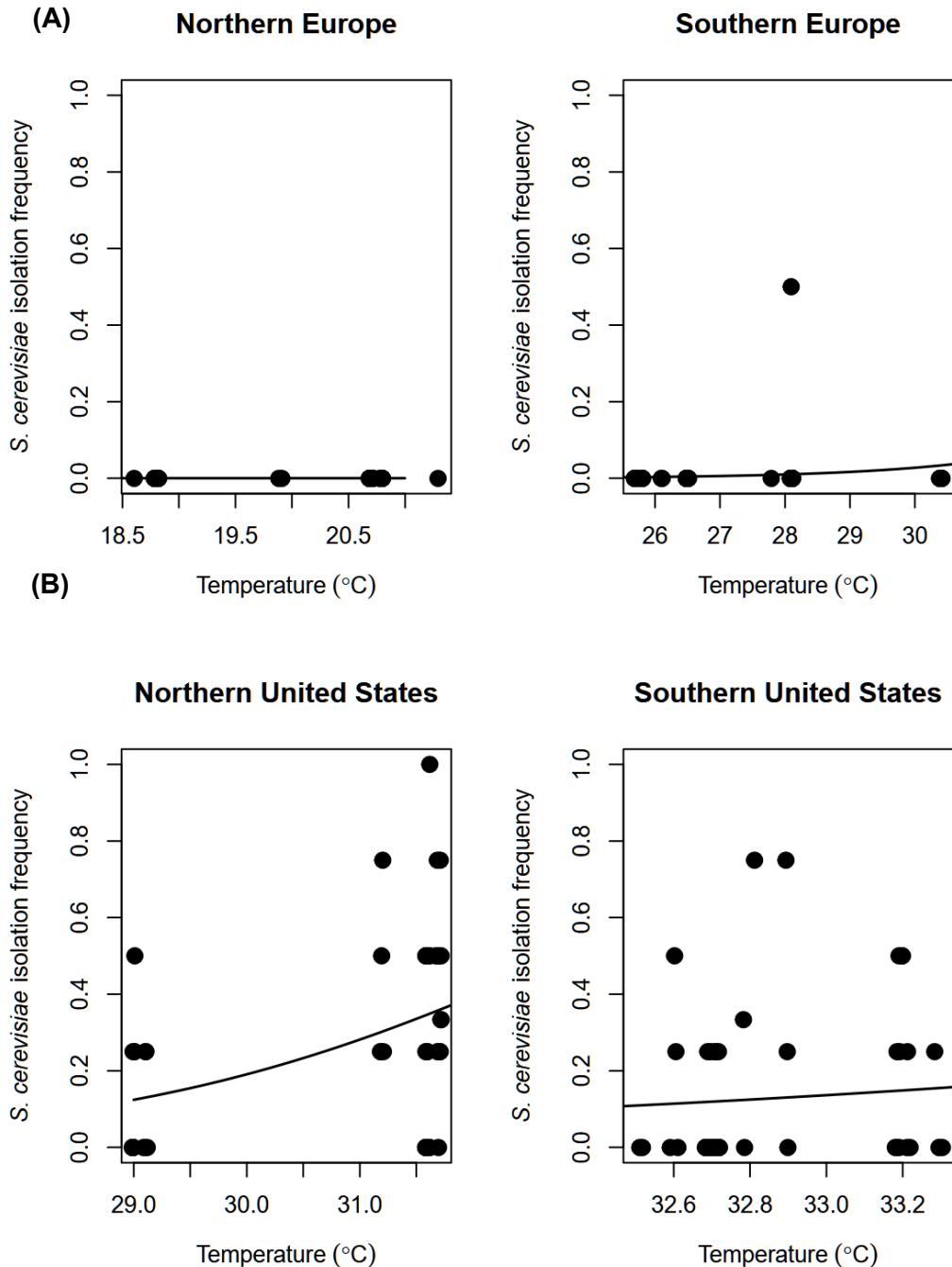


Figure 3.4 *Saccharomyces cerevisiae* isolation frequency increases with increasing temperature (°C) of the warmest month (Tmax) for 196 *Quercus* sp. from Europe (A) and eastern United States (B). Regions are color-coded by north-south splits (blue is for northern regions and red for southern regions). Lines show the predicted probabilities of isolating *S. cerevisiae* estimated from the final GLM (49% deviance). The final GLM showed a significant additive effect for Tmax and region *S. cerevisiae* isolation frequency (Tmax: d.f. = 1, P-value = 3.3×10^{-4} ; region: d.f. = 3, P-value = 2.5×10^{-5}).

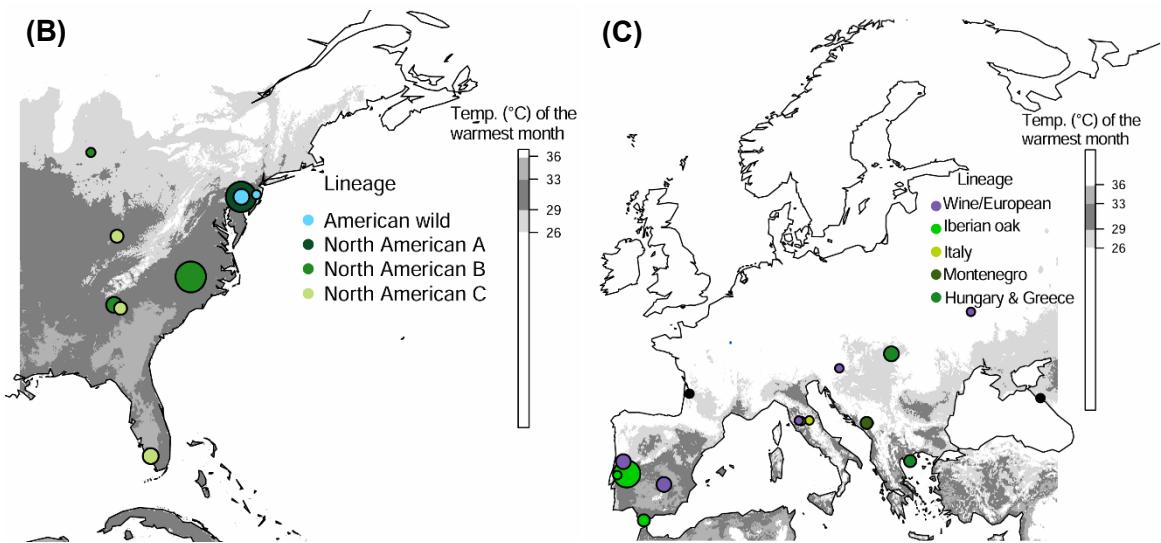
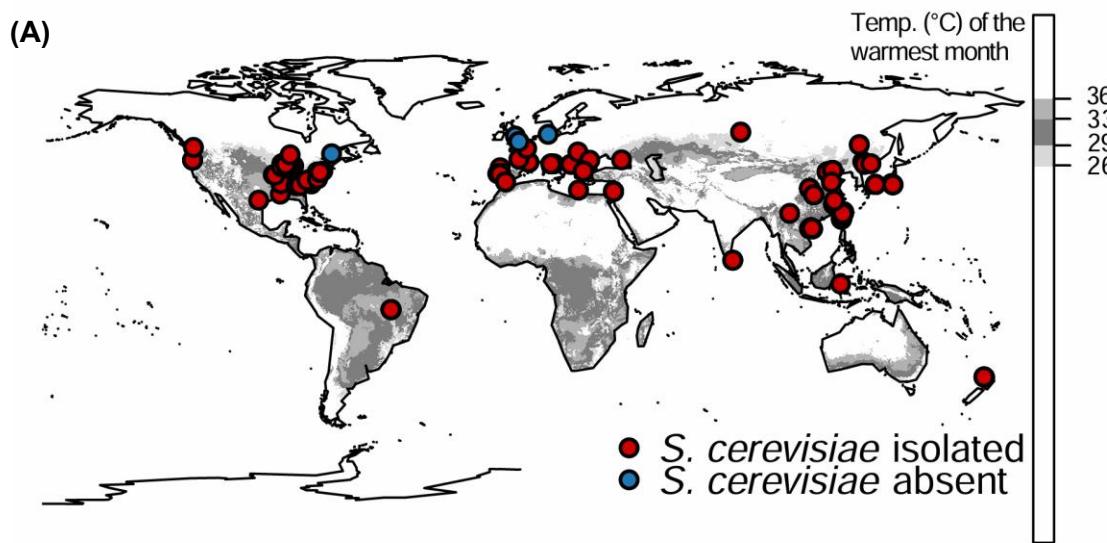


Figure 3.5 Optimal summer temperature ($^{\circ}\text{C}$) can predict the species range of wild *S. cerevisiae* associated with plants and the distribution of wild lineages. (A) Global distributions of strains from published studies. Red circles show the approximate GPS locations of published strains from field surveys (Almeida et al., 2015; Barbosa et al., 2016; Cromie et al., 2013; Diezmann & Dietrich, 2009; Duan et al., 2018; Fay et al., 2019; Han et al., 2021; Kuehne et al., 2007; T. J. Lee et al., 2022; Naumov et al., 1997; Peter et al., 2018; Pontes et al., 2020; Robinson et al., 2016; Sniegowski et al., 2002b; Song et al., 2015; Spurley et al., 2022; Sylvester et al., 2015; Wang et al., 2012; Zhang et al., 2010). Blue circles show GPS locations where field surveys collected over 100 tree bark samples but did not report any *S. cerevisiae* isolate (Charron et al., 2014; Johnson et al., 2004; Kowallik et al., 2015; Robinson et al., 2016). North American forest lineages (B) and European forest lineages defined by Peña et al. (2025) (C) fall within the predicted optimal summer temperature range. Circles are colored by lineage and scaled by the square root of transformed sample sizes.

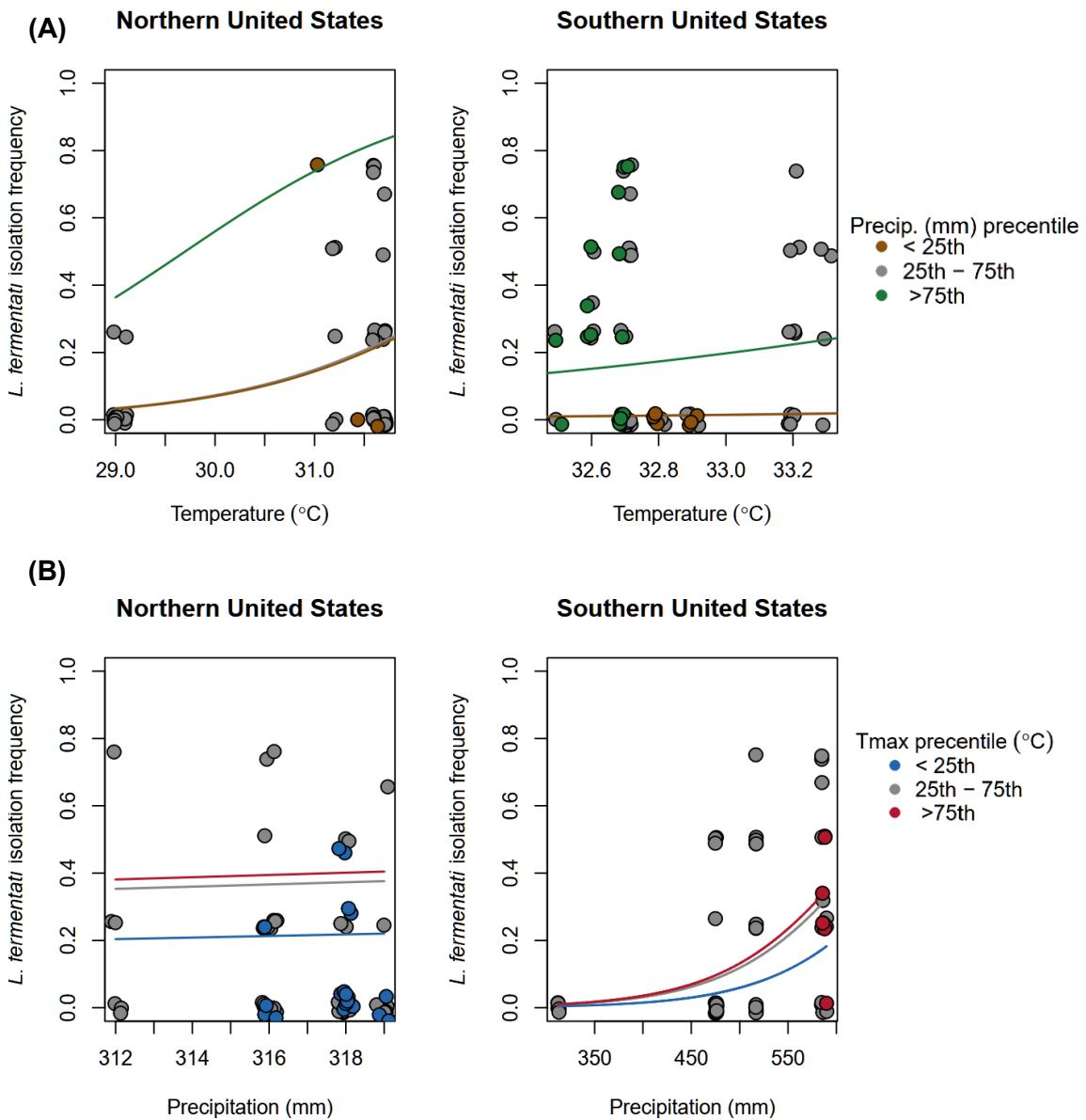


Figure 3.6 The combined effect of temperature and precipitation for the isolation frequency of *L. fermentati* isolation frequency for 102 *Quercus* sp. from northern and southern United States (GLM, 19% deviance). Across all regions, *L. fermentati* isolation frequency increases with increasing temperature ($^{\circ}\text{C}$) of the warmest month (Tmax) (A) and with increasing precipitation (mm) of the warmest quarter (B). Lines show the probability of isolating *L. fermentati* using the overall quartiles for precipitation (lower quartile = 316 mm, median = 319 mm, and upper quartile = 517 mm) and Tmax (lower quartile = 31.6 $^{\circ}\text{C}$, median = 32.55 $^{\circ}\text{C}$, and upper quartile = 32.7 $^{\circ}\text{C}$). Lines for the median (gray) and lower quartile (brown) for precipitation for northern and southern United States overlap.

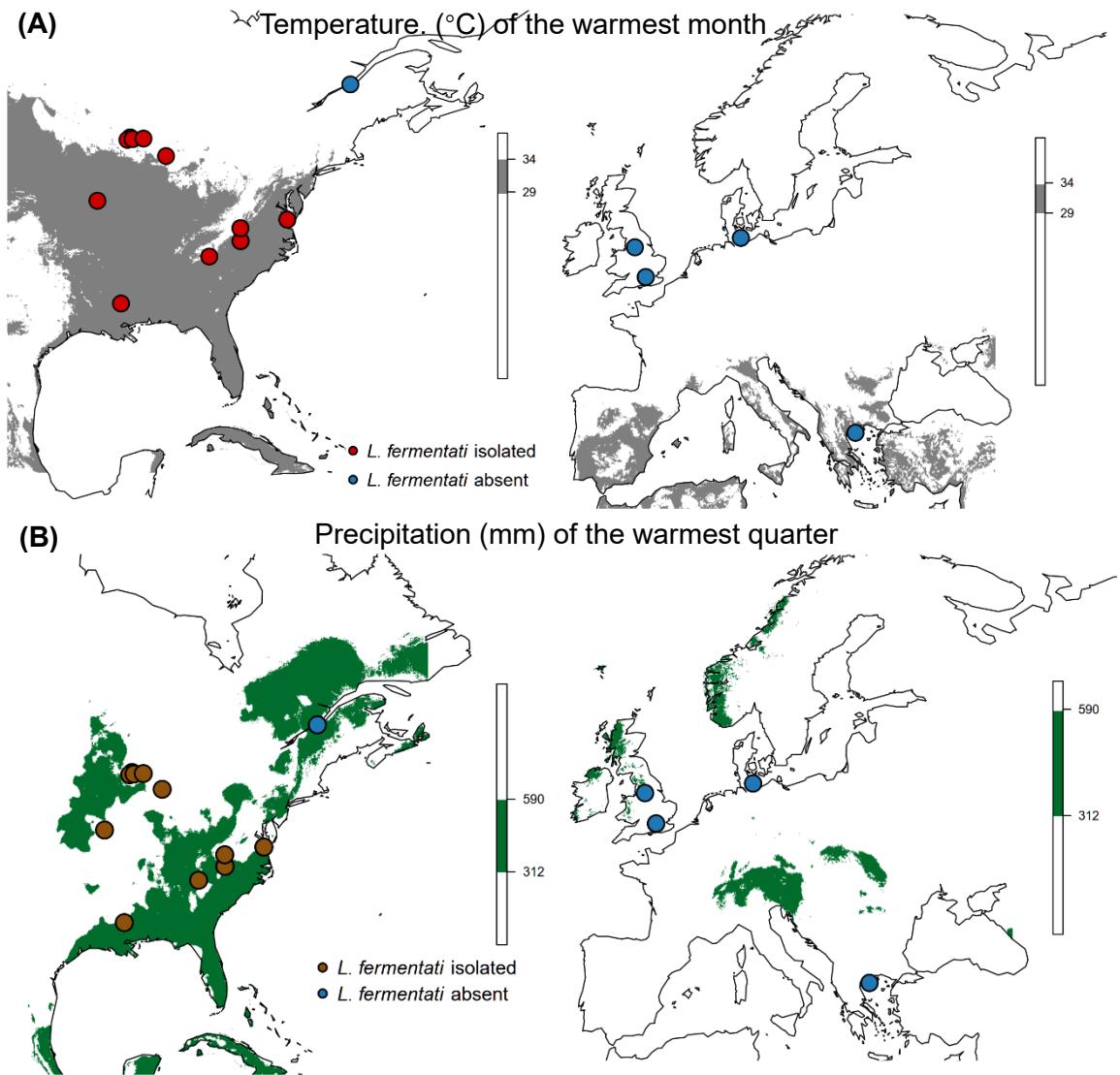


Figure 3.7 There are no regions of Europe with high summer temperature and precipitation that are similar to the regions we sampled in the United States. GPS locations of strains from extensive field surveys (Spurley et al., 2022; Sylvester et al., 2015) for temperature (A) and precipitation (B) in red and brown circles, respectively, in the United States. In Europe, summer temperatures may be too cool and dry for wild *L. fermentati*. Blue circles show GPS locations of extensive field surveys from Quebec, Canada (Charron et al., 2014), the United Kingdom (Johnson et al., 2004), northern Germany (Kowallik et al., 2015), and Greece (Robinson et al., 2016) that collected over 100 tree samples but did not yield any *L. fermentati* isolates.

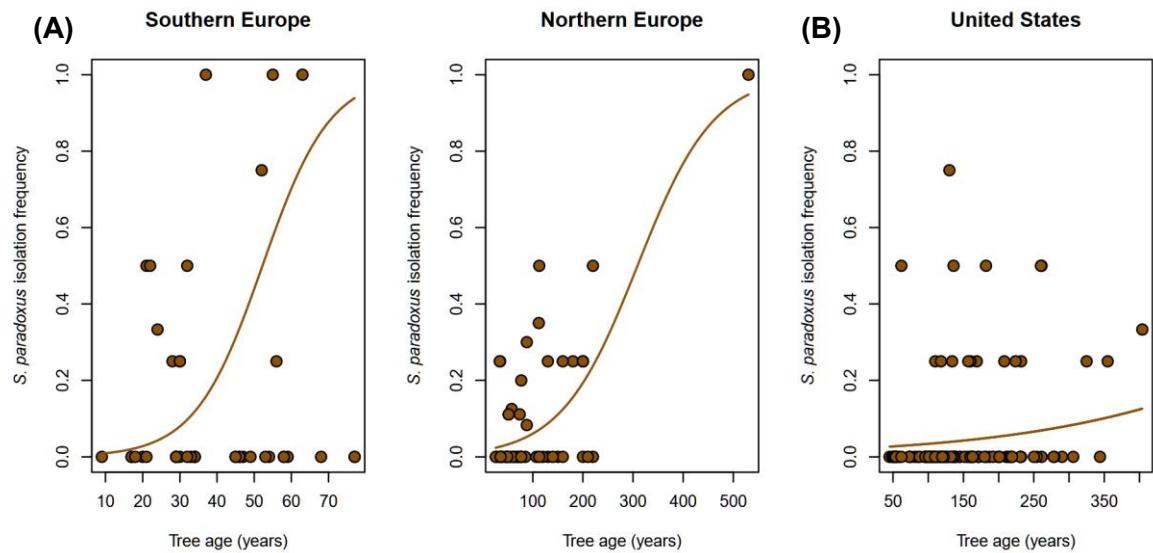


Figure 3.8 *Saccharomyces paradoxus* frequency increases on older trees (years) for 196 *Quercus* species, from southern and northern Europe (A) and the eastern United States (B). Lines show the probability of isolating *S. paradoxus* from the final GLM (35% deviance) using the overall median temperature (29.1°C) of the warmest month (Tmax) and the overall median precipitation (312 mm) of the warmest quarter.

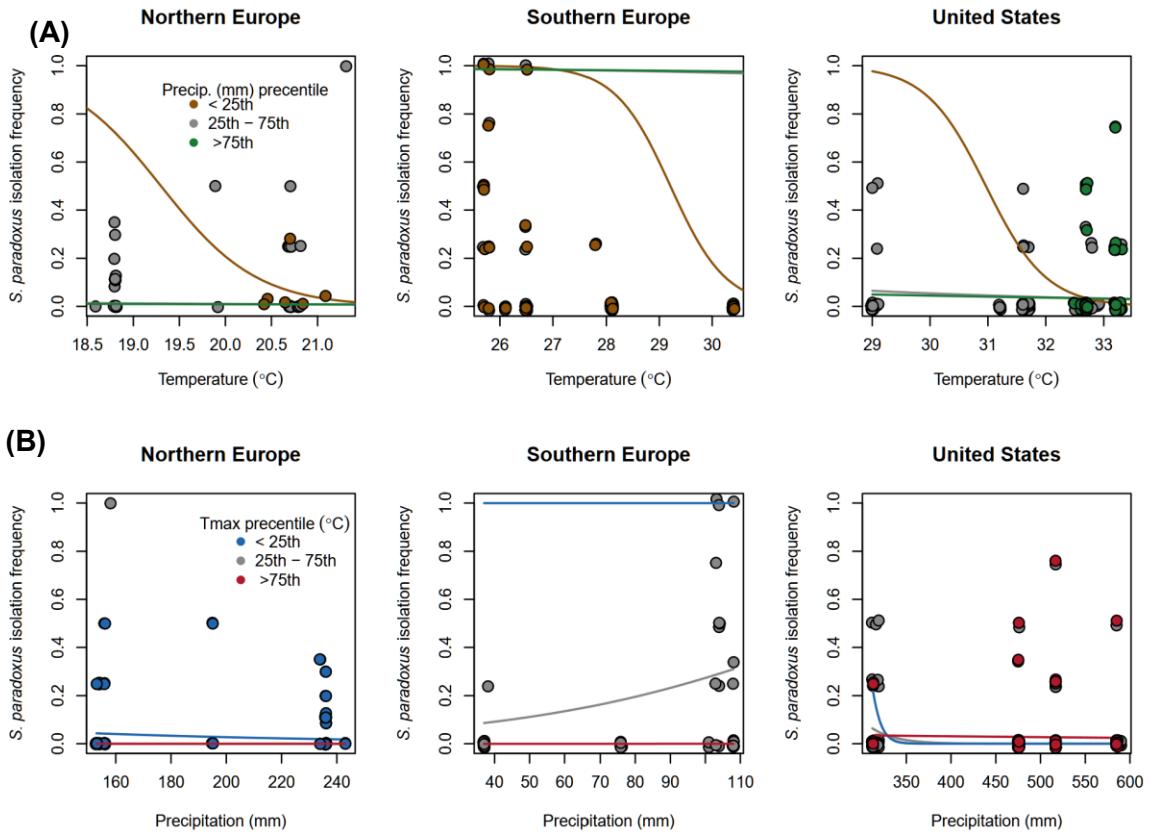


Figure 3.9 The combined effect of temperature ($^{\circ}\text{C}$) of the warmest month (Tmax) (A) and precipitation (mm) of the warmest quarter for the isolation frequency of *S. paradoxus* from 196 *Quercus* species. from Europe and the United States (GLM, 35% deviance). Lines show the probability of isolating *S. paradoxus* using the overall quartiles for precipitation (lower quartile = 154 mm, median = 312 mm, and upper quartile = 319 mm) and Tmax (lower quartile = 20.93 $^{\circ}\text{C}$, median = 29.10 $^{\circ}\text{C}$, and upper quartile = 32.6 $^{\circ}\text{C}$). Lines for the median (gray) and upper quartile (green) overlap for precipitation.

CHAPTER 4

CONCLUSION AND FUTURE DIRECTIONS

Climate change affects every organism, and humans are no exception. Predicting how natural populations respond to climate change is urgent for mitigation strategies. Therefore, we need interdisciplinary approaches to identify what makes a species vulnerable (Boult & Evans, 2021; Pacifici et al., 2015). One approach is to identify which climate associations and ecological factors make a habitat suitable for a species to predict likely species range changes. In this dissertation, I used combined approaches in ecology (field collections and statistical modeling) and evolution (population genomics and phylogenetics) to develop wild yeast as a model. More specifically, I studied phylogeography and species distributions; a necessary baseline for future studies to investigate adaptation. Completion of this research produced hundreds of wild yeast strains and whole genomes of ecological and economic importance. I hope these strains and genomes will be used for future research through the use of functional genomics approaches or for developing new craft beverages. For Chapter 2, we have made our research findings, genome sequences, whole-genome alignments, and other files used for analyses open access to the public. For Chapter 3, we plan to make our strains available in public culture collections, and the Bensasson Lab is happy to share strains with the research community upon request.

In Chapter 2, I used a phylogenetic and population genomic approach to understand the phylogeography of wild *S. cerevisiae* from forests which are probably its

ancestral habitat (Wang et al., 2012). Unexpectedly, we discovered that forest lineages were not entirely independent from human influences despite forests that tend to be isolated from human activity. According to our rough estimates based on laboratory estimates of mutation rate and generation time, the timing when forest lineages diverged out of Asia started after the last glacial maximum and the times when wild forest lineages arrived in North America (~12 kya) and into European forests (~6 kya) roughly coincide with the beginnings of human sedentism (~14 kya), the rise of farming and agriculture (~10 kya), and fermentation practices (~9 kya) (Marsit et al., 2017; Nielsen et al., 2017). Additionally, we discovered that three strains isolated from European forests genetically resemble contemporary North American lineages crossed modern wine strains. We attribute this recent genetic admixture to how humans responded to the Great French Wine Blight. Our results show that footprints of human migration have affected forest populations, albeit rarely. However, when assisted human migration does occur, it can have a significant impact on wild yeast.

In Chapter 3, I expanded my research focus to study the climate associations of multiple yeast species that live sympatrically in forests. By intensely sampling over a hundred trees across a natural temperature gradient along the eastern United States, we have isolated over 450 strains, where *S. cerevisiae*, *S. paradoxus*, and *L. fermentati* were the most isolated species. This allowed me to estimate a rough thermal optimum for wild *S. cerevisiae*, and thus to estimate a climate envelope model based on ecological data. Additionally, we might have identified a yeast species endemic to the United States, *L. fermentati*, though it might occur in southeast Asia or other regions with forests and hot wet summers. Lastly, I determined that *S. paradoxus* is the most ecologically complex in

that tree age and climate are important to its species distribution. Our results illustrate that the sympatric yeasts of U.S. forests differ in their climate associations. For all three species, changes in temperature or precipitation could therefore affect their distribution, and populations could become locally (mal)adapted.

Overall, this dissertation provides a baseline to expand wild yeast as a model by establishing (i) genetically distinct lineages and (ii) showing that the ranges of these natural populations are limited by climate. Future research avenues should test whether wild *S. cerevisiae* forest lineages are locally adapted (or maladapted) through phenotypic screens (a common garden approach) to grow strains across a broad range of temperatures between 18-42°C and under fluctuating temperature conditions. If specific lineages are doing better under certain conditions, then our data could be used to scan for signatures of natural selection (i.e., nucleotide diversity, F_{st} , Tajima's D) that could be associated with changes in thermal optima. If different yeast species show similar signatures of selection this could be signs of convergent evolution. This future research avenue is important for our understanding of eukaryotes and, especially given our lack of knowledge of how fungi may respond to climate change. Given that human pathogenic fungi can be isolated from forests and marshes, it seems important to understand the evolution of pathogenic yeasts in the wild.

4.1 References

- Boult, V. L., & Evans, L. C. (2021). Mechanisms matter: Predicting the ecological impacts of global change. *Global Change Biology*.
<https://doi.org/10.1111/gcb.15527>
- Marsit, S., Leducq, J.-B., Durand, É., Marchant, A., Filteau, M., & Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nature Reviews Genetics*, 18(10), 581–598. <https://doi.org/10.1038/nrg.2017.49>

Nielsen, R., Akey, J. M., Jakobsson, M., Pritchard, J. K., Tishkoff, S., & Willerslev, E. (2017). Tracing the peopling of the world through genomics. *Nature*, 541(7637), Article 7637. <https://doi.org/10.1038/nature21347>

Pacifci, M., Foden, W. B., Visconti, P., Watson, J. E. M., Butchart, S. H. M., Kovacs, K. M., Scheffers, B. R., Hole, D. G., Martin, T. G., Akçakaya, H. R., Corlett, R. T., Huntley, B., Bickford, D., Carr, J. A., Hoffmann, A. A., Midgley, G. F., Pearce-Kelly, P., Pearson, R. G., Williams, S. E., ... Rondinini, C. (2015). Assessing species vulnerability to climate change. *Nature Climate Change*, 5(3), Article 3. <https://doi.org/10.1038/nclimate2448>

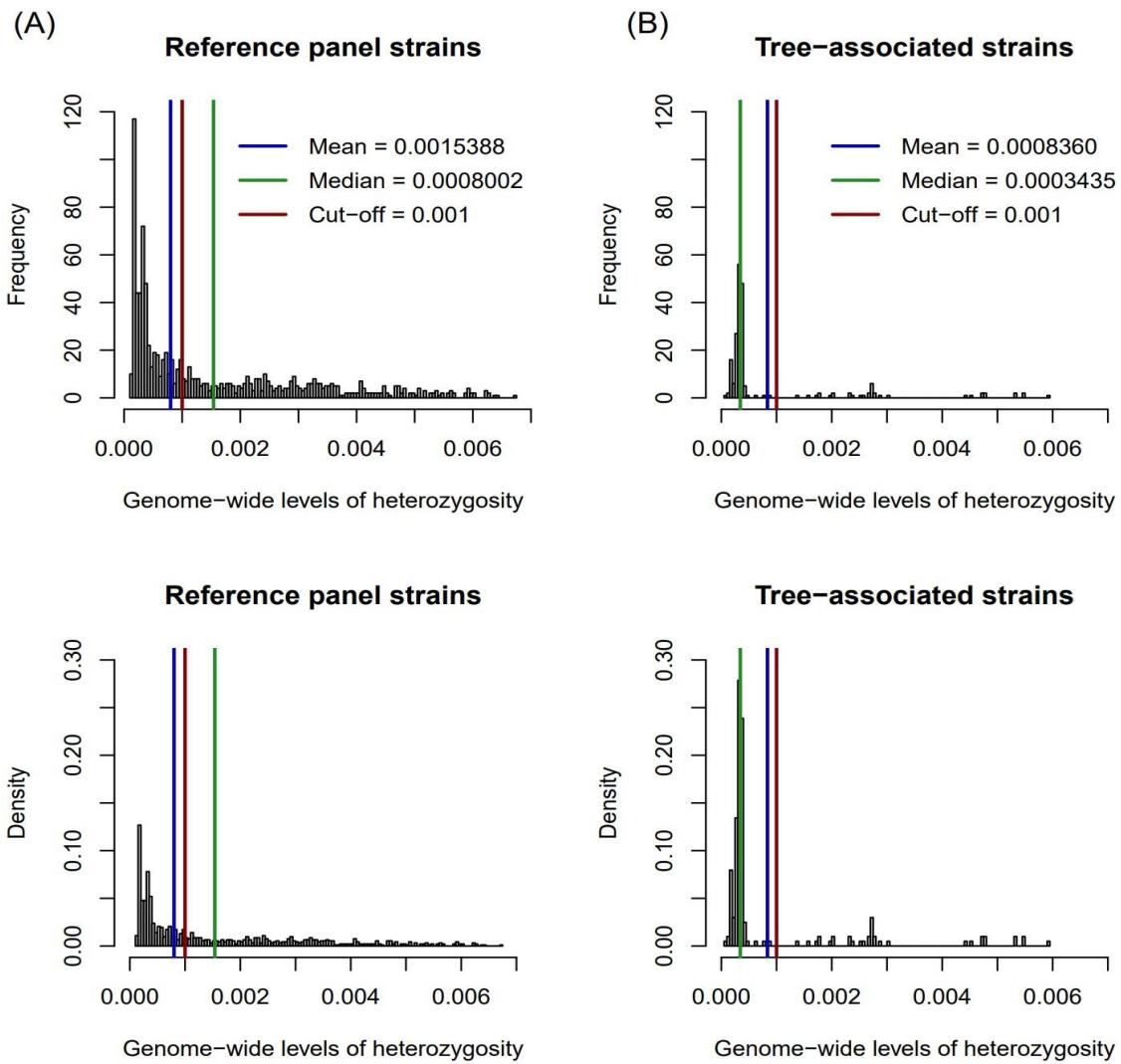
Wang, Q.-M., Liu, W.-Q., Liti, G., Wang, S.-A., & Bai, F.-Y. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular Ecology*, 21(22), 5404–5417. <https://doi.org/10.1111/j.1365-294x.2012.05732.x>

APPENDIX A

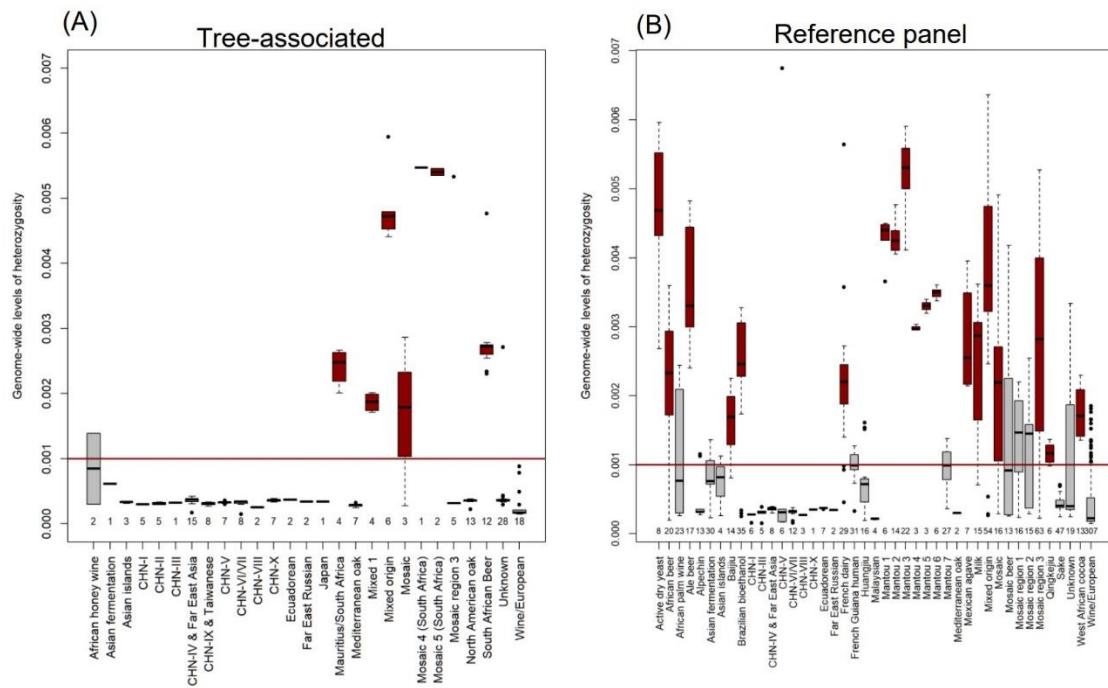
SUPPLEMENTAL MATERIALS FOR CHAPTER 2

Supplemental tables 2.1-2.8 can be found at

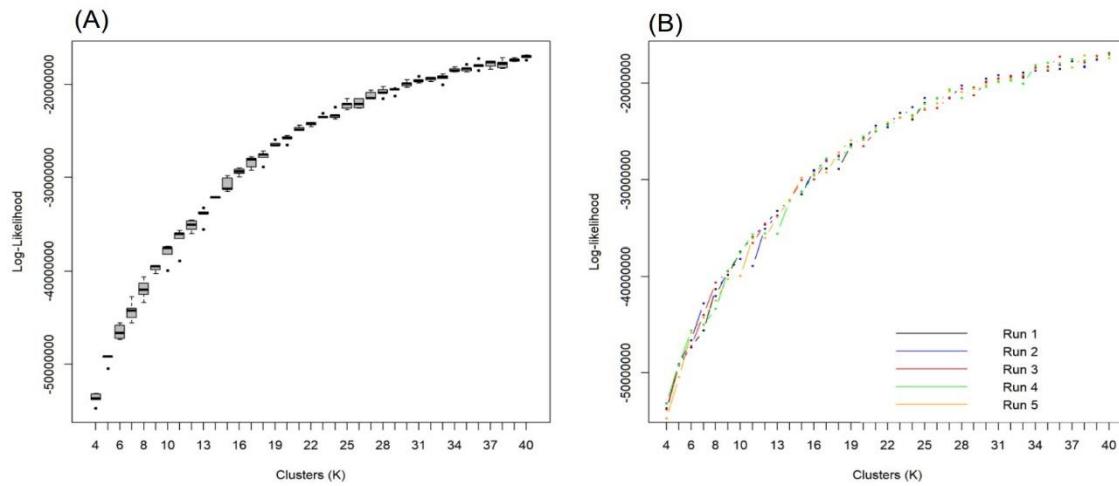
https://github.com/bensassonlab/data/blob/master/Pena_dissertation25/PenaCh2_SupplementalTables.xlsx



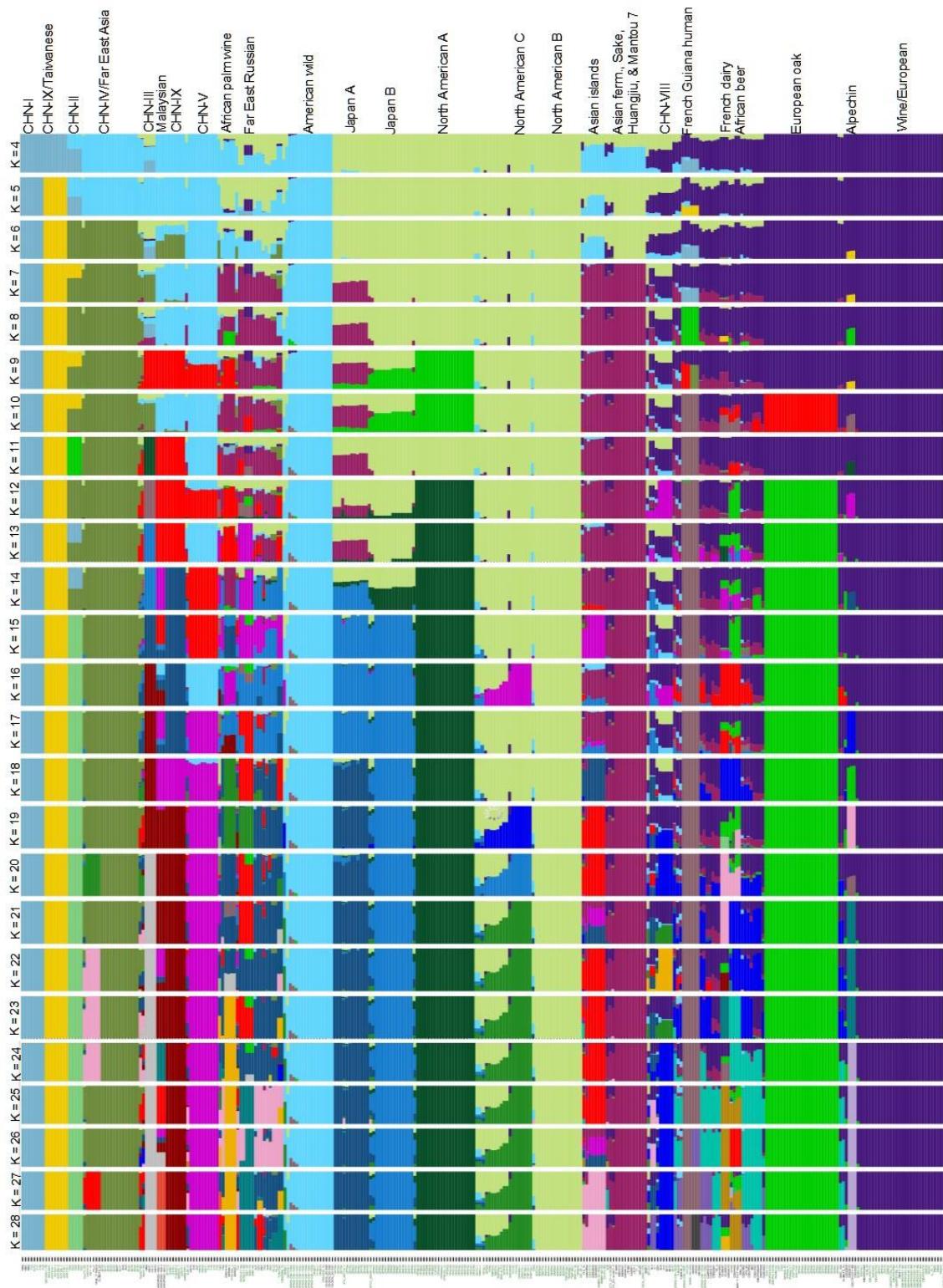
Supplemental Figure 2.1 Genome-wide levels of heterozygosity of *Saccharomyces cerevisiae* partitioned by tree-associated ($N = 172$) (A) and non-tree-associated strains ($N = 881$) (B) after excluding monosporic derivatives. For downstream population structure analyses, we applied a 0.001 heterozygosity cutoff (red vertical line) to exclude strains that are likely to be inter-clade hybrids, which would obscure phylogenetic relationships.

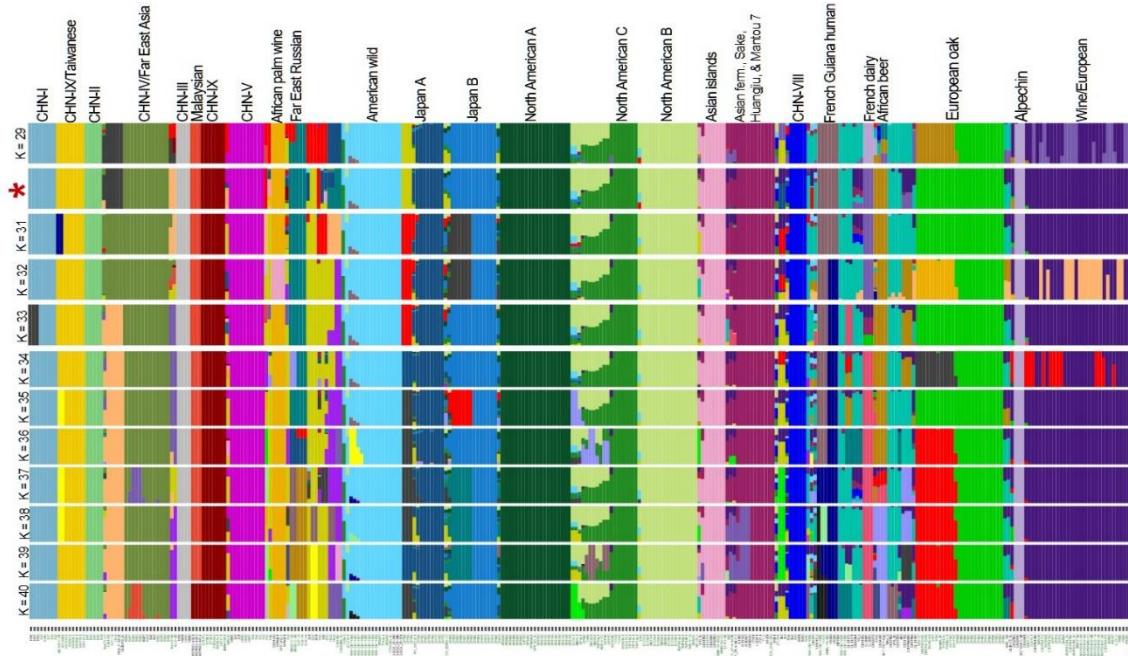


Supplemental Figure 2.2 Genome-wide levels of heterozygosity for *Saccharomyces cerevisiae* of tree-associated ($N = 172$) (A) and non-tree-associated strains ($N = 881$) (B) partition by published clades after excluding monosporic derivatives. Heterozygous clades are highlighted in red where 90% or more of the strains within that clade have a genome-wide level of heterozygosity greater than 0.001 (horizontal red line). Numbers on the plots indicate the number of strains per clade. Heterozygous clades are prevalent in non-tree-associated clades ($N = 37$) compared to tree-associated clades ($N = 21$) (Fisher's exact test, $P = 0.04$). For statistical analyses, we excluded published clades that were defined as 'mosaic clades' (i.e. Mosaic region 1) because these lineages represent multiple source populations ($N = 5$ tree-associated clades excluded and $N = 3$ nontree-associated clades) and strains that have an unknown lineage. A reference panel of non-tree-associated strains from published clades was created by randomly selecting three strains per homozygous clade (highlighted in gray) to see how tree-associated strains genetically cluster with non-tree associated strains. For example, all the strains within the Active dry yeast lineage are heterozygous and thus this lineage is not represented in the reference panel because heterozygous strains are likely intra-clade hybrids, which would obscure phylogenetic relationships.

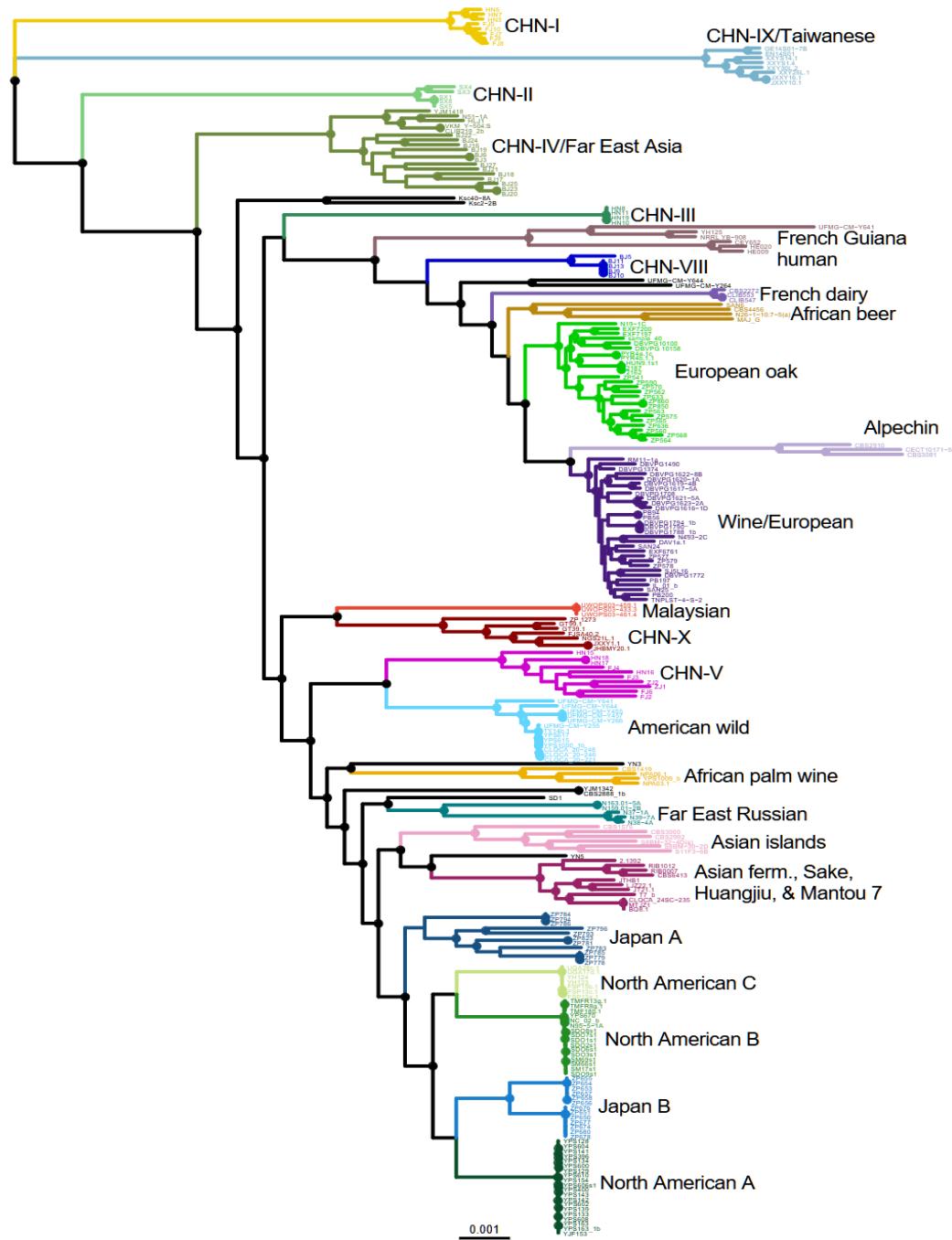


Supplemental Figure 2.3 Loglikelihood values from ADMIXTURE analyses for each cluster (K) across five replicate runs among 313 *S. cerevisiae* tree-associated and reference panel strains (non-tree-associated) strains. (A) The loglikelihood values as a function of five replicate runs. (B) The loglikelihood values for each replicate run across Ks.

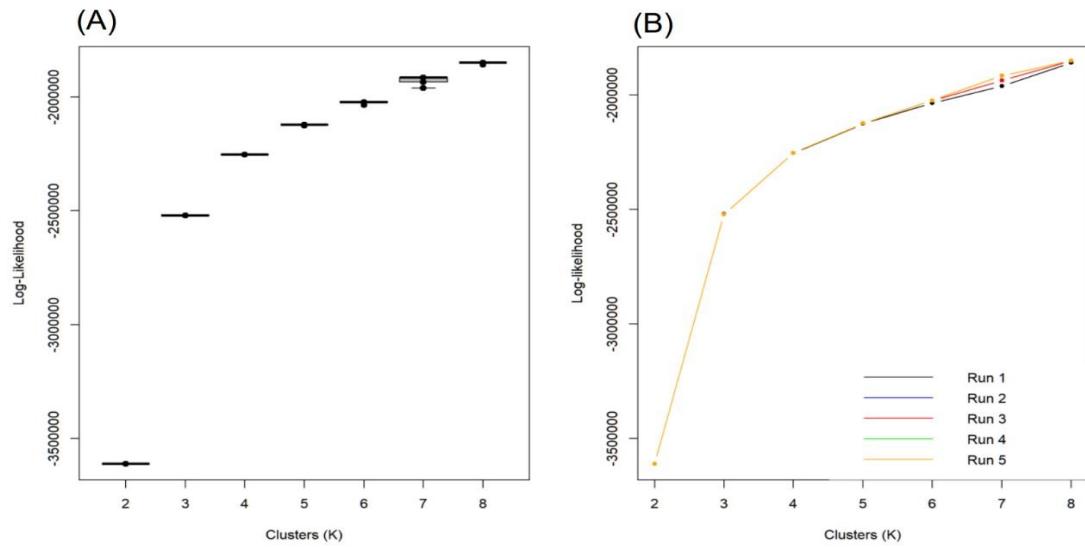




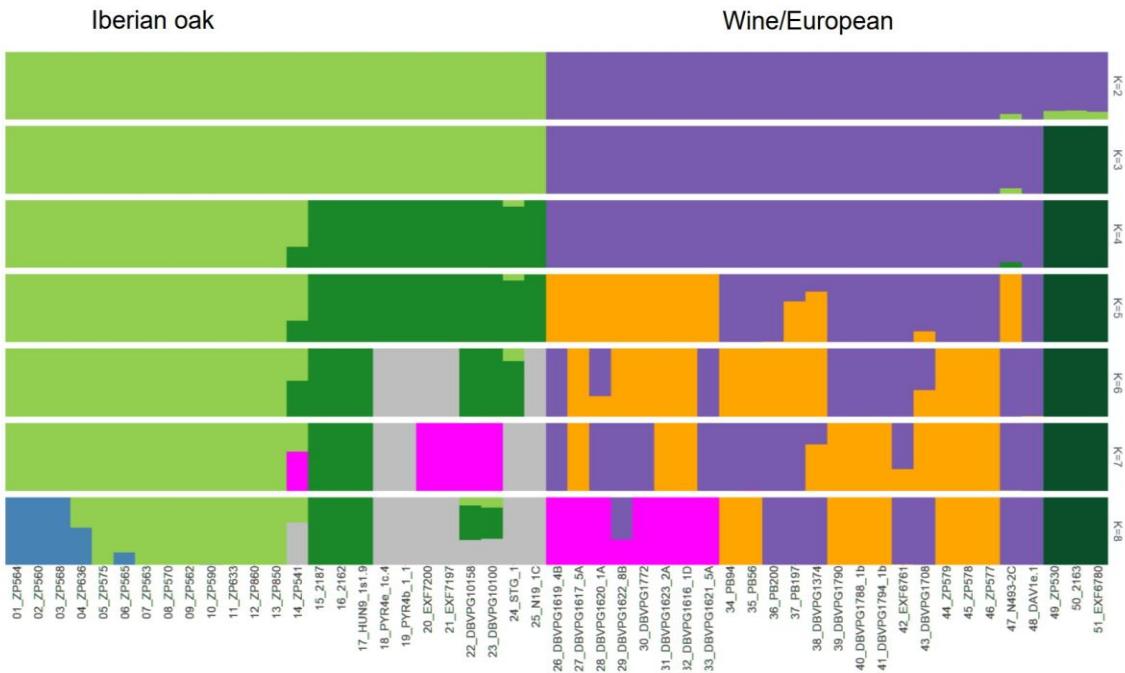
Supplemental Figure 2.4 Population structure and admixture of *Saccharomyces cerevisiae* tree-sampled strains and reference panel strains after excluding strains with heterozygosity > 0.001. Lineages were estimated using ADMIXTURE from varying cluster (K) 4-40 with five replicate runs per each K. We selected the run that had the highest loglikelihood value from each K and strains are ordered by their position in the neighbor-joining tree. Lineages that only have tree-sampled strains are highlighted in green text. The ADMIXTURE plot highlighted with a red asterisk (K = 30) is the model that had distinct genetic clusters that matched monophyletic clades in the neighbor-joining tree (>95% bootstrap support).



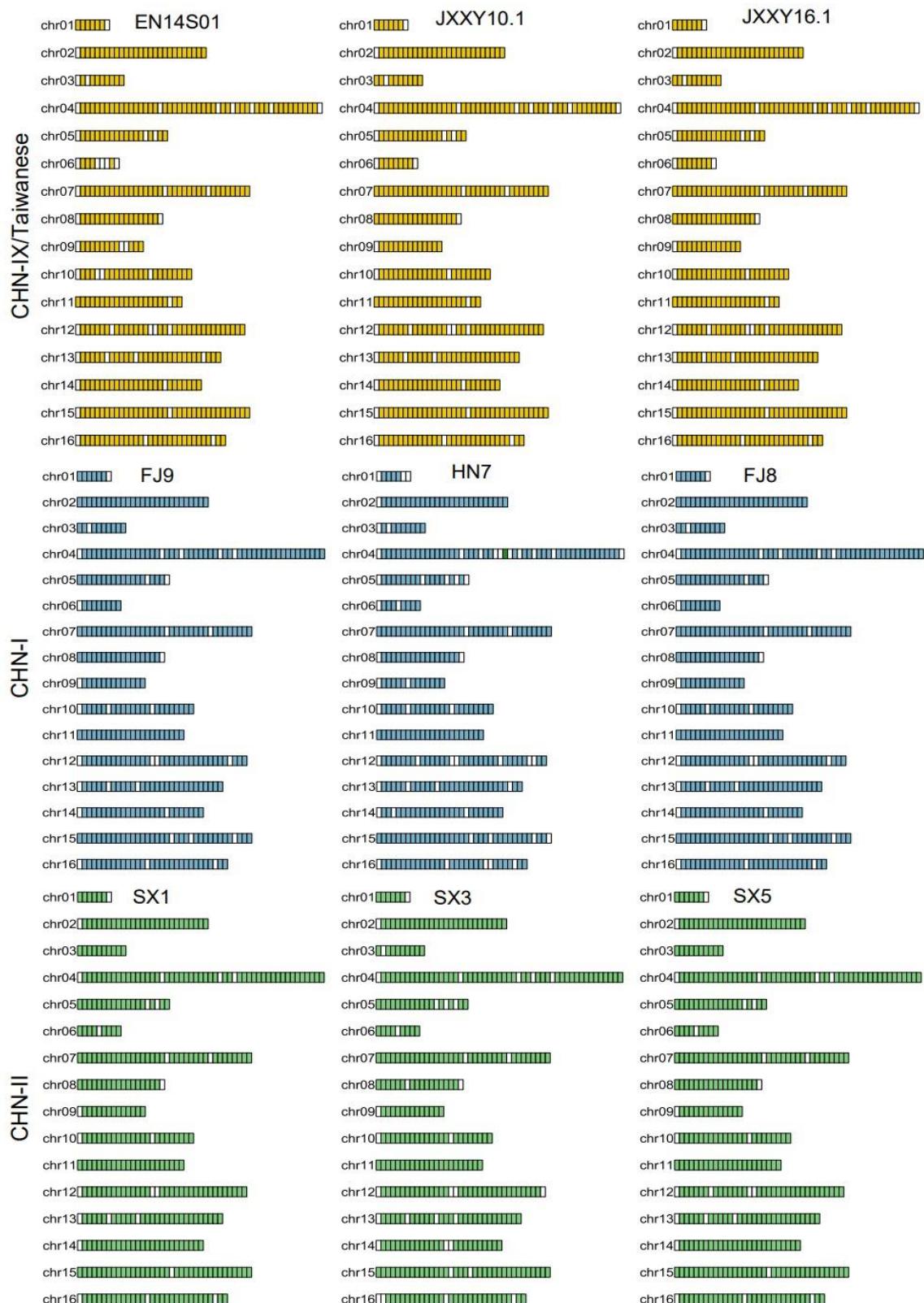
Supplemental Figure 2.5 Phylogenetic relationships of *Saccharomyces cerevisiae*. Whole-genome phylogeny of tree-sampled and reference panel strains after excluding admixed strains (percent ancestry to a single lineage is <90% when K = 30; Figure 1B). The phylogeny was constructed using maximum likelihood estimation using IQtree ultrafast bootstrapping (1000 bootstraps) using a general-time reversible model with a gamma distribution. Filled circles at nodes show monophyletic clades with >99% bootstrap support.

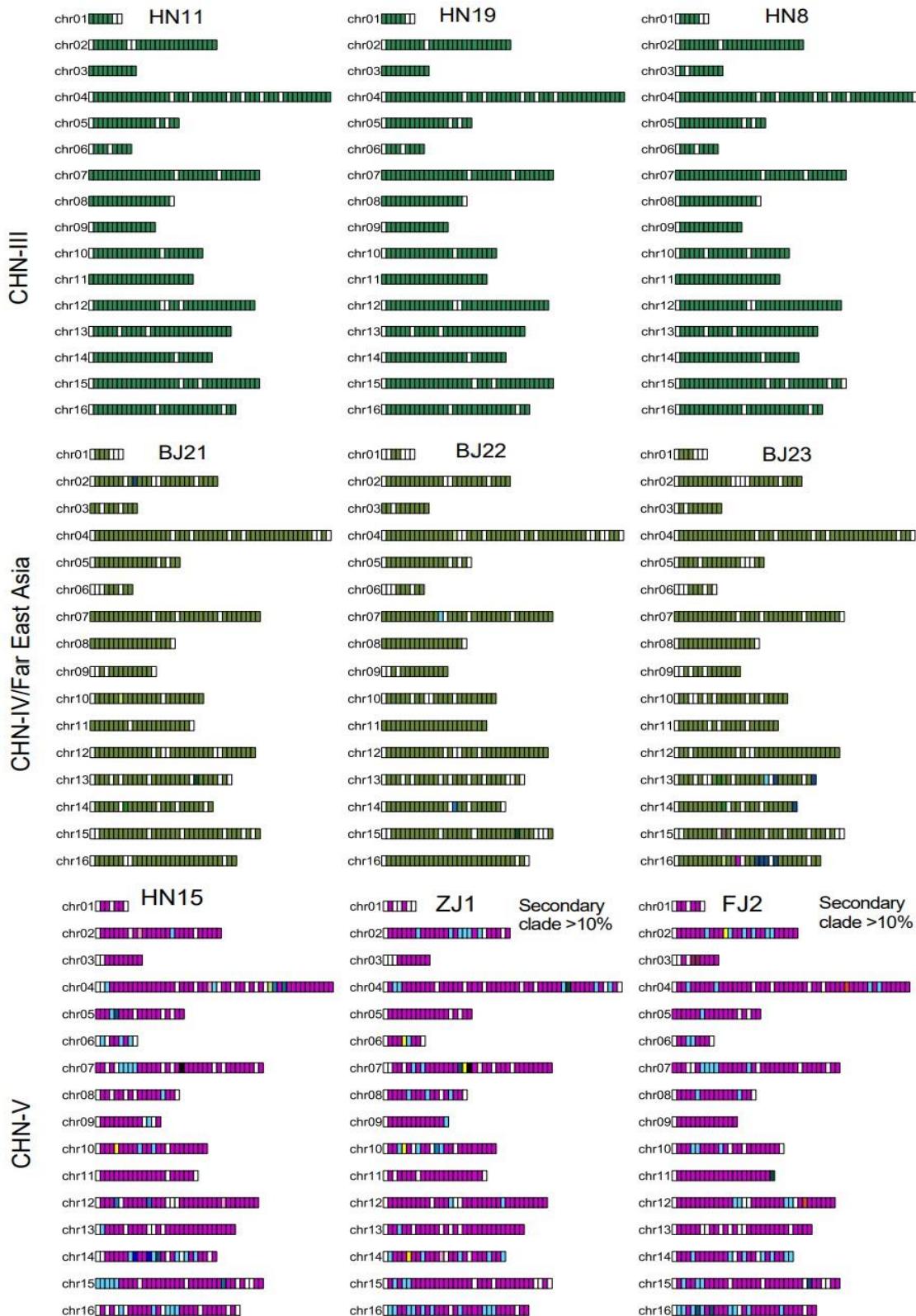


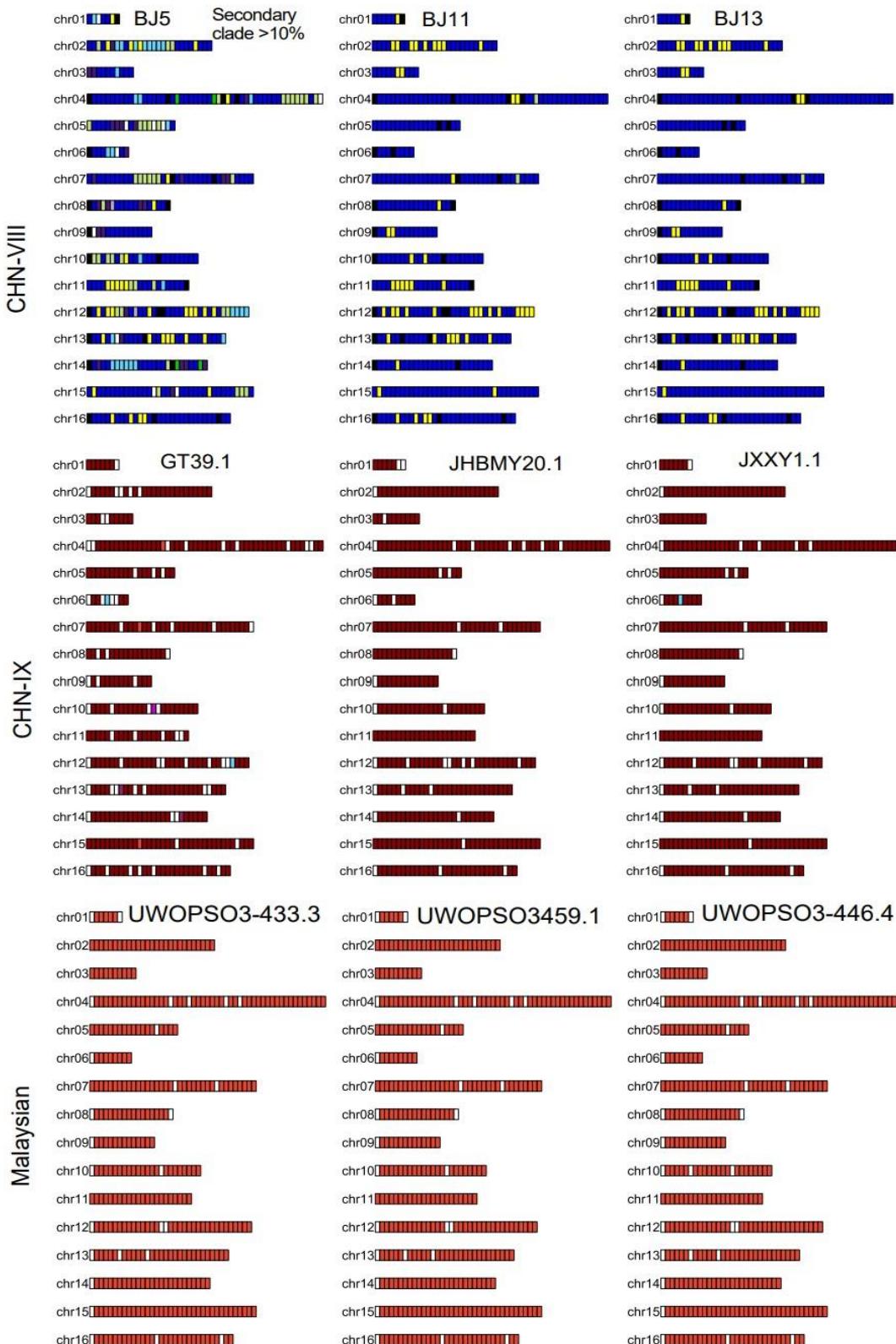
Supplemental Figure 2.6 Loglikelihood values from ADMIXTURE analyses for each cluster (K) across five replicate runs among 51 wild *S. cerevisiae* from European woodlands. (A) The loglikelihood values as a function of five replicate runs. (B) The loglikelihood values for each replicate run across Ks.

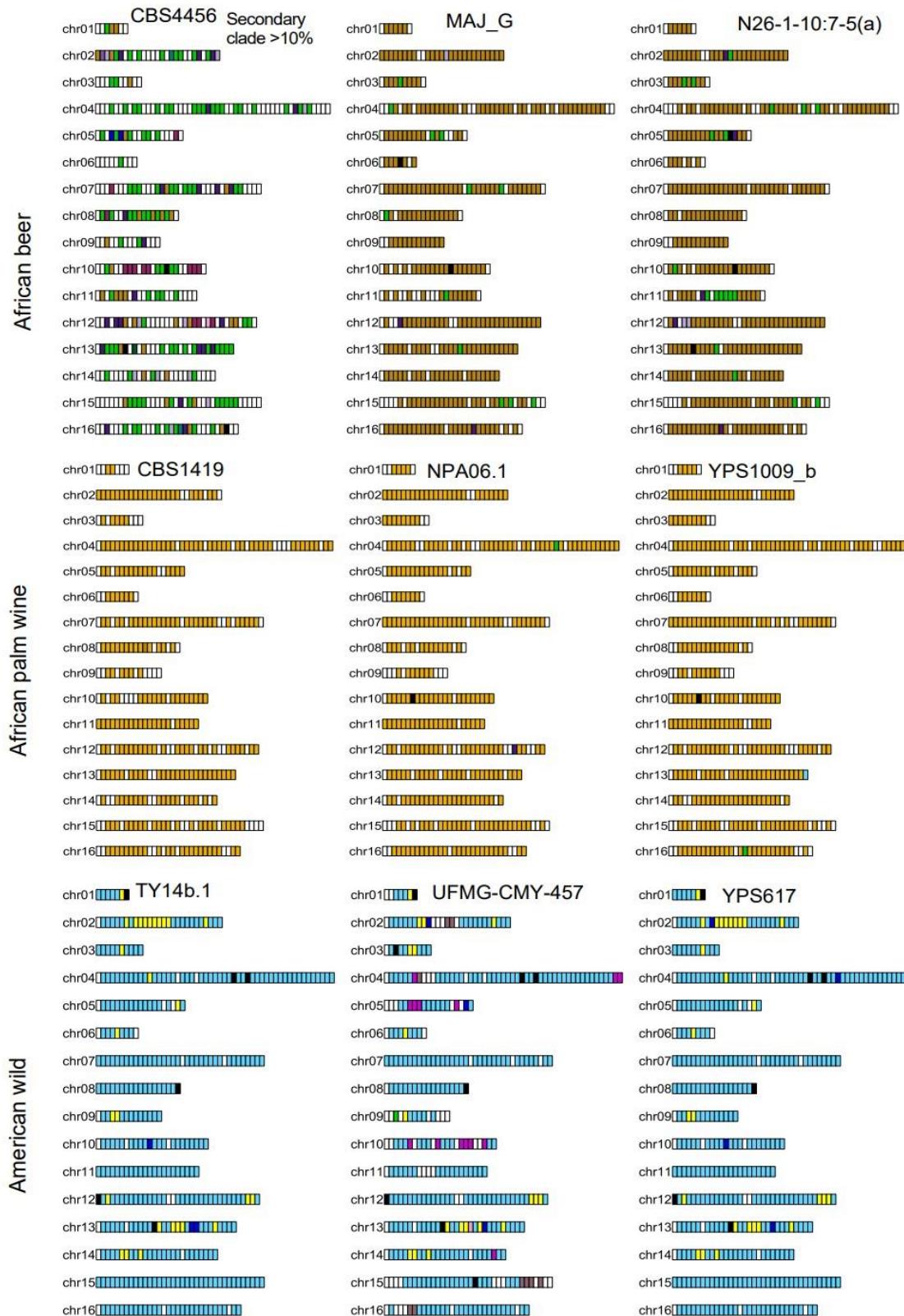


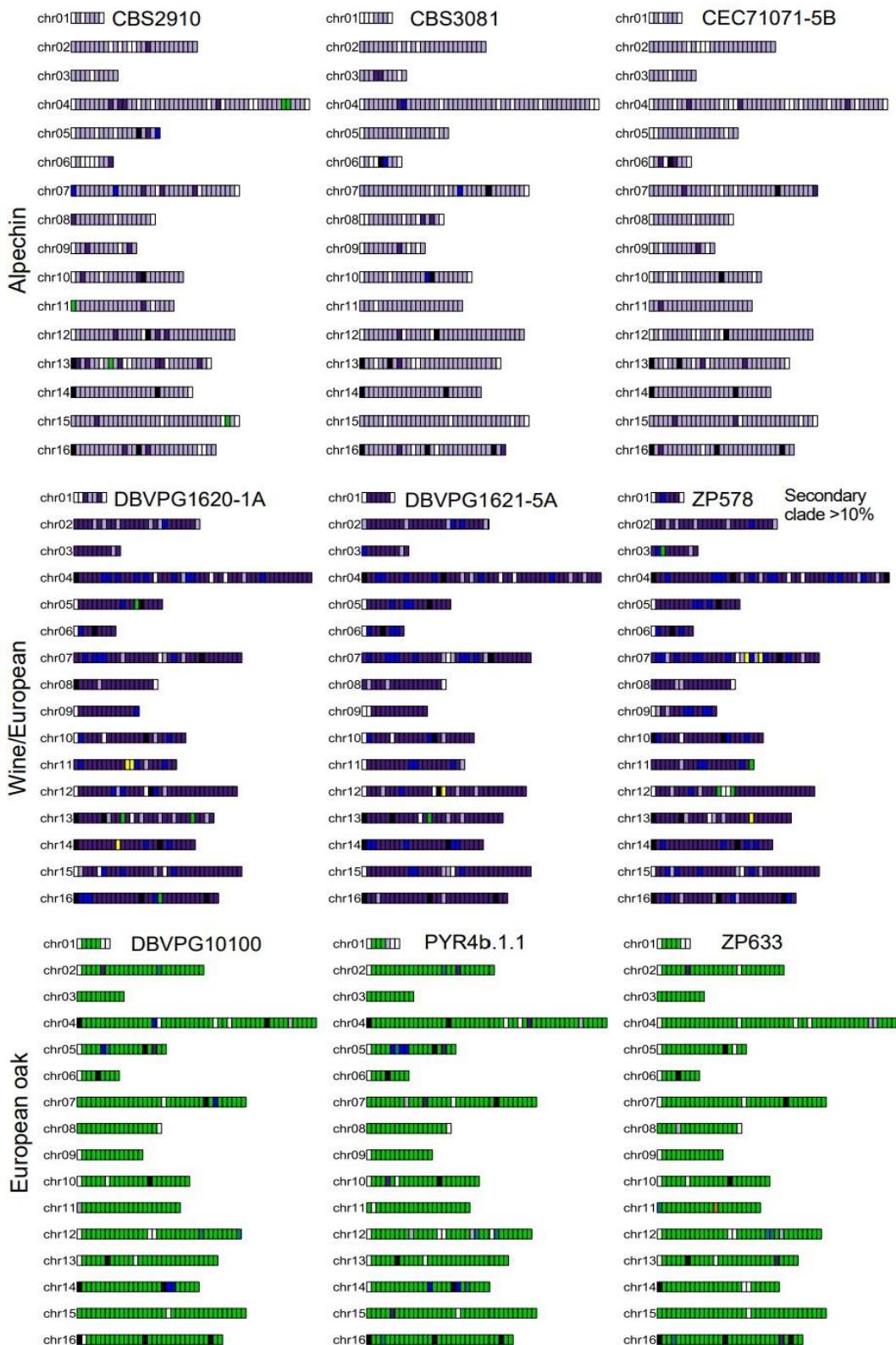
Supplemental Figure 2.7 The population structure and admixture of wild *Saccharomyces cerevisiae* from Europe. Populations were estimated using ADMIXTURE from varying cluster (K) 2-8 with five replicate runs for each K. We selected the run that had the highest loglikelihood value from each K and strains are ordered by their position in the neighbor-joining tree.

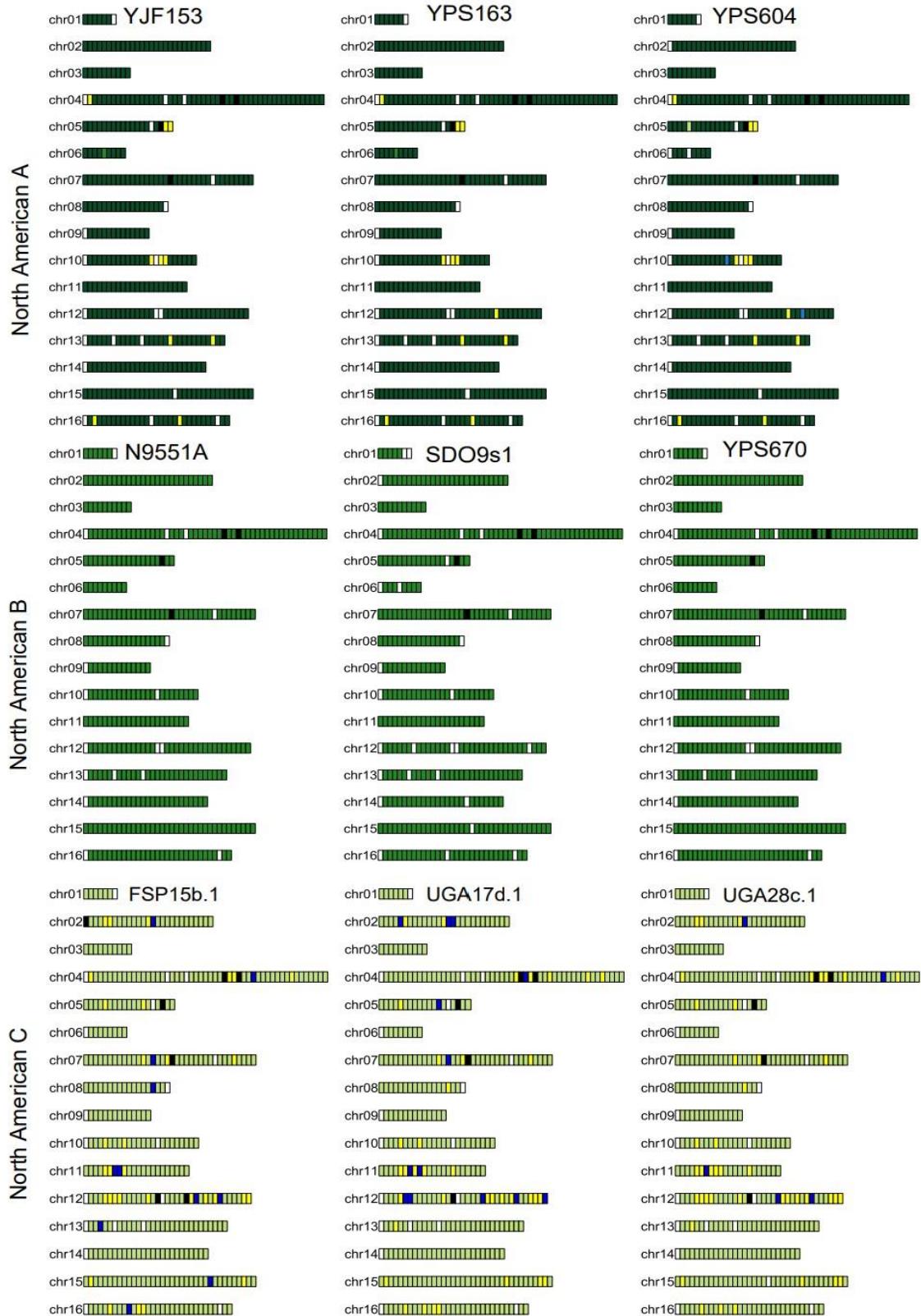


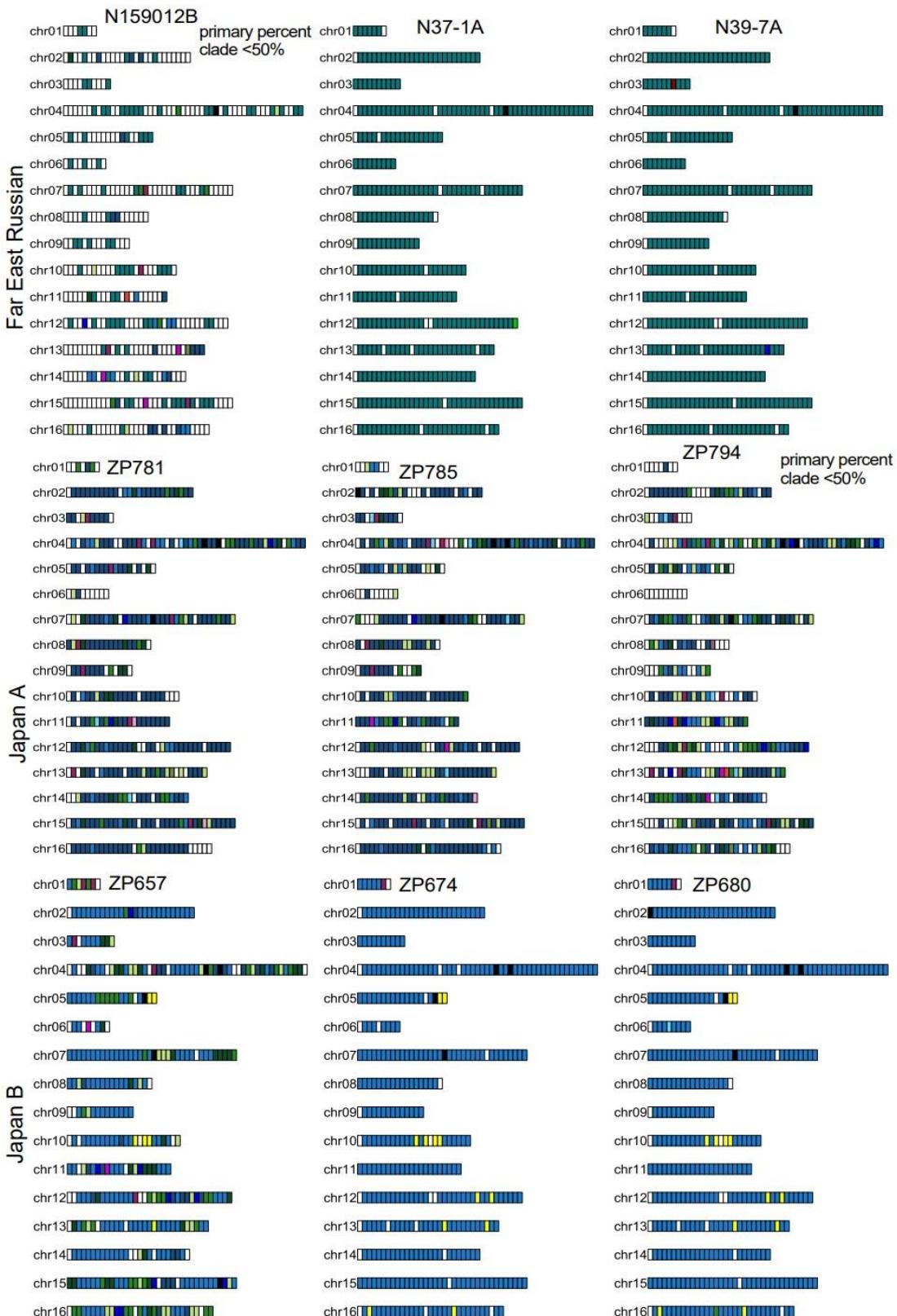


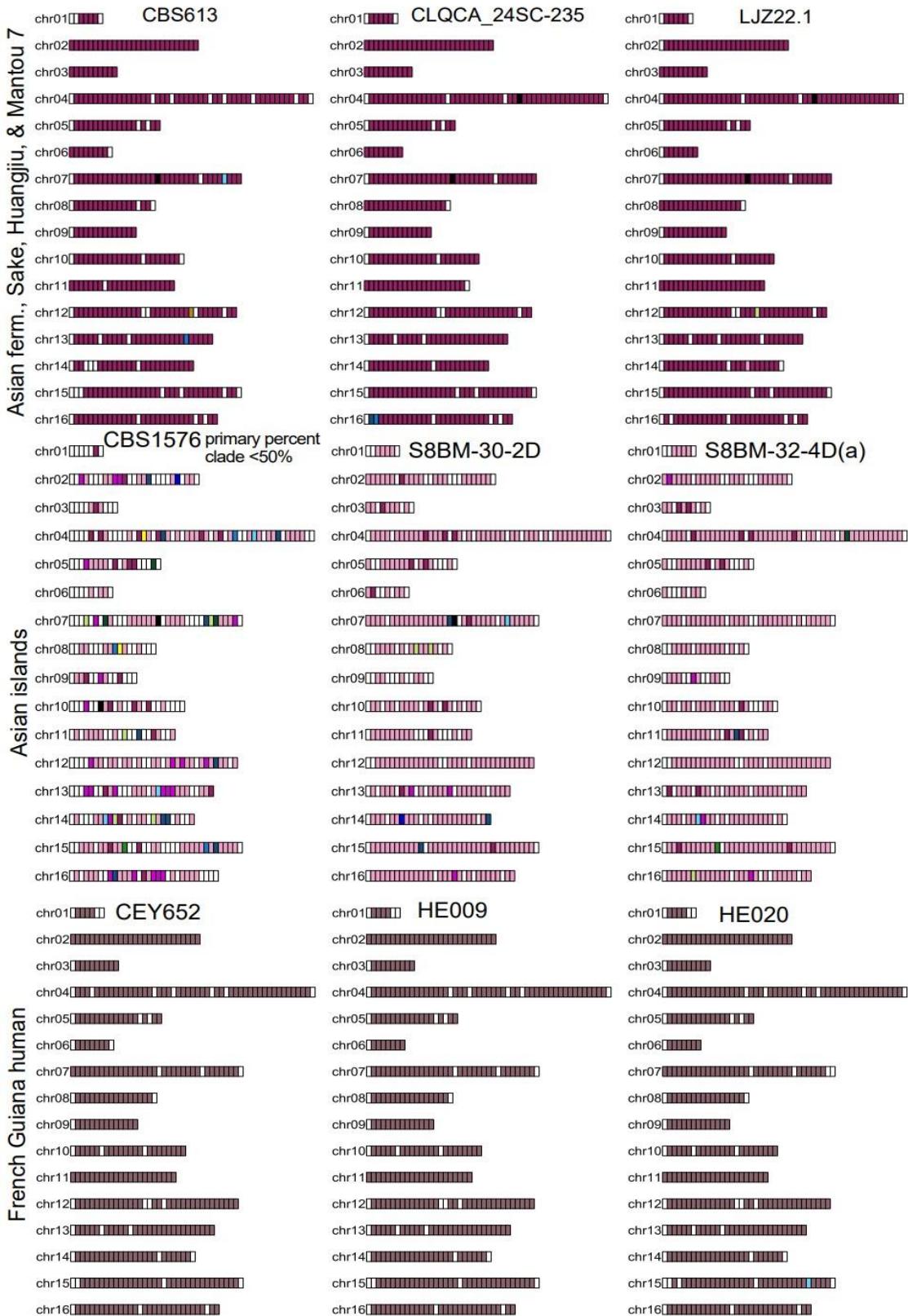






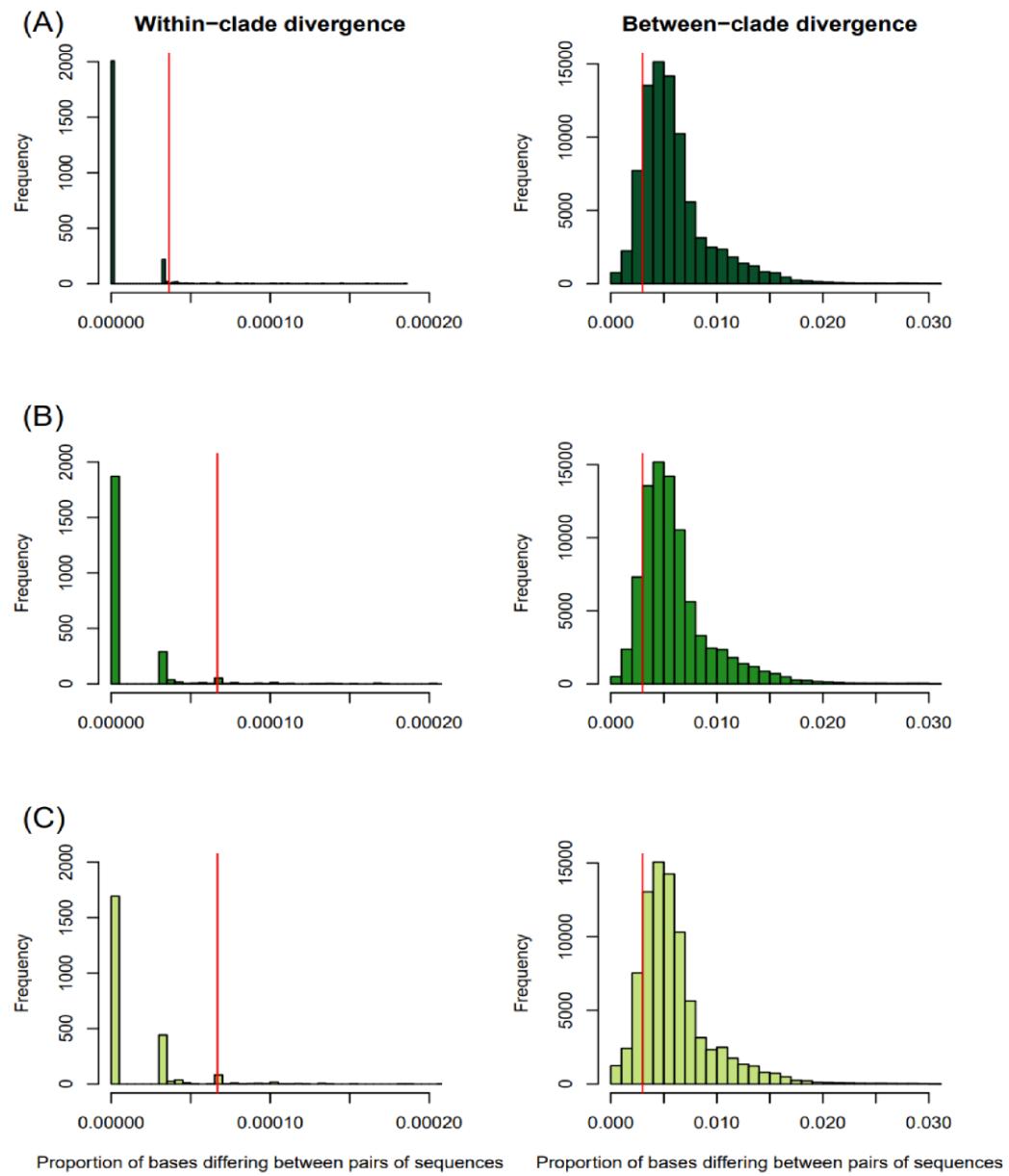




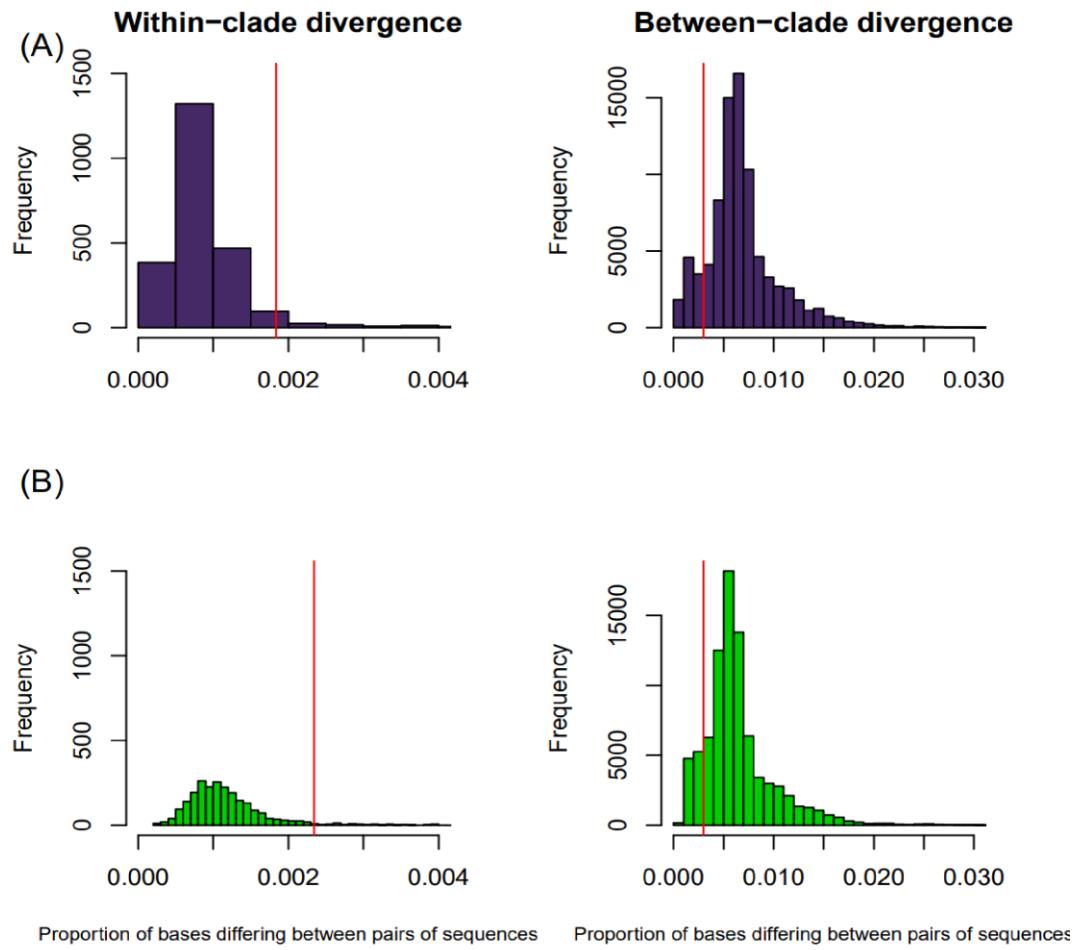




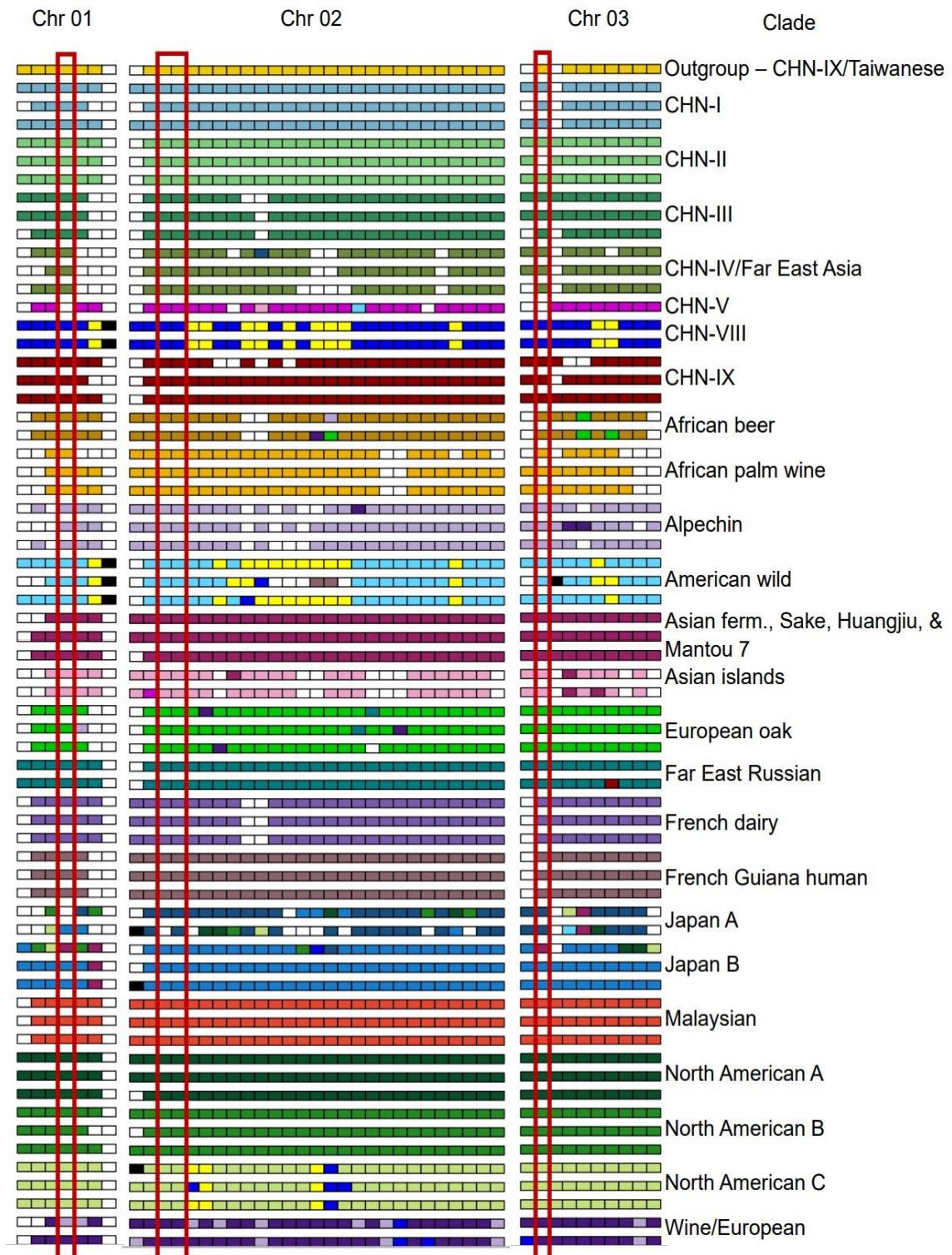
Supplemental Figure 2.8 Chromosome painting for 77 strains used as a backbone phylogeny (Table S3). Genomic regions were “painted” based on the clade assignment of the most similar strain in 30 kb nonoverlapping windows. Diverged regions were not colored (white) and were defined as regions that had a maximum proportion of sites that differed by 0.003 from all other strains in the backbone phylogeny. Black colored regions indicate low coverage, and yellow windows show equal similarity to multiple lineages.



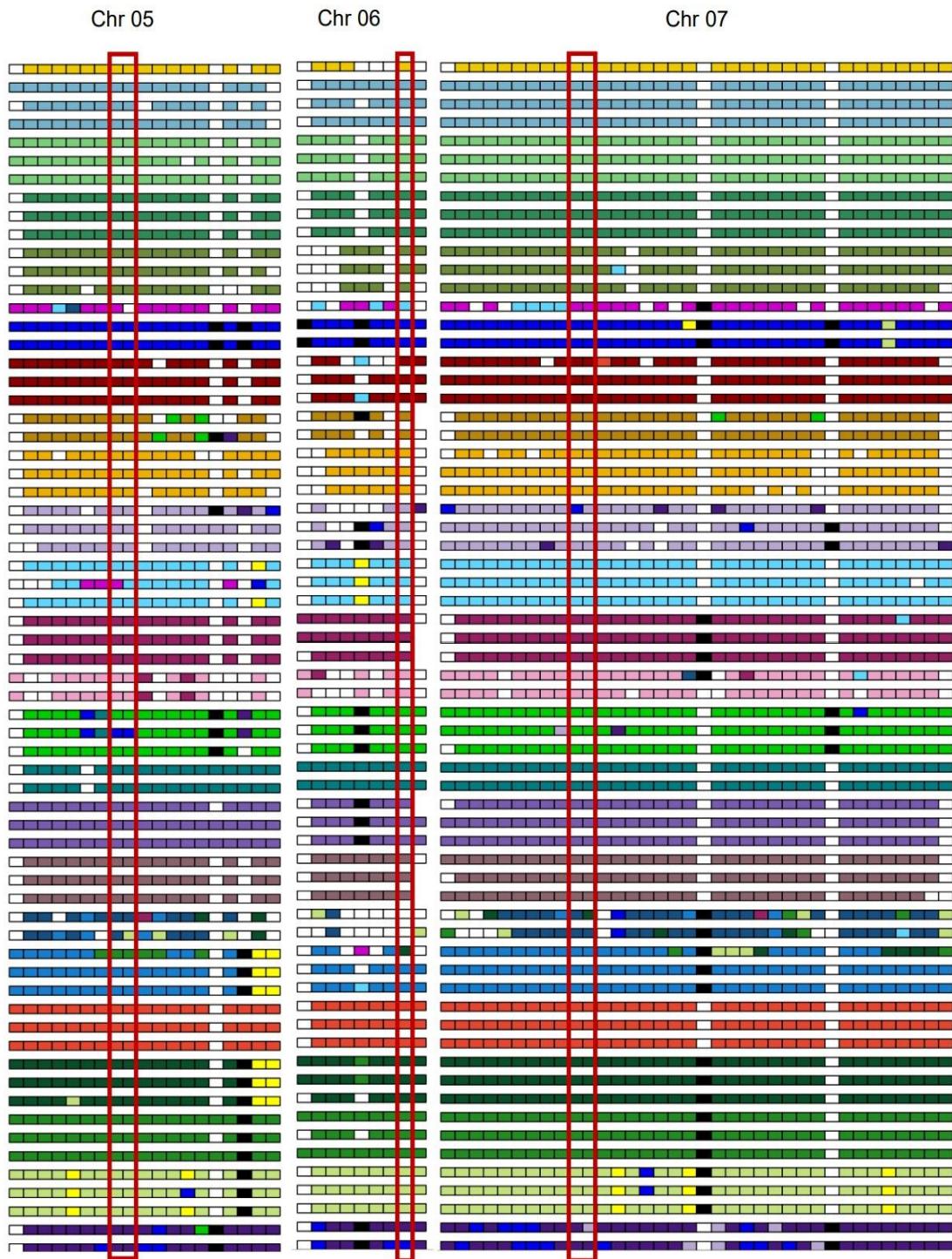
Supplemental Figure 2.9 Histograms of within-clade and between-clade comparisons for North American A-C (A-C). Within-clade divergences are below the 95th quantile (vertical red line), while between-clade divergences are above 0.003 (vertical red line).

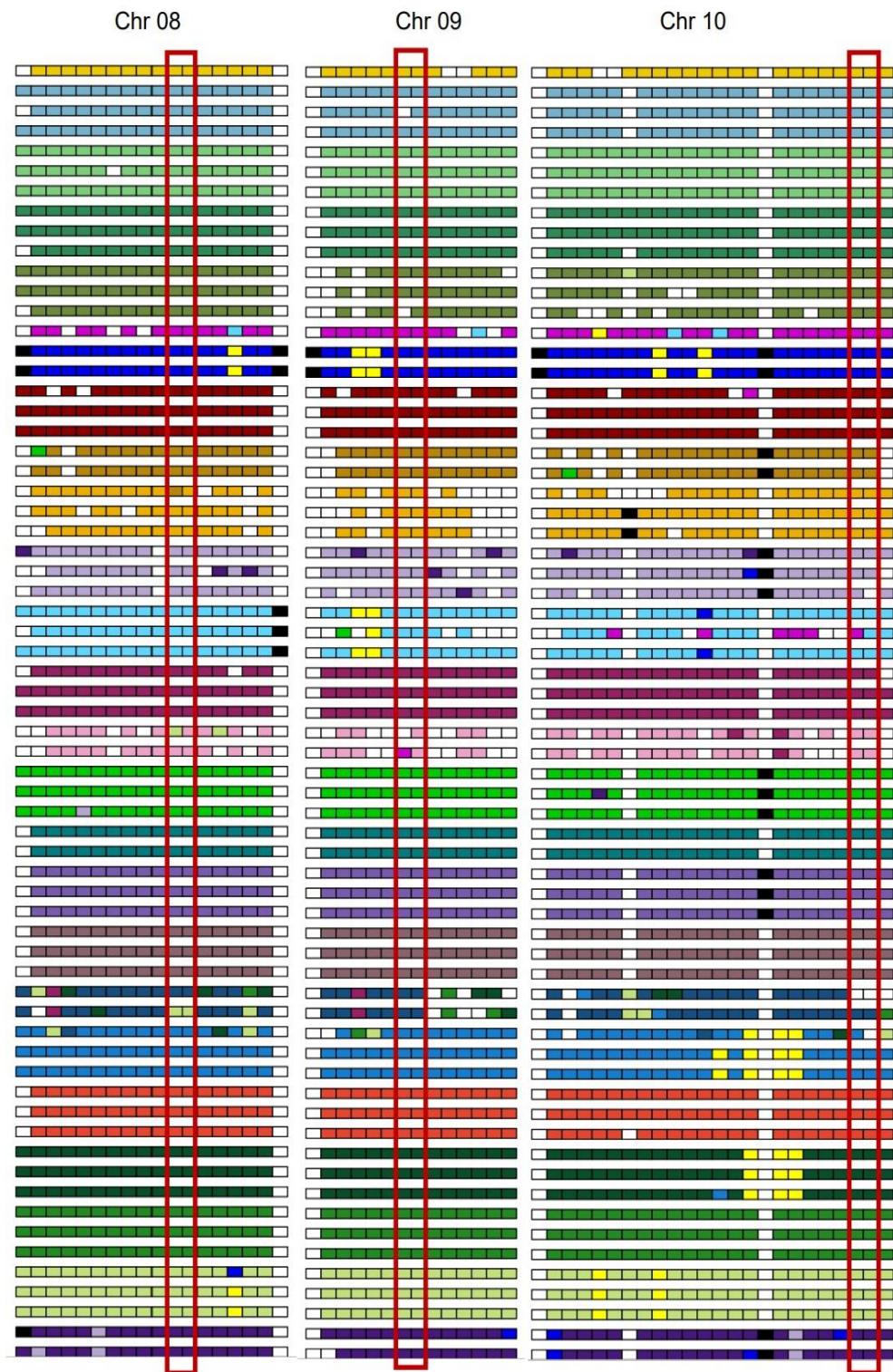


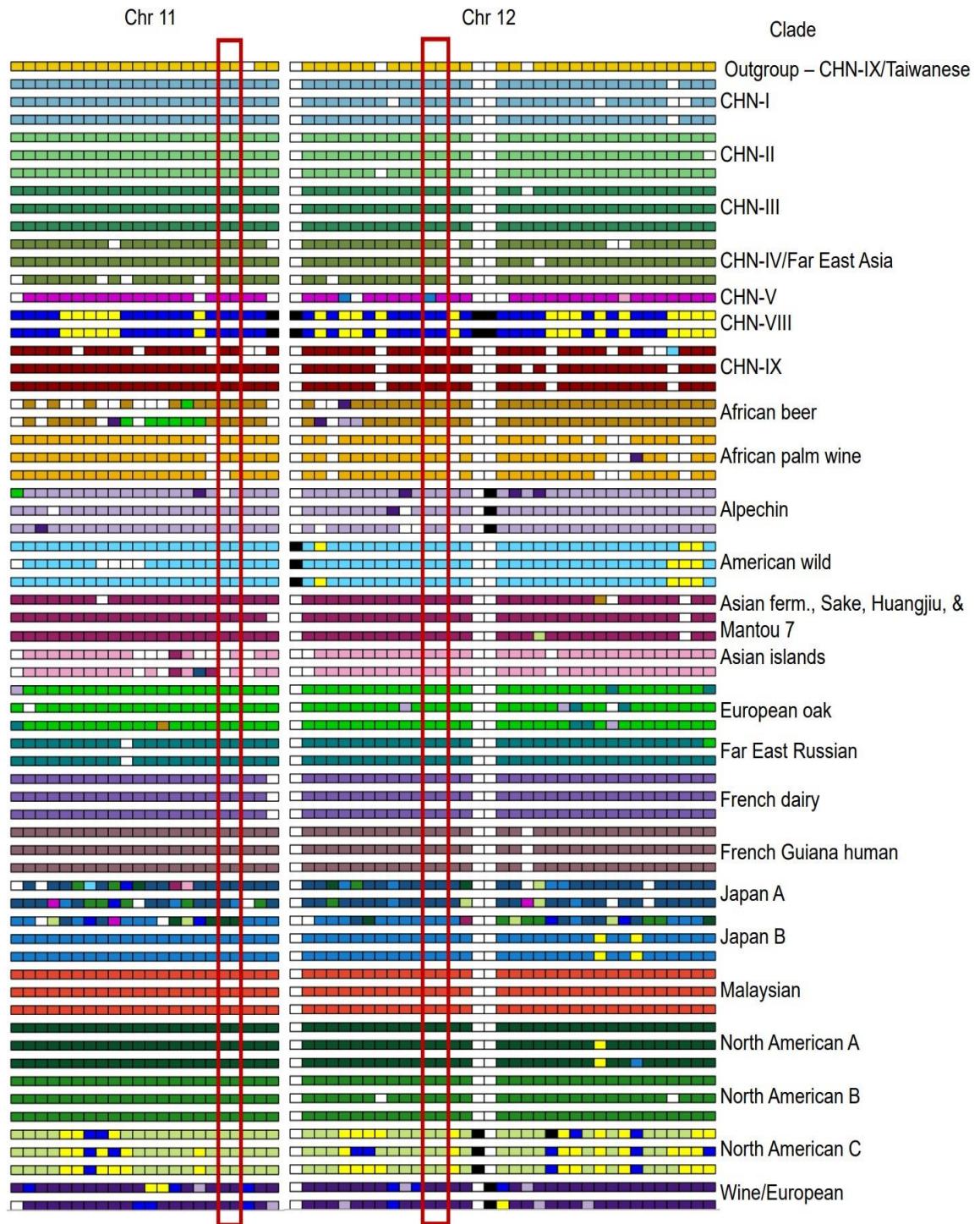
Supplemental Figure 2.10 Histograms of within-clade and between-clade comparisons for Wine/European (A) and European oak (B). Within-clade divergences are below the 95th quantile (vertical red line), while most between-clade divergences are above 0.003 (vertical red line).

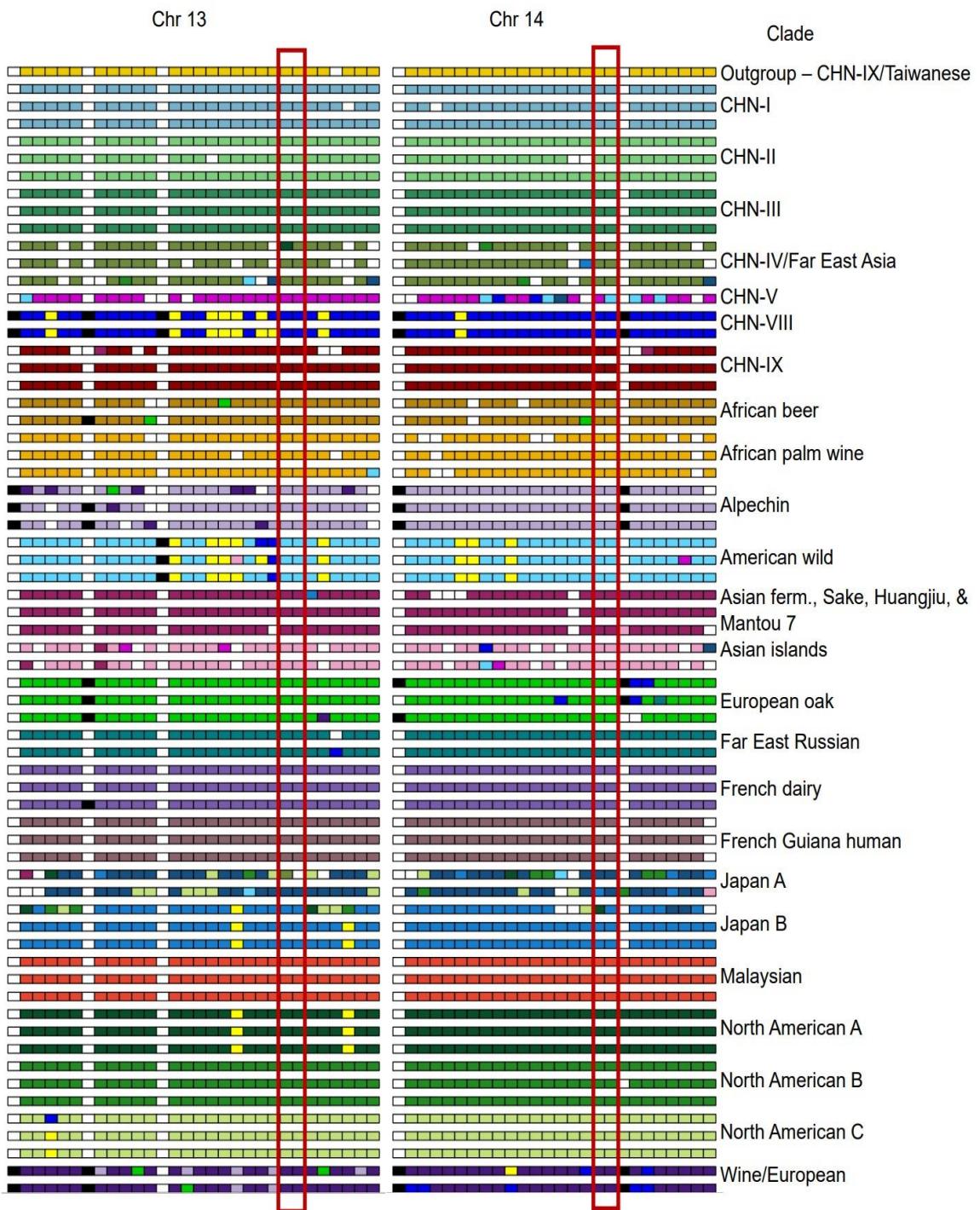


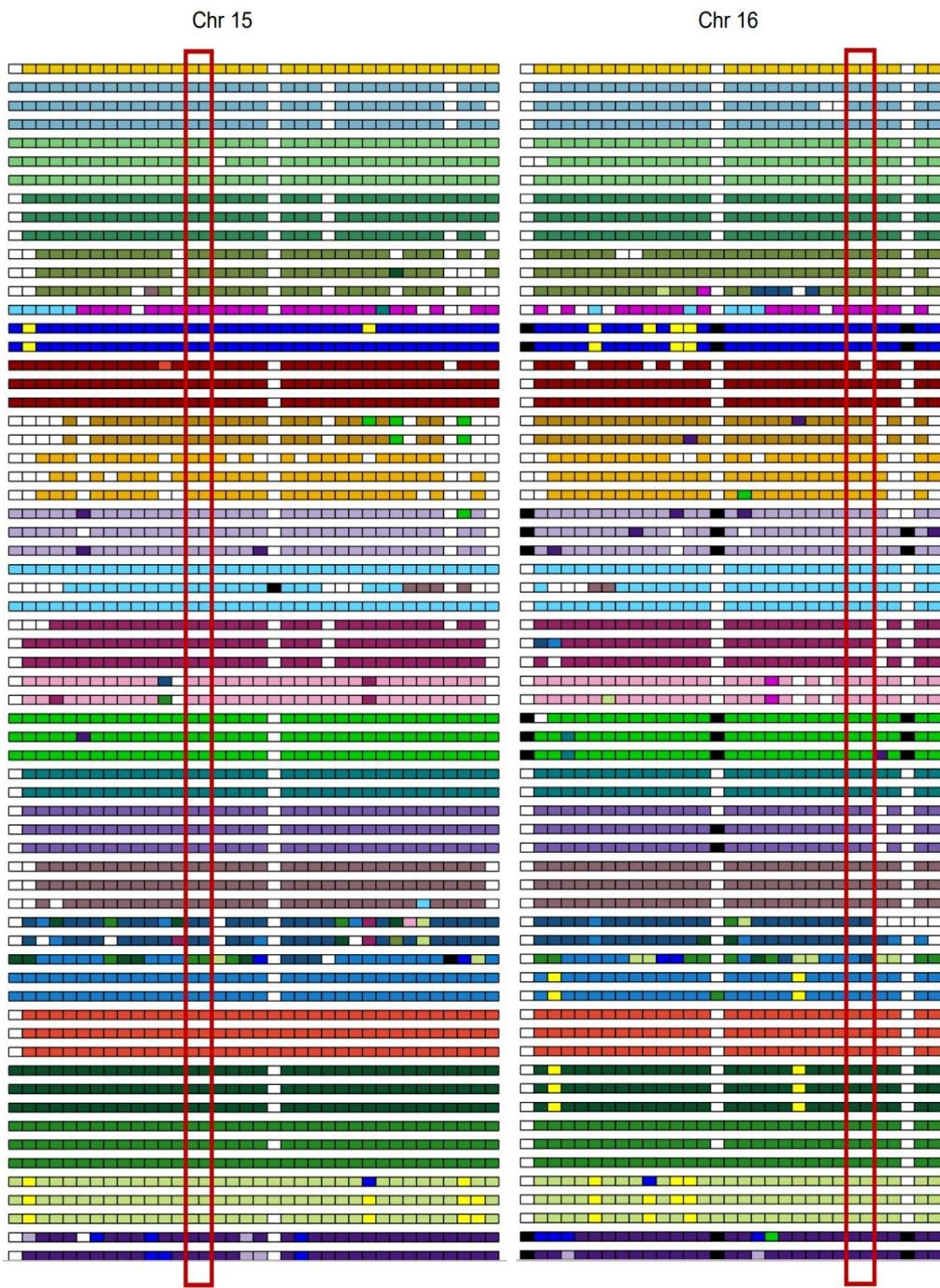






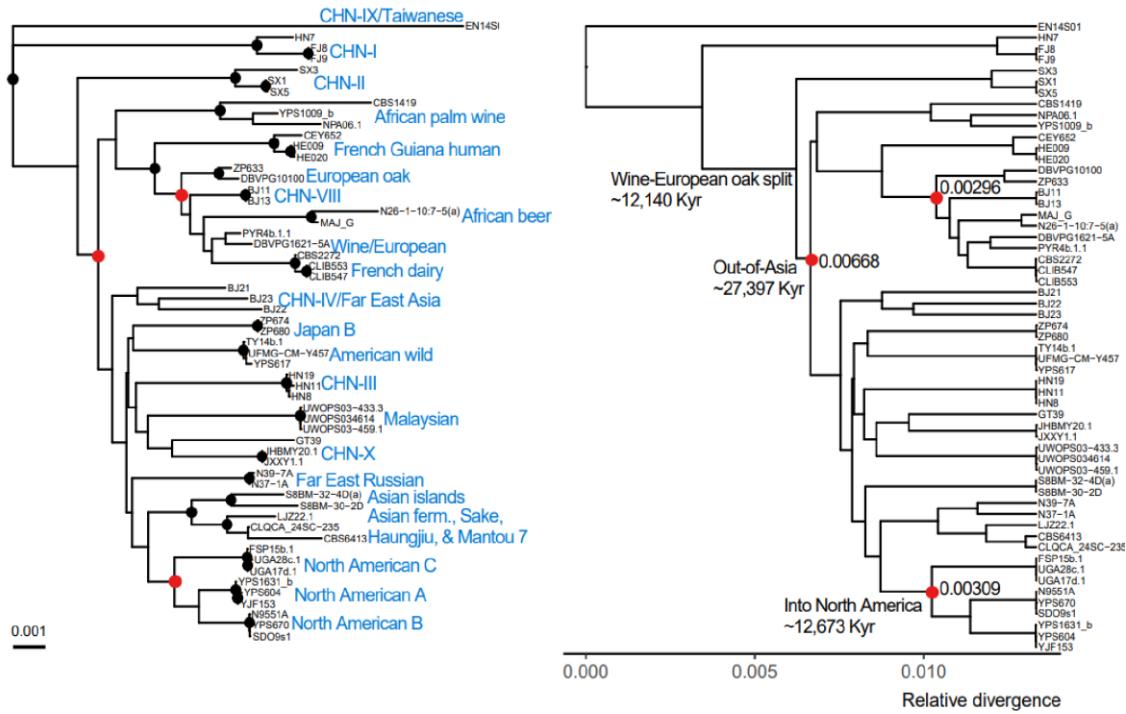




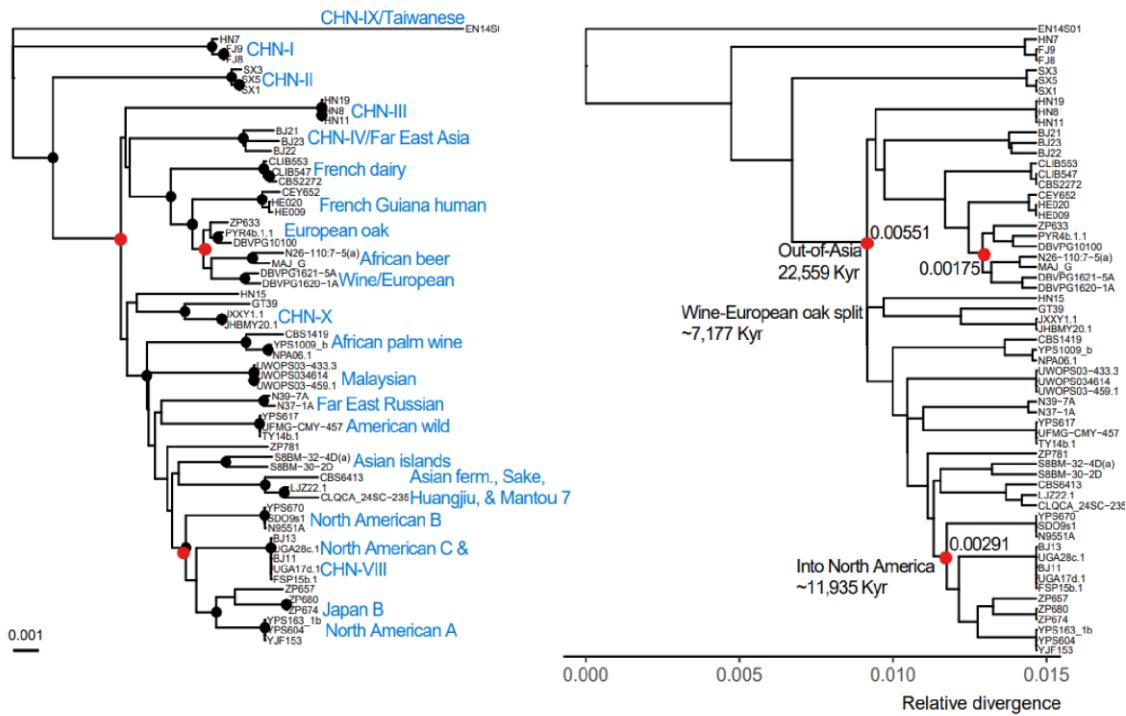


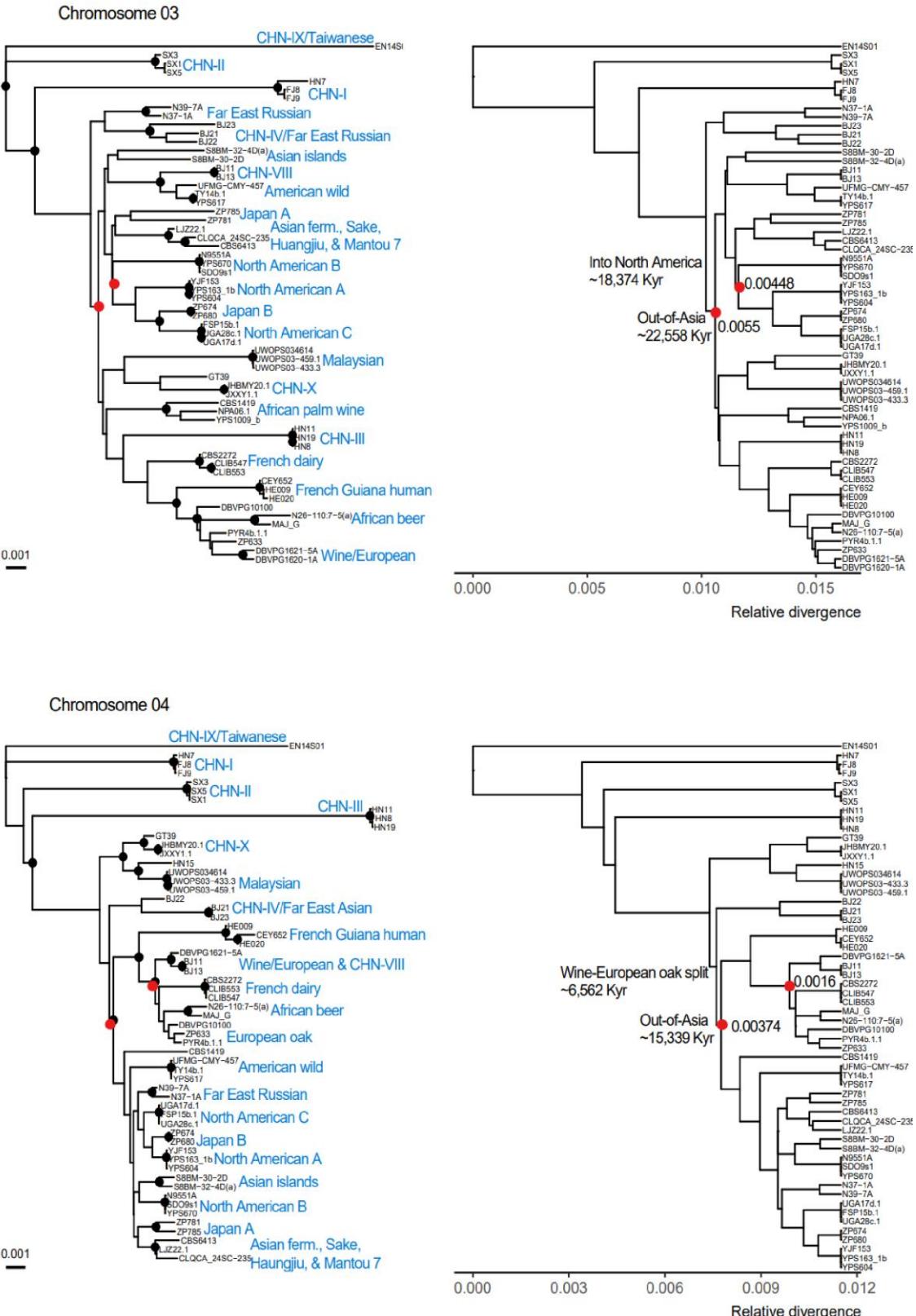
Supplemental Figure 2.11 in silico chromosome painting results of backbone phylogeny strains grouped by chromosome of strains that made it past quality filters for time divergence analysis (Table S5). Red boxes indicate 30-60 kb genomic regions selected for time divergence analysis when strains are assigned to their primary clade ($>50\%$). Diverged regions were not colored (white) and were defined as regions that had a maximum proportion of sites that differed by 0.003 from all other strains in the backbone phylogeny. Black colored regions indicate low coverage. Genomic regions that show equal genetic similarity to strains from multiple clades are colored yellow.

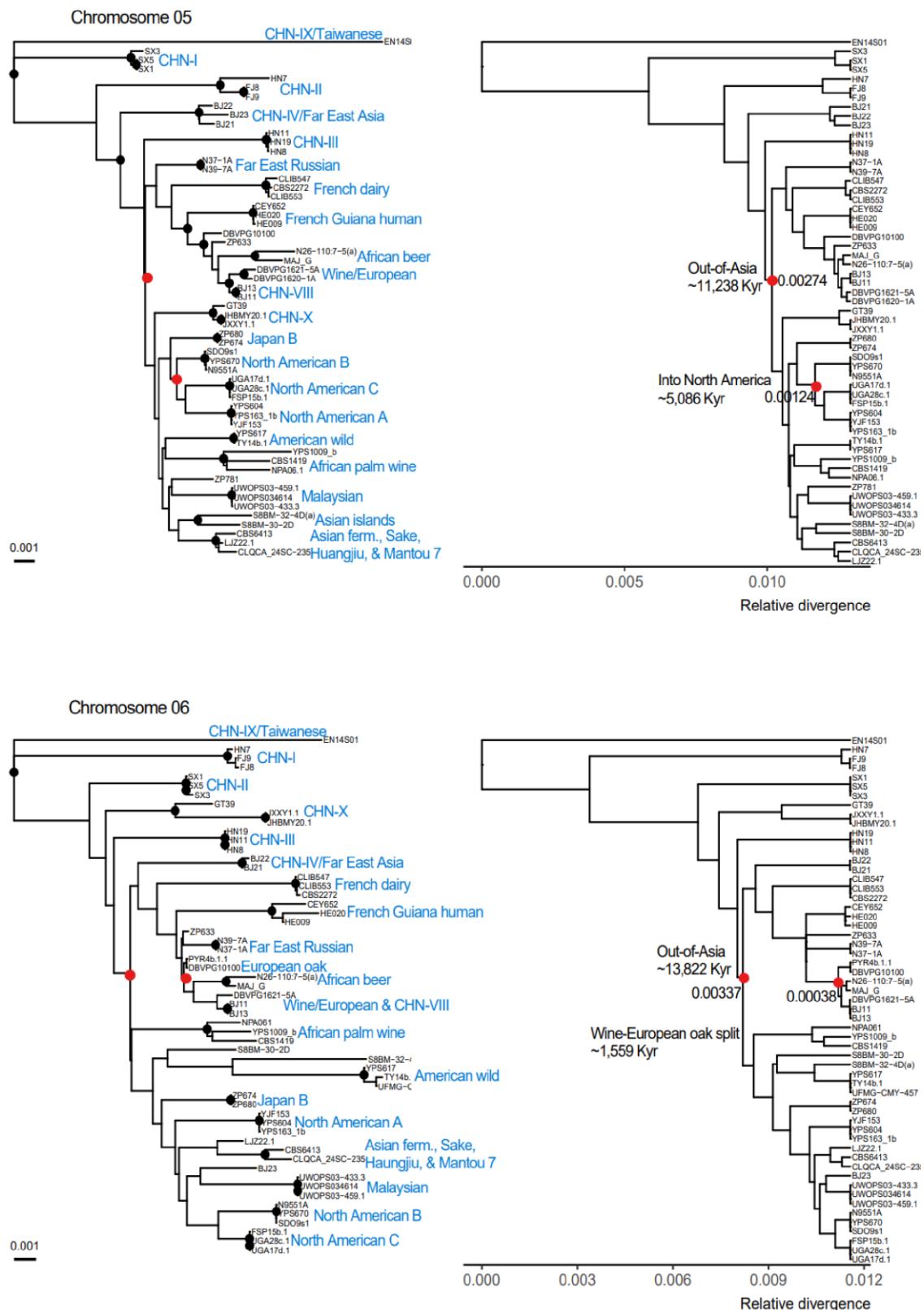
Chromosome 01

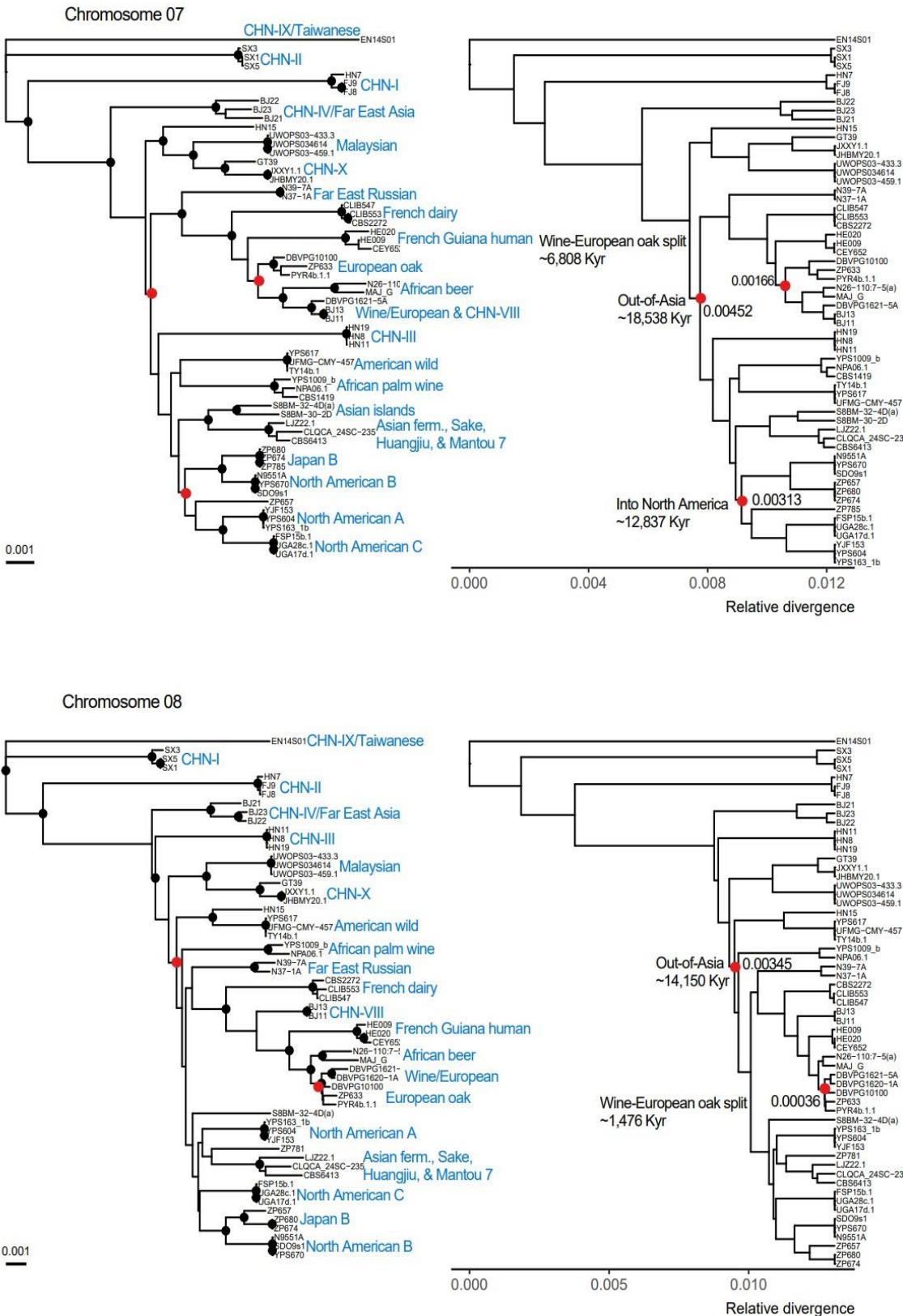


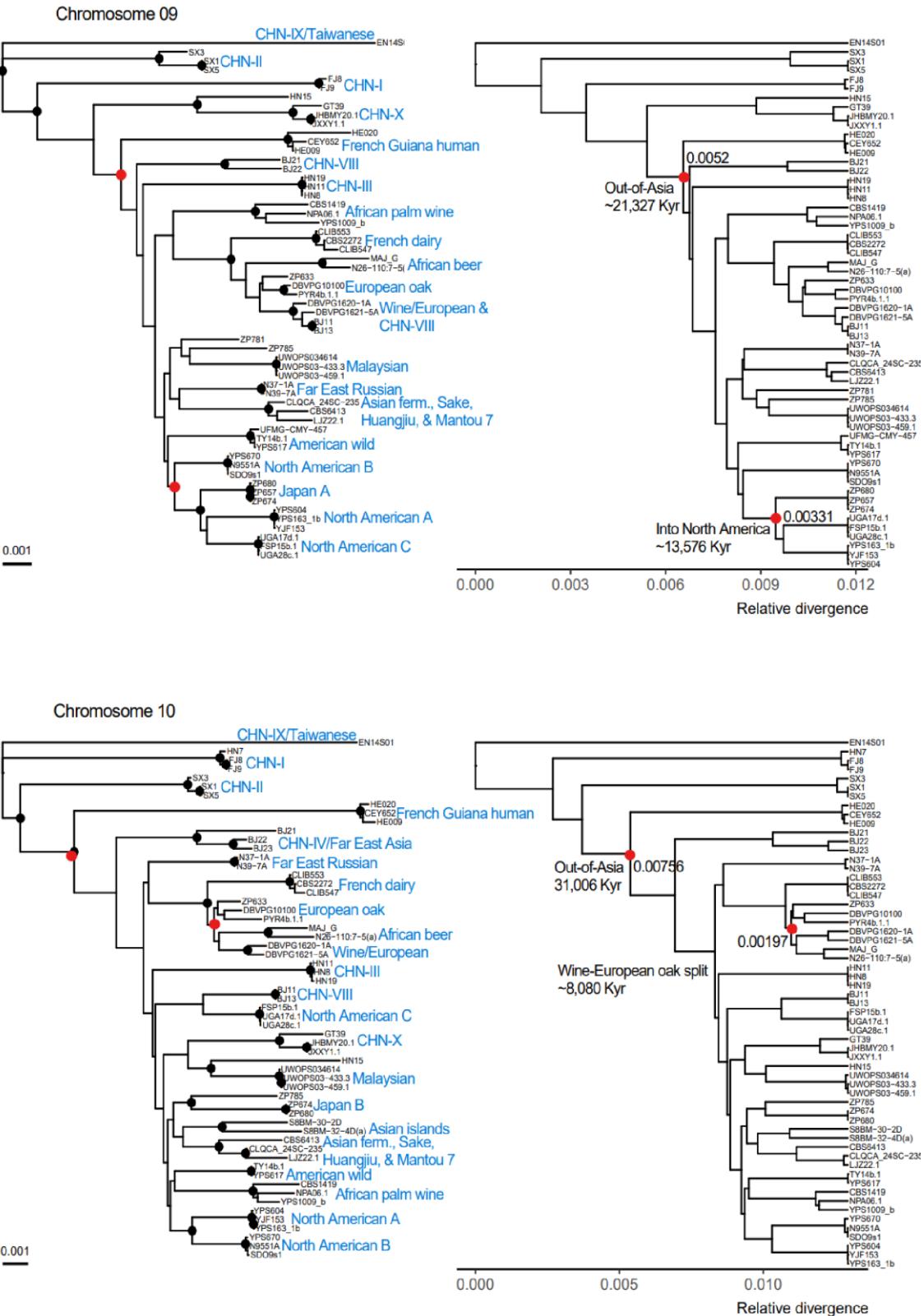
Chromosome 02

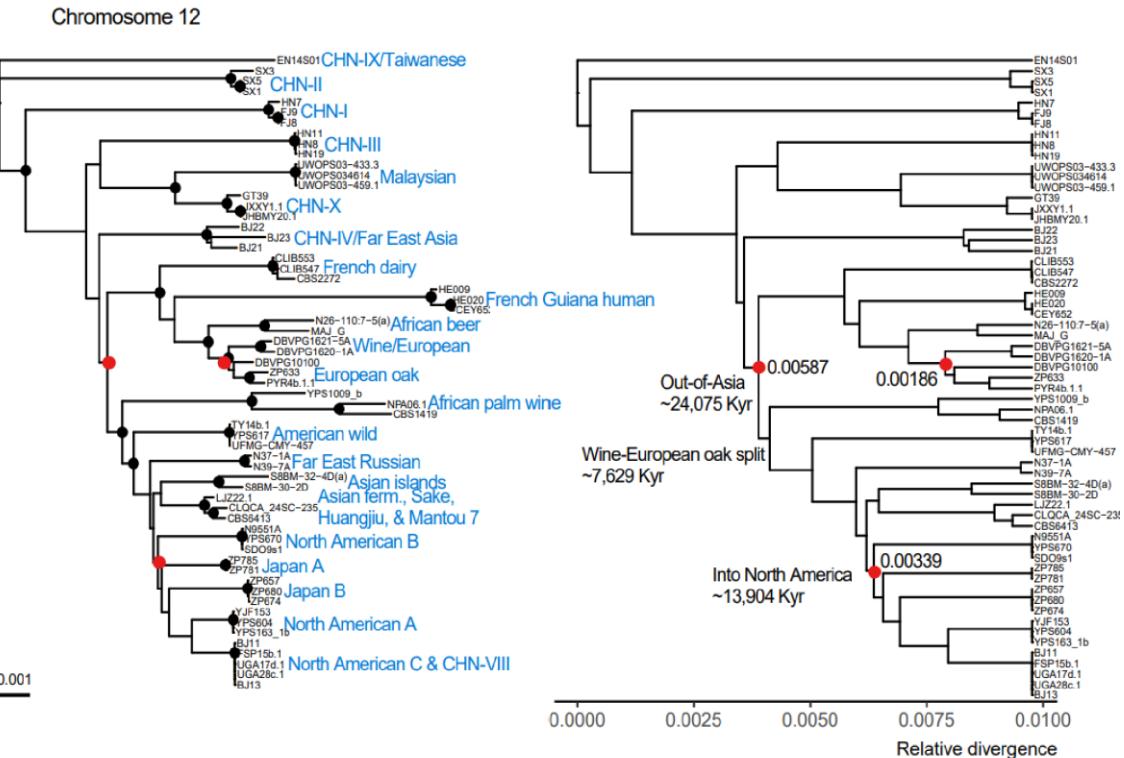
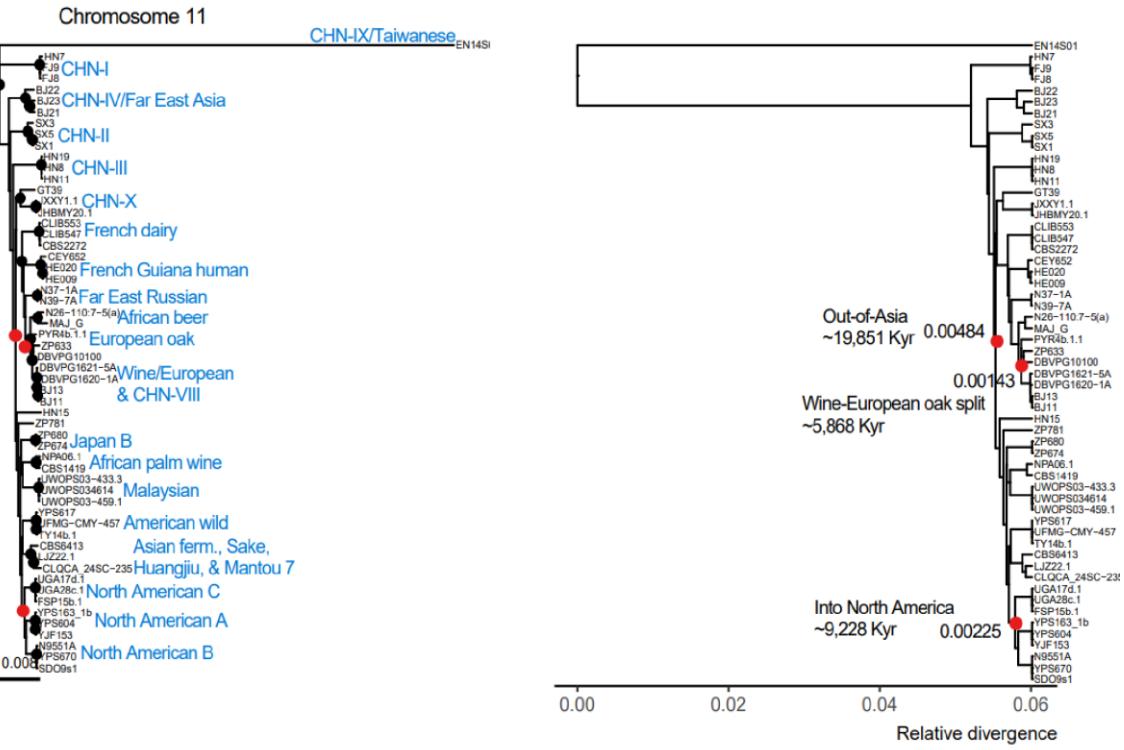


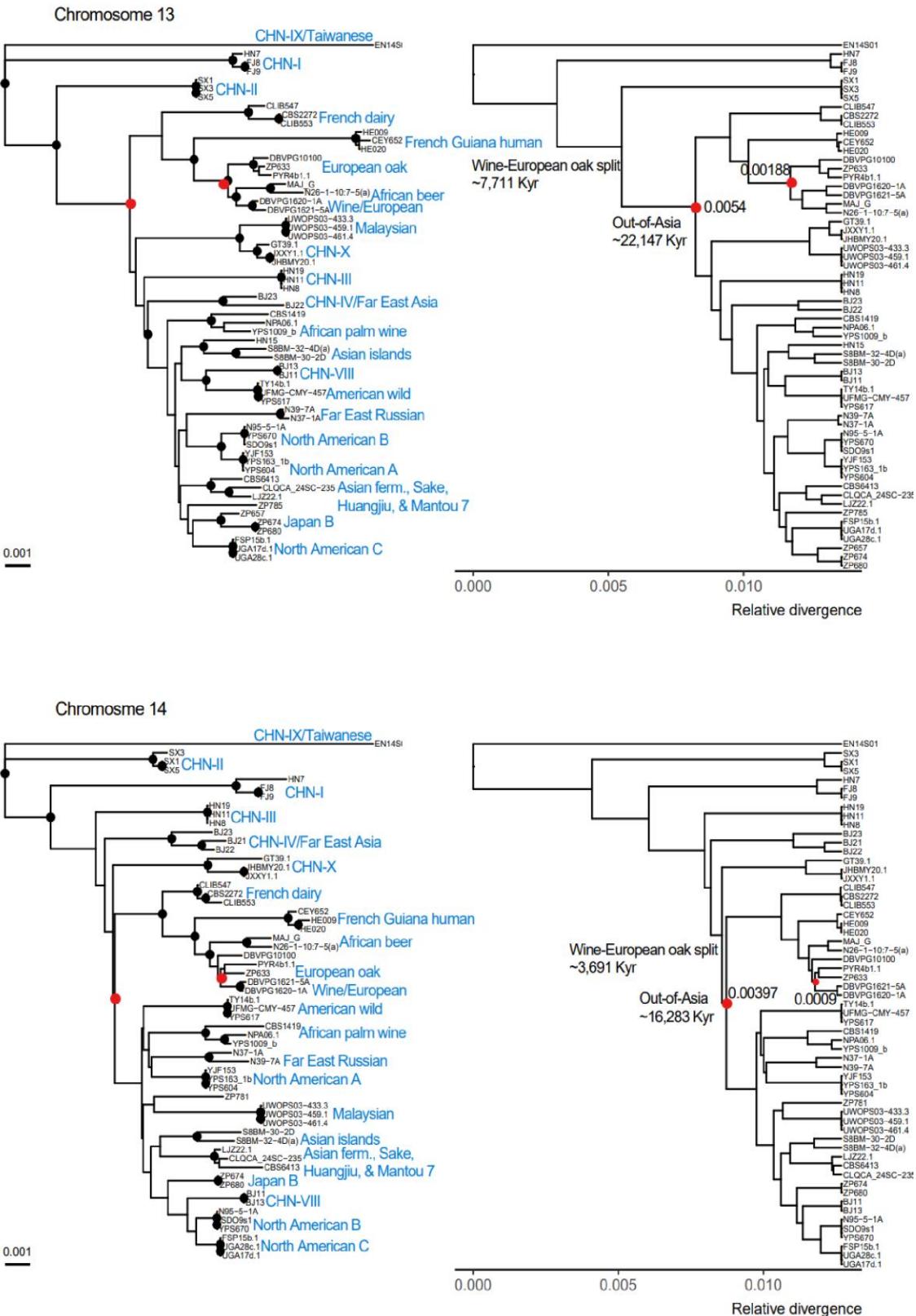


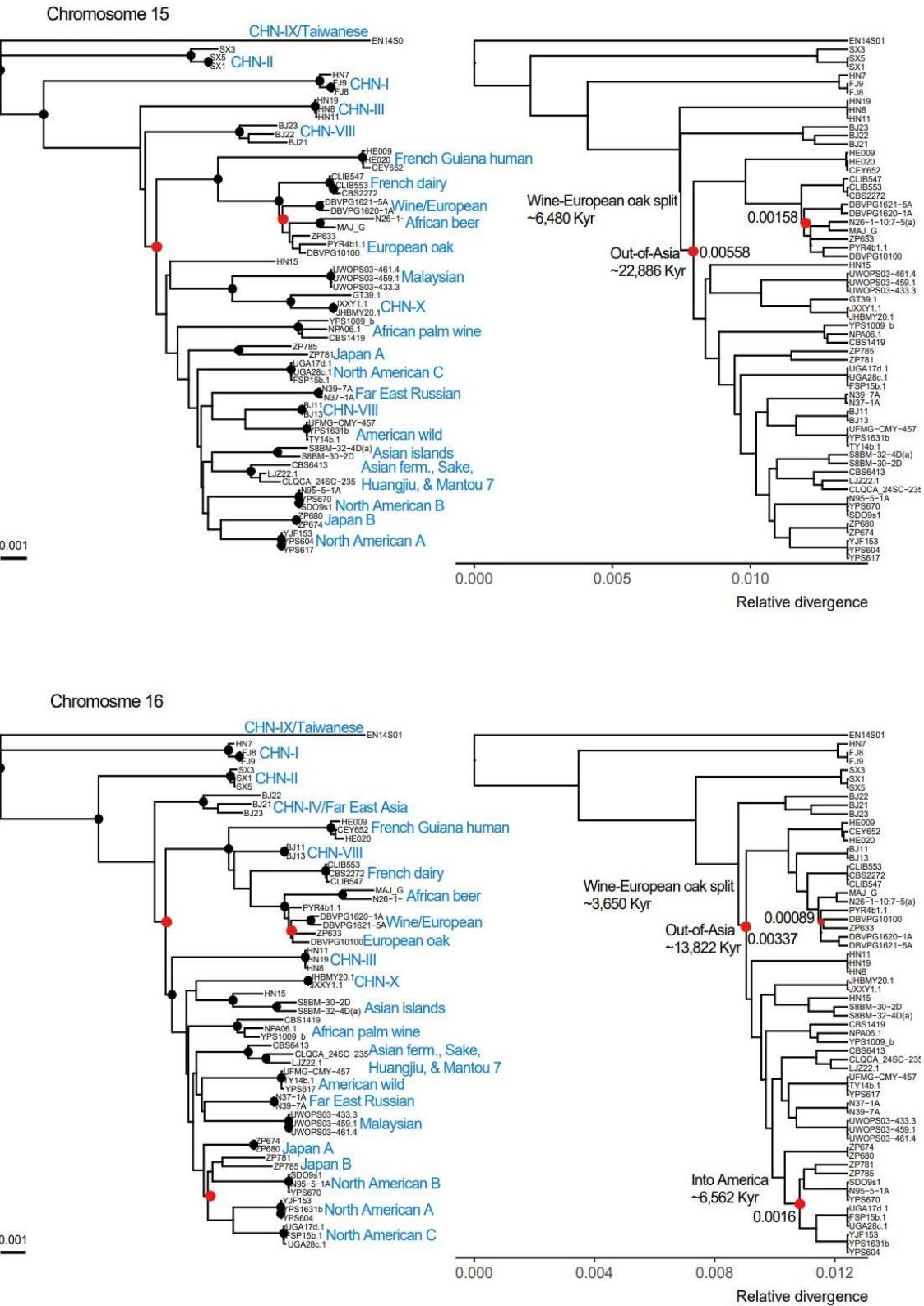












Supplemental Figure 2.12 Time divergence results of selected 60-90 kb genomic regions of backbone phylogeny strains for each chromosome. Phylogenetic trees on the left are neighborjoining trees of the selected genomic region. The tree was inferred using a Tamura-Nei substitution model (Tamura & Nei, 1993) with a gamma distribution using 100 bootstrap replicates. Black circles at nodes indicate bootstrap support >95%. Clades are labelled in blue text. Time calibrated trees are on the right and were estimated using the RelTime-ML option with default settings (Tamura et al., 2012). We used a CHN-IX/Taiwanese strain (EN14S01) as an outgroup to root the time tree. We calculated the time (T) since the most recent common ancestor (MRCA) in generations per year, $T_{MRCA} = k / \mu$ / generations per year, where k is the genetic distance to the MRCA of strains in the clade for that node and μ is the point mutation rate per bp. We used the mutation rate 1.84×10^{-10} from (Fay & Benavides, 2005) where they accounted for 82% of mutations being single base substitutions (Kang et al., 1992) for a mutation rate of 2.25×10^{-10} (Drake, 1991). Time divergence events are indicated with a red circle at a node for the following events: Out-of-Asia, Wine-European oak split, and North America/Japan split. We calculated time estimates for each chromosome from TMRCA estimates (Table S7). Time estimates were not calculated for events when clades did not form monophyletic groups.

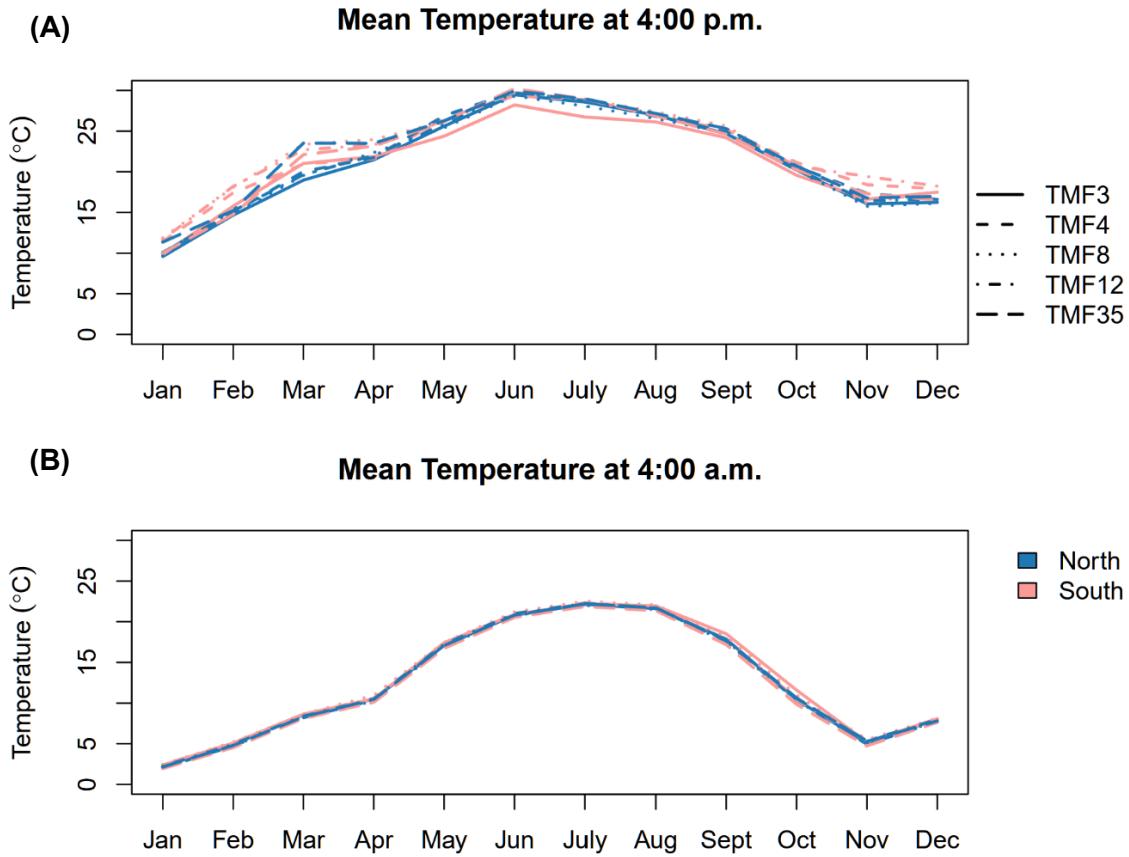
APPENDIX B

SUPPLEMENTAL MATERIALS FOR CHAPTER 3

Supplemental tables 3.1-3.5 can be found at

https://github.com/bensassonlab/data/blob/master/Pena_dissertation25/PenaCh3_Supple

[mentalTables.xlsx](#)



Supplemental Figure 3.1 Mean monthly temperature (°C) where tree hosts are likely to experience the (A) hottest (4:00 pm) and (B) coldest (4:00 am) temperatures during a 24-hour cycle from the north and south sides of *Quercus alba* at Thompson Mills Forest in Braselton, Georgia. The temperature was monitored between November 2021 and October 2022 using the iButtons data logger to record the daily temperature every four hours (8:00 am, 12:00 pm, 4:00 pm, 8:00 pm, 12:00 am, and 4:00 am) from five tree hosts (TMF3, TMF4, TMF8, TMF12, and TMF35). For clarity, we aligned months in calendar order starting with January.