DEVELOPING THE WINE YEAST *LACHANCEA THERMOTOLERANS*

AS A MODEL FOR EVOLUTIONARY GENOMICS

by

AUDREY KATHERINE WARD

(Under the Direction of Douda Bensasson)

ABSTRACT

How do species adapt to their environments? This question has been a major driver of evolutionary genetics research for a century. Developing cost-efficient models to address how traits are shaped by their environments and how species adapt to distinct ecological niches is critical as the impact of human influence on the environment becomes clearer. This dissertation develops the use of the wine yeast *Lachancea thermotolerans* as a model for population genomics by investigating the phylogenetics, population history, and phenotypic variation within the species. Firstly, I summarized population structure using over 300 genome sequences from strains with a broad range of geographic and environmental origins. Expansion of available genomic resources to include 90 more isolates from European and North American woodland habitats across two continents revealed several new tree-associated lineages. Additionally, I found evidence for recent gene flow between continents, providing a more complete view of population structure and the impact of environment on genetic variation. The addition of wild strains suggested that copy number variation previously associated with adaptation to domestic environments may be more prevalent across ecological and geographical origins than previously thought. Secondly, analysis of growth rates at a range of temperatures showed natural genetic variation within *L. thermotolerans*. Strains from one North American lineage grew at a

significantly lower rate than others at high temperatures. This suggests a single change within the species that appears maladaptive at high temperatures has occurred. The lack of adaptation seems surprising because there was natural genetic variation in growth rates among *L. thermotolerans* strains, suggesting that standing variation exists for adaptation to high temperature growth. Population genomic analyses require high-quality data to determine differences within a species, and the data used here did not show intraspecies contamination. This is important because, using simulated data, I found that contamination between 5 and 10% can alter phylogenetic tree topology and gene flow. Overall, the results presented here emphasize the importance of screening for intra-species contamination prior to phylogenetic or population genomic work and demonstrate the potential of *L. thermotolerans* as a model system to increase understanding of the genetic mechanisms of adaptation.

INDEX WORDS: yeast ecology, population genomics, environmental adaptation, phylogenetics

DEVELOPING THE WINE YEAST *LACHANCEA THERMOTOLERANS* AS A MODEL FOR EVOLUTIONARY GENOMICS

by

AUDREY KATHERINE WARD

BS, University of Alabama, 2019

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

DEVELOPING WINE YEAST *LACHANCEA THERMOTOLERANS*

AS A MODEL FOR EVOLUTIONARY GENOMICS

by

AUDREY KATHERINE WARD

Major Professor:	Douda Bensasson
Committee:	Kelly Dyer
	Zachary Lewis
	Michelle Momany
	Andrea Sweigart

Electronic Version Approved:

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

ACKNOWLEDGEMENTS

It has been a great privilege to be surrounded by a community of talented, dedicated, and supportive mentors and peers at the University of Georgia over the course of my PhD. Special thanks to my advisor, Douda Bensasson, for her guidance and mentorship in scientific thinking, evolutionary biology, population genomics, statistics, and bioinformatics. Thank you for taking me in and helping me navigate this journey – it has been an honor and a privilege to learn from you.

For their encouragement, advice, and patience, I would like to thank my advisory committee: Kelly Dyer, Zachary Lewis, Michelle Momany, and Andrea Sweigart. I would also like to thank Kelly Dyer and Robert Schmitz for their support and encouragement in their roles as department graduate coordinators. Thank you to Laramie Lemon and to Peggy Brickman for your mentorship and for serving as such an inspiration and a testament to the power of science education. Thank you to Ryan Earley, Laurie Stevison, and Cheng-Yu Li for encouraging and cultivating my enthusiasm for scientific research and for your consistent support throughout my career.

For scientific advice and engagement throughout my projects, thank you to all members of the Ecology/Diversity/Genetics/Evolution and Fungal Groups. To the larger community of genetics (and genetics-adjacent!) graduate students at UGA – thank you for your camaraderie and your perspectives over the years. Thank you to Susan White, without whom every graduate student in the department would be lost. I have had the distinct privilege to learn from, grow with, and work alongside some truly fabulous people: Paige Duffin, Andrew Duitsman, Katie Duval, Hannah Ericson, Trevor Haskins, Noah Legall, Holly McQueary, Hannah Nichols, Rida

Osman, Benjamin Phipps, and Daniel Shaw. Thank you for your friendship, scientific perspectives, advice, and at times, commiseration.

To Momany Lab members, past and present – thank you for your collaboration, friendship, support, and scientific insight. Special thanks to Bran Celia-Sanchez and Justina Stanislaw for listening to me ramble when I needed a moment to pull myself together.

Thank you to all my lab-mates, past and present, for your camaraderie, encouragement, and lively scientific and philosophical discussions: Eduardo Scopel Ferreira da Costa, Jacqueline Peña, Sydney McCall, Miranda McKibben, Rosemary Wills, Diana Ambrocio, Linda Habersham, Oliver Nemeth, Frema Owusu-Ansah, Grady Waple, and Domenic Won. Light is the burden shared among friends, indeed – I am so grateful to have worked with you and for your assistance and support.

To the Bio Babes (Theresa Erlenbach, Megan Motley, Miranda Tolbert and Madelaine Usey) – you are a constant source of inspiration and joy. To my stellar crew (Grace Beasley, Caroline Bloodworth, and Andrew Bonacci) and my intrepid eclectics (Sydney Bledsoe, Matthew Farnitano, Matthew Treaster, and Erika Turflinger) – thank you for reminding me not to take myself too seriously and to embrace life as an adventure. Elizabeth Coffin – thank you for making video blogs to keep me laughing and for your encouragement and support. Sydnie Lundy – it is such a blessing that a friendship begun through maternal scheming and Waffle House has persisted so beautifully for over a decade. Emma Fay and Nadia DelMedico – I am so glad to have you in my corner, ready for whatever comes. Thank you for having my back and for sharing your zeal for life with me whether we were across town or across the country.

I would like to thank my healthcare team and the University of Georgia Counseling and Psychiatric Center: graduate school is a long and arduous journey and access to mental

healthcare professionals was essential for my success. Similarly, I would like to thank the staff at the Science Learning Center Einstein Brother's Bagels for ready access to caffeine and a warm sandwich – as essential to my mental wellbeing as any medical care.

None of this would have been possible without the love and support of my family. Thank you to my family in love for welcoming me with open arms: Ms. Jackie, Mr. Steve, Jessa, Jay, Joe, and the entire Andrews and Sadle clans. Aunt Samantha, Uncle Dave Hall, Madie, Nora Jane, Bronwyn, Joey, Thalia, Nana, Aunt Kim, and Uncle Dave Ward: thank you for the walks in the park, the dinners, the sit-down chats, and everything else that's kept me sane. Thank you to my Mimi for emphasizing the importance of education and building a table big enough for everyone to have a seat. To my sisters, Meredith and Anne – thank you for keeping pinky-promises. Mama and Dad: thank you for encouraging me to pursue each new curiosity the world presented and for taking my panicked phone calls with grace and reassurance.

I only have this opportunity because of the courage and persistence of my late grandfather Dr. Bill Ward. Papa, thank you for everything. I miss you. I hope I make you proud.

To my dog, Angus: I am glad that you came into my life and apologize for tripping over you the day you came home, even if you cannot read this. Your shenanigans and willingness to wear a sweater have been a particular balm to me the past six months – you are a very good dog.

To my husband John: thank you for helping me manage a chapter of life that often felt unmanageable, for your selflessness, for your patience, for acing your duties as my outside world liaison and technology warden, and for your unwavering belief in me. I love you more today than I did yesterday, and I will love you even more tomorrow.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Adaptation, human influence, and microbes

Underlying mechanisms of genetic adaptation have been a focus of evolutionary biology since Darwin published *The Origin of Species* in 1859. Questions concerning species' adaptation to distinct ecological niches and the genetic architecture shaping those traits have driven the field for nearly 100 years (Orr, 2005). As the impact of the Anthropocene on the environment becomes more apparent, understanding ways in which lifeforms are influenced by their environments becomes even more critical. Domestication, or the mutualistic process by which one species controls the growth and reproduction of another for some benefit (Meyer & Purugganan, 2013; Purugganan, 2022), is frequently conflated with artificial selection by humans in colloquial parlance (Purugganan, 2022; Warwick et al., 2024). However, this understanding does not consider the influence humans have inadvertently had across the Tree of Life, especially on microbes (Friedrich et al., 2023; Warwick et al., 2024).

Adaptation of microorganisms to industrial environments began prior to the discovery of microbial life by Hooke and van Leeuwenhoek in the late 17th century; evidence suggests humans fermented food and beverages as early as the Neolithic period (Dupont et al., 2017; Marsit et al., 2017). This ancient relationship between microbes and humans, as well as their importance in multiple human-associated processes, is a major driver of microbial adaptation and niche specialization (Marsit et al., 2017; Villarreal et al., 2022). Fungi and yeasts have been widely used as food sources and in the production of victuals since at least 7000 BCE (Dupont et

al., 2017; Marsit et al., 2017; Peña et al., 2025). These organisms therefore offer unique insight into the means of genomic adaptation and especially human influence on these processes.

Using yeast to study genetics of adaptation

Yeasts in the fungal phylum *Ascomycota* (i.e., ascomycetes) are a diverse group that originated between 317-523 million years ago (mya). Species within this phylum have divergence times comparable to that between humans and roundworms (Shen et al., 2018). Within the ascomycetes, evolution of phenotypes in the subphylum *Saccharomycotina* is driven by gain and loss of metabolic traits allowing yeasts to adapt to a variety of specialized niches within a wide number of habitats (Clowers, Heilberger, et al., 2015; Gonçalves et al., 2020; Harrison et al., 2024; Mozzachiodi et al., 2022; Opulente et al., 2018; Rosa & Gábor, 2006; Samarasinghe et al., 2021; Shen et al., 2018). The species *Saccharomyces cerevisiae* is well known as a eukaryotic model system due to its tractability, large genetic toolkit, and collaborative community of biologists (Botstein & Fink, 2011).

Saccharomyces cerevisiae has a long history of human association, appearing in fermented foodstuffs from China dating to approximately 7000 BCE (Marsit et al., 2017). Genetic differences between wild and domestic lineages of *S. cerevisiae* have shaped the population structure of the species (Liti et al., 2009; Marsit et al., 2017; Peña et al., 2025). Even between domestic populations, there are significant differences – beer and bakery lineages are genetically diverse with polyphyletic structure, while sake and wine lineages are each monophyletic (Gallone et al., 2016; Peter et al., 2018). Aneuploidy variation also differs between domestic populations - there is high aneuploidy variation within the sake lineage and low aneuploidy variation within the wine lineage (Scopel et al., 2021). There is evidence that domestication of

beer lineages greatly impacted the life cycle of *S. cerevisiae*, changing growth patterns under stress and impairing or completely removing the sexual reproductive cycle (De Chiara et al., 2022). Beyond domestication, there is evidence for convergent evolution in *S. cerevisiae* as they transitioned between woodland and fruit habitats in the United States and Europe (Almeida et al., 2017; Clowers, Heilberger, et al., 2015; Clowers, Will, et al., 2015). This suggests independent populations within yeast species are effective models for studying processes driving adaptation and convergent evolution.

Recent studies focusing on the roles of non-model yeasts in production of beer and wine (Binati et al., 2020; Canonico et al., 2019; Capece et al., 2018; Domizio et al., 2016; Feng et al., 2020; Galaz & Franco, 2023; Giannakou et al., 2020; Jolly et al., 2014; Masneuf-Pomarede et al., 2016; Mateus et al., 2020; Molinet & Cubillos, 2020; Postigo et al., 2022; Vejarano & Gil-Calderón, 2021), other industrial processes (Fernández-Pacheco et al., 2023; Giannakou et al., 2020; Mukherjee et al., 2017; Pretscher et al., 2018), and their clinical and ecological significance (Agarbati et al., 2020; Freel et al., 2015, 2016; Harrison et al., 2024; Kunyeit et al., 2024; Schikora-Tamarit & Gabaldón, 2022; Vitanović et al., 2019; Yilmazer Aktar et al., 2023) emphasize ways in which broadening the scope of yeast research may impact our understanding of evolutionary and population dynamics in eukaryotes. As whole genome resources for non-model yeast species continue to become readily available, they provide a means for addressing questions about genetic factors shaping adaptation (Libkind et al., 2020; Peter & Schacherer, 2016). One species of particular interest for its use in beer (Canonico et al., 2019; Capece et al., 2018; Domizio et al., 2016; Molinet & Cubillos, 2020; Postigo et al., 2022; Zdaniewicz et al., 2020), wine (Benito, 2018; Gatto et al., 2020; Morata et al., 2018; Porter et al., 2019; Vilela,

2018), and its potential as a biocontrol agent (Mioranza et al., 2021; Nally et al., 2018; Zeidan et al., 2018) is *Lachancea thermotolerans*.

Non-model yeast Lachancea thermotolerans as a system to study adaptation

The yeast *L. thermotolerans*, type species for the genus *Lachancea*, was previously part of the *Kluyveromyces* yeasts (Lachance & Kurtzman, 2011). The genus is estimated to have diverged approximately 125-150 mya (Shen et al., 2018), meaning that the genetic distance between *L. thermotolerans* and *S. cerevisiae* is roughly equivalent to that between humans and starfish (Shen et al., 2018). *Lachancea thermotolerans* can be found on a variety of substrates, including tree and vine bark, flower, and fruit (Osburn et al., 2018; Porter et al., 2019; Robinson et al., 2016). Further, it has been isolated on every continent, apart from Antarctica, and from a wide variety of climates. It has been noted for its production of lactic acid during fermentation (Gatto et al., 2020; Morata et al., 2018) and for possibly providing a biological method to deal with the effects of climate change's impact on wine flavor profiles (Morata et al., 2018; Vaquero et al., 2020; Vicente et al., 2023). Previous work with mitochondrial and microsatellite data showed that *L. thermotolerans* live in geographically structured populations with distinct populations associated with wine and trees (Banilas et al., 2016; Freel et al., 2014; Hranilovic et al., 2018). *Lachancea thermotolerans* from oenological environments display metabolic traits suggesting that wine-association lineages are domesticated (Hranilovic et al., 2018; Vicente et al., 2025). The geographic structure and presence of domestication traits within the species makes it an interesting model for investigating i) local adaptation and ii) differences in drivers of adaptation between domesticated and wild environments; however, previous studies using

whole-genome sequencing have focused primarily on European winemaking strains (Vicente et al., 2025).

Saccharomyces cerevisiae regularly co-occurs with *L. thermotolerans* in both natural and winemaking environments, where mixed fermentations of the species are frequently used to improve wine quality (Binati et al., 2020; Gobbi et al., 2013). Domestication of *S. cerevisiae* is associated with the expansion of viticulture within the Mediterranean Basin (Almeida et al., 2015), the same geographic origin as the lineage of *L. thermotolerans* showing hallmarks of domestication (Banilas et al., 2016; Hranilovic et al., 2018; Vicente et al., 2025). Further, causal genes for traits exhibited by this domestic lineage show significant homology with genes also characterized as part of the *S. cerevisiae* domestication suite (Liti et al., 2009; Vicente et al., 2025). These ecological and genetic similarities, along with the large genetic distance between the two, make *L. thermotolerans* and *S. cerevisiae* a compelling model system for understanding repeated evolution (Cerca, 2023). Further investigation into the population history, genetic architecture, and phenotypic profiles of *L. thermotolerans* is needed as the basis for future investigations into processes of parallel evolution between the two species.

Thesis outline

The goal of this dissertation is to use phylogenetics and population genomics approaches to study phylogeography and population structure in *L. thermotolerans* and identify thermal growth profiles to assess whether local adaptation may be occurring within the species.

Chapter 2 presents an expanded understanding of the phylogeography and population structure of *L. thermotolerans*. Here, I expand available whole-genome resources for wild forest populations within *L. thermotolerans* and use publicly available sequences (Vicente et al., 2025;

Xia et al., 2017) to identify 12 genetically distinct lineages within the species. I additionally find evidence for historical migration and gene flow between continents likely driven by human migration events. A preliminary finding shows that increased copy number of genes associated with the success of *L. thermotolerans* in winemaking environments are present in wild forest populations. This work provides a more complete view of the population structure and impact of human-associated environments on genetic variation within *L. thermotolerans*.

In Chapter 3, I investigated the natural variation in the growth rates of *L. thermotolerans* strains at different temperatures and explored factors that could affect a strain's growth within these conditions. Using a combination of high-throughput phenotyping (via collaboration with the Gasch lab at the University of Wisconsin – Madison), phylogenetics, and linear and non-parametric modeling with a subset of 58 strains from 8 distinct genetic lineages (identified in chapter 2), I found that strains from North American lineages grow at a significantly lower rate than those from other lineages at high temperatures. This was the only statistically significant difference between lineages, suggesting a single change in growth rate as a response to temperature change has occurred in the species. This is perhaps surprising because I also observed standing genetic variation in high temperature growth rates among closely related strains within *L. thermotolerans*.

Chapter 4 assesses the impact of low-level intra-species contamination on genome data and downstream population genomic analyses. In collaboration with Bensasson lab member Eduardo F. C. Scopel and the Momany lab in the Fungal Biology Group at the University of Georgia, I found that while contaminated genomes appear to produce good quality genome data, 5-10% intraspecies contamination can significantly alter tree topologies and estimations of admixture in multiple fungal species. A previous study in prokaryotes has suggested phylogenetic analyses are

more robust to contamination, with a threshold between 40 and 50% contamination before significant impact (Pightling et al., 2019). This work emphasizes the importance of screening for intra-species contamination and presents the use of B-allele frequency plots to assess potential contamination prior to downstream analyses. All sequences used in analyses within this dissertation were screened before use.

In Chapter 5, I assess the results of chapters two through four at a broader level and discuss future directions for analyses using these data. My dissertation shows the potential of *L. thermotolerans* as a population genomics model to further understand domestication and the potential for local adaptation in natural environments. Additional research into this species on both ecological and genetic scales is essential to more completely develop a cost-effective and tractable eukaryotic model for repeated evolution.

REFERENCES

Agarbati, A., Canonico, L., Marini, E., Zannini, E., Ciani, M., & Comitini, F. (2020). Potential Probiotic Yeasts Sourced from Natural Environmental and Spontaneous Processed Foods. *Foods*, *9*(3), 287. https://doi.org/10.3390/foods9030287

Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2017). Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites. *Molecular Ecology*, *26*(7), 2167–2182. https://doi.org/10.1111/mec.14071

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

Banilas, G., Sgouros, G., & Nisiotou, A. (2016). Development of microsatellite markers for *Lachancea thermotolerans* typing and population structure of wine-associated isolates. *Microbiological Research*, *193*, 1–10. https://doi.org/10.1016/j.micres.2016.08.010

Benito, S. (2018). The impacts of *Lachancea thermotolerans* yeast strains on winemaking. *Applied Microbiology and Biotechnology*, *102*(16), 6775–6790. https://doi.org/10.1007/s00253-018-9117-z

Binati, R. L., Lemos Junior, W. J. F., Luzzini, G., Slaghenaufi, D., Ugliano, M., & Torriani, S. (2020). Contribution of non-*Saccharomyces* yeasts to wine volatile and sensory diversity: A study on *Lachancea thermotolerans*, *Metschnikowia spp*. and *Starmerella bacillaris* strains isolated in Italy. *International Journal of Food Microbiology*, *318*, 108470. https://doi.org/10.1016/j.ijfoodmicro.2019.108470

Botstein, D., & Fink, G. R. (2011). Yeast: An Experimental Organism for 21st Century Biology. *Genetics*, *189*(3), 695–704. https://doi.org/10.1534/genetics.111.130765

Canonico, L., Galli, E., Ciani, E., Comitini, F., & Ciani, M. (2019). Exploitation of Three Non-Conventional Yeast Species in the Brewing Process. *Microorganisms*, *7*(1). https://doi.org/10.3390/microorganisms7010011

Capece, A., Romaniello, R., Siesto, G., & Romano, P. (2018). Conventional and Non-Conventional Yeasts in Beer Production. *Fermentation*, *4*(2), 38. https://doi.org/10.3390/fermentation4020038

Cerca, J. (2023). Understanding natural selection and similarity: Convergent, parallel and repeated evolution. *Molecular Ecology*, *32*(20), 5451–5462. https://doi.org/10.1111/mec.17132

Clowers, K. J., Heilberger, J., Piotrowski, J. S., Will, J. L., & Gasch, A. P. (2015). Ecological and Genetic Barriers Differentiate Natural Populations of *Saccharomyces cerevisiae*. *Molecular Biology and Evolution*, *32*(9), 2317–2327. https://doi.org/10.1093/molbev/msv112

Clowers, K. J., Will, J. L., & Gasch, A. P. (2015). A unique ecological niche fosters hybridization of oak-tree and vineyard isolates of *Saccharomyces cerevisiae*. *Molecular Ecology*, *24*(23), 5886–5898. https://doi.org/10.1111/mec.13439

De Chiara, M., Barré, B. P., Persson, K., Irizar, A., Vischioni, C., Khaiwal, S., Stenberg, S., Amadi, O. C., Žun, G., Doberšek, K., Taccioli, C., Schacherer, J., Petrovič, U., Warringer, J., & Liti, G. (2022). Domestication reprogrammed the budding yeast life cycle. *Nature Ecology & Evolution*, *6*(4), 448–460. https://doi.org/10.1038/s41559-022-01671-9

Domizio, P., House, J. F., Joseph, C. M. L., Bisson, L. F., & Bamforth, C. W. (2016). *Lachancea thermotolerans* as an alternative yeast for the production of beer. *Journal of the Institute of Brewing*, *122*(4), 599–604. https://doi.org/10.1002/jib.362

Dupont, J., Dequin, S., Giraud, T., Le Tacon, F., Marsit, S., Ropars, J., Richard, F., & Selosse, M.-A. (2017). Fungi as a Source of Food. *Microbiology Spectrum*, *5*(3), 5.3.09. https://doi.org/10.1128/microbiolspec.FUNK-0030-2016

Feng, L., Wang, J., Ye, D., Song, Y., Qin, Y., & Liu, Y. (2020). Yeast population dynamics during spontaneous fermentation of icewine and selection of indigenous *SACCHAROMYCES*

CEREVISIAE strains for the winemaking in QILIAN, CHINA. *Journal of the Science of Food and Agriculture*, *100*(15), 5385–5394. https://doi.org/10.1002/jsfa.10588

Fernández-Pacheco, P., Ramos Monge, I. M., Poveda, J. M., Díaz-Maroto, M. C., & Arévalo-Villena, M. (2023). Use of probiotic yeasts with biocontrol activity for fermentation of ewe's milk. *Journal of the Science of Food and Agriculture*, jsfa.12394. https://doi.org/10.1002/jsfa.12394

Freel, K. C., Charron, G., Leducq, J.-B., Landry, C. R., & Schacherer, J. (2015). *Lachancea quebecensis sp. Nov.*, a yeast species consistently isolated from tree bark in the Canadian province of Québec. *International Journal of Systematic and Evolutionary Microbiology*, *65*(Pt_10), 3392–3399. https://doi.org/10.1099/ijsem.0.000426

Freel, K. C., Friedrich, A., Hou, J., & Schacherer, J. (2014). Population Genomic Analysis Reveals Highly Conserved Mitochondrial Genomes in the Yeast Species *Lachancea thermotolerans*. *Genome Biology and Evolution*, *6*(10), 2586–2594. https://doi.org/10.1093/gbe/evu203

Freel, K. C., Friedrich, A., Sarilar, V., Devillers, H., Neuvéglise, C., & Schacherer, J. (2016). Whole-Genome Sequencing and Intraspecific Analysis of the Yeast Species *Lachancea quebecensis*. *Genome Biology and Evolution*, *8*(3), 733–741. https://doi.org/10.1093/gbe/evv262

Friedrich, A., Gounot, J.-S., Tsouris, A., Bleykasten, C., Freel, K., Caradec, C., & Schacherer, J. (2023). Contrasting Genomic Evolution Between Domesticated and Wild *Kluyveromyces lactis* Yeast Populations. *Genome Biology and Evolution*, *15*(2), evad004. https://doi.org/10.1093/gbe/evad004

Galaz, V., & Franco, W. (2023). *Lachancea quebecensis* a Novel Isolate for the Production of Craft Beer. *Foods*, *12*(18), 3347. https://doi.org/10.3390/foods12183347

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

Gatto, V., Binati, R. L., Lemos Junior, W. J. F., Basile, A., Treu, L., de Almeida, O. G. G., Innocente, G., Campanaro, S., & Torriani, S. (2020). New insights into the variability of lactic acid production in *Lachancea thermotolerans* at the phenotypic and genomic level. *Microbiological Research*, *238*, 126525. https://doi.org/10.1016/j.micres.2020.126525

Giannakou, K., Cotterrell, M., & Delneri, D. (2020). Genomic Adaptation of *Saccharomyces* Species to Industrial Environments. *Frontiers in Genetics*, *11*. https://doi.org/10.3389/fgene.2020.00916

Gobbi, M., Comitini, F., Domizio, P., Romani, C., Lencioni, L., Mannazzu, I., & Ciani, M. (2013). *Lachancea thermotolerans* and *Saccharomyces cerevisiae* in simultaneous and sequential co-fermentation: A strategy to enhance acidity and improve the overall quality of wine. *Food Microbiology*, *33*(2), 271–281. https://doi.org/10.1016/j.fm.2012.10.004

Gonçalves, P., Gonçalves, C., Brito, P. H., & Sampaio, J. P. (2020). The *Wickerhamiella/Starmerella* clade—A treasure trove for the study of the evolution of yeast metabolism. *Yeast*, *37*(4), 313–320. https://doi.org/10.1002/yea.3463

Harrison, M.-C., Opulente, D. A., Wolters, J. F., Shen, X.-X., Zhou, X., Groenewald, M., Hittinger, C. T., Rokas, A., & LaBella, A. L. (2024). Exploring *Saccharomycotina* Yeast

Ecology Through an Ecological Ontology Framework. *Yeast*, *41*(n/a), 615–628. https://doi.org/10.1002/yea.3981

Hranilovic, A., Gambetta, J. M., Schmidtke, L., Boss, P. K., Grbin, P. R., Masneuf-Pomarede, I., Bely, M., Albertin, W., & Jiranek, V. (2018). Oenological traits of *Lachancea thermotolerans* show signs of domestication and allopatric differentiation. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-33105-7

Jolly, N. P., Varela, C., & Pretorius, I. S. (2014). Not your ordinary yeast: Non- *Saccharomyces* yeasts in wine production uncovered. *FEMS Yeast Research*, *14*(2), 215–237. https://doi.org/10.1111/1567-1364.12111

Kunyeit, L., Rao, R. P., & Anu-Appaiah, K. A. (2024). Yeasts originating from fermented foods, their potential as probiotics and therapeutic implication for human health and disease. *Critical Reviews in Food Science and Nutrition*, *64*(19), 6660–6671. https://doi.org/10.1080/10408398.2023.2172546

Lachance, M.-A., & Kurtzman, C. P. (2011). *Lachancea* Kurtzman (2003). In *Yeasts: A Taxonomic Study* (5th ed., Vols. 1–3, pp. 511–519). Elsevier. doi.org/10.1016/B978-0-444-52149-1.00041-0

Libkind, D., Peris, D., Cubillos, F. A., Steenwyk, J. L., Opulente, D. A., Langdon, Q. K., Rokas, A., & Hittinger, C. T. (2020). Into the wild: New yeast genomes from natural environments and new tools for their analysis. *FEMS Yeast Research*, *20*(2), foaa008. https://doi.org/10.1093/femsyr/foaa008

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, 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

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

Masneuf-Pomarede, I., Bely, M., Marullo, P., & Albertin, W. (2016). The Genetics of Non-conventional Wine Yeasts: Current Knowledge and Future Challenges. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.01563

Mateus, D., Sousa, S., Coimbra, C., S. Rogerson, F., & Simões, J. (2020). Identification and Characterization of Non-*Saccharomyces* Species Isolated from Port Wine Spontaneous Fermentations. *Foods*, *9*(2), 120. https://doi.org/10.3390/foods9020120

Meyer, R. S., & Purugganan, M. D. (2013). Evolution of crop species: Genetics of domestication and diversification. *Nature Reviews Genetics*, *14*(12), 840–852. https://doi.org/10.1038/nrg3605

Mioranza, T. M., Miamoto, A., Mattos, A. P., Rissato, B. B., Dias-Arieira, C. R., & Schwan-Estrada, K. R. F. (2021). *Lachancea thermotolerans* acts as a resistance inducer in soybean infected with *Meloidogyne incognita*. *European Journal of Plant Pathology*, *159*(3), 511–523. https://doi.org/10.1007/s10658-020-02178-4

Molinet, J., & Cubillos, F. A. (2020). Wild Yeast for the Future: Exploring the Use of Wild Strains for Wine and Beer Fermentation. *Frontiers in Genetics*, *11*, 589350. https://doi.org/10.3389/fgene.2020.589350

Morata, A., Loira, I., Tesfaye, W., Bañuelos, M. A., González, C., & Suárez Lepe, J. A. (2018). *Lachancea thermotolerans* Applications in Wine Technology. *Fermentation*, *4*(3), 53. https://doi.org/10.3390/fermentation4030053

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*, *39*(1–2), 4–24. https://doi.org/10.1002/yea.3699

Mukherjee, V., Radecka, D., Aerts, G., Verstrepen, K. J., Lievens, B., & Thevelein, J. M. (2017). Phenotypic landscape of non-conventional yeast species for different stress tolerance traits desirable in bioethanol fermentation. *Biotechnology for Biofuels*, *10*(1), 216. https://doi.org/10.1186/s13068-017-0899-5

Nally, M. C., Ponsone, M. L., Pesce, V. M., Toro, M. E., Vazquez, F., & Chulze, S. (2018). Evaluation of behaviour of *Lachancea thermotolerans* biocontrol agents on grape fermentations. *Letters in Applied Microbiology*, *67*(1), 89–96. https://doi.org/10.1111/lam.13001

Opulente, D. A., Rollinson, E. J., Bernick-Roehr, C., Hulfachor, A. B., Rokas, A., Kurtzman, C. P., & Hittinger, C. T. (2018). Factors driving metabolic diversity in the budding yeast subphylum. *BMC Biology*, *16*(1), 26. https://doi.org/10.1186/s12915-018-0498-3

Orr, H. A. (2005). The genetic theory of adaptation: A brief history. *Nature Reviews Genetics*, *6*(2), 119–127. https://doi.org/10.1038/nrg1523

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

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. https://doi.org/10.1111/mec.17669

Peter, J., De Chiara, M., 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), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Peter, J., & Schacherer, J. (2016). Population genomics of yeasts: Towards a comprehensive view across a broad evolutionary scale. *Yeast*, *33*(3), 73–81. https://doi.org/10.1002/yea.3142

Pightling, A. W., Pettengill, J. B., Wang, Y., Rand, H., & Strain, E. (2019). Within-species contamination of bacterial whole-genome sequence data has a greater influence on clustering analyses than between-species contamination. *Genome Biology*, *20*(1), 286. https://doi.org/10.1186/s13059-019-1914-x

Porter, T. J., Divol, B., & Setati, M. E. (2019). *Lachancea* yeast species: Origin, biochemical characteristics and oenological significance. *Food Research International*, *119*, 378–389. https://doi.org/10.1016/j.foodres.2019.02.003

Postigo, V., Sánchez, A., Cabellos, J. M., & Arroyo, T. (2022). New Approaches for the Fermentation of Beer: Non-*Saccharomyces* Yeasts from Wine. *Fermentation*, *8*(6), 280. https://doi.org/10.3390/fermentation8060280

Pretscher, J., Fischkal, T., Branscheidt, S., Jäger, L., Kahl, S., Schlander, M., Thines, E., & Claus, H. (2018). Yeasts from Different Habitats and Their Potential as Biocontrol Agents. *Fermentation*, *4*(2), 31. https://doi.org/10.3390/fermentation4020031

Purugganan, M. D. (2022). What is domestication? *Trends in Ecology & Evolution*, *37*(8), 663–671. https://doi.org/10.1016/j.tree.2022.04.006

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

Rosa, C., & Gábor, P. (Eds.). (2006). *Biodiversity and ecophysiology of yeasts*. Springer.

Samarasinghe, H., Lu, Y., Aljohani, R., Al-Amad, A., Yoell, H., & Xu, J. (2021). Global patterns in culturable soil yeast diversity. *iScience*, *24*(10), 103098. https://doi.org/10.1016/j.isci.2021.103098

Schikora-Tamarit, M. À., & Gabaldón, T. (2022). Using genomics to understand the mechanisms of virulence and drug resistance in fungal pathogens. *Biochemical Society Transactions*, *50*(3), 1259–1268. https://doi.org/10.1042/BST20211123

Scopel, E. F. C., Hose, J., Bensasson, D., & Gasch, A. P. (2021). Genetic variation in aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages. *Genetics*, *217*(4), 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

Vaquero, C., Loira, I., Bañuelos, M. A., Heras, J. M., Cuerda, R., & Morata, A. (2020). Industrial Performance of Several *Lachancea thermotolerans* Strains for pH Control in White Wines from Warm Areas. *Microorganisms*, *8*(6), 830. https://doi.org/10.3390/microorganisms8060830

Vejarano, R., & Gil-Calderón, A. (2021). Commercially Available Non-*Saccharomyces* Yeasts for Winemaking: Current Market, Advantages over *Saccharomyces*, Biocompatibility, and Safety. *Fermentation*, *7*(3), 171. https://doi.org/10.3390/fermentation7030171

Vicente, J., Friedrich, A., Schacherer, J., Freel, K., Marquina, D., & Santos, A. (2025). Whole-Genome Sequencing and Phenotyping Reveal Specific Adaptations of *Lachancea thermotolerans* to the Winemaking Environment. *Molecular Ecology*, e17667. https://doi.org/10.1111/mec.17667

Vicente, J., Kelanne, N., Rodrigo-Burgos, L., Navascués, E., Calderón, F., Santos, A., Marquina, D., Yang, B., & Benito, S. (2023). Influence of different *Lachancea thermotolerans* strains in the wine profile in the era of climate challenge. *FEMS Yeast Research*, *23*, foac062. https://doi.org/10.1093/femsyr/foac062

Vilela, A. (2018). *Lachancea thermotolerans*, the Non-*Saccharomyces* Yeast that Reduces the Volatile Acidity of Wines. *Fermentation*, *4*(3), 56. https://doi.org/10.3390/fermentation4030056

Villarreal, P., Villarroel, C. A., O'Donnell, S., Agier, N., Quintero-Galvis, J. F., Peña, T. A., Nespolo, R. F., Fischer, G., Varela, C., & Cubillos, F. A. (2022). Late Pleistocene-dated divergence between South Hemisphere populations of the non-conventional yeast *L.*

cidri. *Environmental Microbiology*, *24*(12), 5615–5629. https://doi.org/10.1111/1462-2920.16103

Vitanović, E., Aldrich, J. R., Winterton, S. L., Boundy-Mills, K., Lopez, J. M., & Zalom, F. G. (2019). Attraction of the Green Lacewing *Chrysoperla comanche* (*Neuroptera: Chrysopidae*) to Yeast. *Journal of Chemical Ecology*, *45*(4), 388–391. https://doi.org/10.1007/s10886-019-01060-w

Warwick, C., Steedman, C., & Grant, R. (2024). Is the domestication phenomenon more perceived than real? *Discover Animals*, *1*(1), 36. https://doi.org/10.1007/s44338-024-00036-9

Xia, W., Nielly-Thibault, L., Charron, G., Landry, C. R., Kasimer, D., Anderson, J. B., & Kohn, L. M. (2017). Population genomics reveals structure at the individual, host-tree scale and persistence of genotypic variants of the undomesticated yeast *Saccharomyces paradoxus* in a natural woodland. *Molecular Ecology*, *26*(4), 995–1007. https://doi.org/10.1111/mec.13954

Yilmazer Aktar, B., Georgakis, N., Labrou, N., Turunen, O., & Binay, B. (2023). Comparative structural and kinetic study for development of a novel candidate L-asparaginase based pharmaceutical. *Biochemical Engineering Journal*, *191*, 108806. https://doi.org/10.1016/j.bej.2023.108806

Zdaniewicz, M., Satora, P., Pater, A., & Bogacz, S. (2020). Low Lactic Acid-Producing Strain of *Lachancea thermotolerans* as a New Starter for Beer Production. *Biomolecules*, *10*(2), 256. https://doi.org/10.3390/biom10020256

Zeidan, R., Ul-Hassan, Z., Al-Thani, R., Balmas, V., & Jaoua, S. (2018). Application of Low-Fermenting Yeast *Lachancea thermotolerans* for the Control of Toxigenic Fungi

Aspergillus parasiticus, *Penicillium verrucosum* and *Fusarium graminearum* and Their Mycotoxins. *Toxins*, *10*(6), 242. https://doi.org/10.3390/toxins10060242

CHAPTER 2

GEOGRAPHY, DOMESTICATION, AND POPULATION GENOMICS OF A WINE YEAST, *LACHANCEA THERMOTOLERANS* [1]

---

[1] Ward, A.K., Peña, J.J., McKibben, M.M., Ambrocio, D.R., Celia-Sanchez, B.N., & Bensasson, D. To be submitted to a peer-reviewed journal.

ABSTRACT

Environmental adaptation has played a major role in the evolution of yeasts due to selective pressures from both wild and domestic environments. The non-model yeast species *Lachancea thermotolerans*, used in wine and other fermentative processes, has been the subject of recent oenological studies focusing on its adaptation to the domestic wine-making environment. Despite its frequent use in industrial fermentation, *L. thermotolerans* is a ubiquitous species with isolates found across six continents and multiple hosts. Here we use over 300 genomes, including 90 new wild strains predominately from trees, to determine whether wild populations are genetically distinct from domestic lineages. We found that wild woodland populations show more population structure than previously recognized and identified 12 distinct clades delineated by geographic and habitat origin. Our analyses suggest occasional recent migration and gene flow between American and European lineages. With the addition of wild strains, we have found evidence that copy number variation in genes previously associated with adaptation to domestic environments may be more prevalent across geographical and ecological origin than previously thought. Additionally, we found that lineage divergences within *L. thermotolerans* are quite old, with nucleotide divergence of approximately 4% between the most diverged lineages. This is approximately twice the distance seen in *Saccharomyces cerevisiae* but resembles distances seen within the wild yeast *Saccharomyces paradoxus*. These results contribute to better understanding of the population structure and evolutionary history of a non-model yeast.

KEYWORDS: population genetics, phylogeography, yeast ecology, genetic admixture

INTRODUCTION

The use of the model yeast, *Saccharomyces cerevisiae*, to ferment food and beverage throughout human history is well recognized (Peña et al., 2025), as are its habitat-associated genetic lineages (Almeida et al., 2015; Liti et al., 2009; Peña et al., 2025; Peter et al., 2018; Tilakaratna & Bensasson, 2017). In recent years, molecular ecology research has expanded to non-*Saccharomyces* yeasts involved in fermentation, industrial processes, or with ecological or clinical relevance (Almeida et al., 2014; Bensasson et al., 2019; Capece et al., 2018; Jolly et al., 2014; Masneuf-Pomarede et al., 2016; Villarreal et al., 2022, 2024). Increased access to whole-genome sequencing data provides further insight into genetic diversity and factors shaping population and evolutionary dynamics within fungal species, including selective pressures that may drive niche-specific adaptation across environments (Dauphin et al., 2023).

Lachancea thermotolerans, a yeast species that co-occurs with *S. cerevisiae* in natural environments (Robinson et al., 2016) and is involved in winemaking (Jolly et al., 2014), is of particular interest in oenology for its ability to produce lactic acid during alcoholic fermentation (Benito 2018; Gobbi et al., 2013). Climate change can lead to earlier fruiting times in wine grapes which can impact wine flavor, and the lactic acid production of *L. thermotolerans* can offset some of these impacts by improving aroma, decreasing aldehyde and fatty acid concentration, and reducing volatile acidity (Vicente et al., 2023; Vilela, 2018). Increased lactic acid production has also contributed to its use in craft brewing, where it is used as an alternative to bacterial souring (Domizio et al., 2016; Osburn et al., 2018) or to enhance ethanol content and aroma (Canonico et al., 2019). In addition to its use in alcoholic beverages, *L. thermotolerans* could be used in agriculture where it has potential as a biocontrol agent against toxigenic fungi (Zeidan et al., 2018).

Previous work using mitochondrial and microsatellite data (Banilas et al., 2016; Freel et al., 2014; Hranilovic et al., 2017) assessed intraspecific diversity and highlighted geography and local adaptation as driving evolution within *L. thermotolerans*. These studies, along with a recent whole-genome study, suggest a role for the oenological environment in lineage differentiation (Freel et al., 2014; Hranilovic et al., 2017, 2018; Vicente et al., 2025) with genetic and phenotypic differences distinguishing winemaking strains from wild strains (Gatto et al., 2020; Hranilovic et al., 2018; Vicente et al., 2025). Despite its global distribution and occurrence in a broad range of habitats (Barnett et al., 2000), most *L. thermotolerans* strains in these studies were from European wine-producing regions. Increasing the number of isolates from natural environments or wine-producing regions in other parts of the world may provide additional information about environmental adaptation and drivers of domestication. For example, analysis of genomes from forest strains would allow a better test of whether domestication explains reports (Vicente et al., 2025) of recent gene gains or losses in *L. thermotolerans*.

Here, we utilize the recent expansion of genomic resources for *L. thermotolerans* to identify population structure and phylogenetic relationships within the species. We identified previously unknown substructure, estimated potential gene flow events between populations, and identified copy number variation of genes conferring increased fitness in domesticated environments. We show (i) increased population structure within Europe and North America, (ii) evidence for historical gene flow between populations, and (iii) preliminary evidence for increased copy number of genes previously associated with adaptation to winemaking environment (Vicente et al., 2025) in wild populations.

MATERIALS AND METHODS

Yeast strains and genomic data

Whole-genome data for *Lachancea thermotolerans* and for sister species *Lachancea quebecensis* were compiled from publicly available data (N =155, Table 2.S1; Freel et al., 2014, 2016; Vicente et al., 2025; Xia et al., 2017). We also generated genome data for 90 more strains (Table 2.S1) from: (i) previously published studies (N = 60, Table 2.S1; Sampaio & Gonçalves, 2008; Robinson et al., 2016; Osburn et al., 2018), (ii) new strains of *L. thermotolerans* from oak and pine bark from Alabama, Florida, Georgia, North Carolina, and Pennsylvania (N = 29, Table 2.S1; Bensasson lab), (iii) and a *L. thermotolerans* strain from a wild azalea flower (Table 2.S1; kindly provided by Jeff Rapp, Athens Technical College). The methods for yeast isolation from bark and species identification were as described in Robinson *et al.* (2016).

DNA was extracted from single yeast colonies using the Promega Wizard Genomic DNA purification kit and the manufacturer's protocol for yeast with the exception that only 75 units of lyticase (Sigma) were used in an overnight incubation at 37 °C. Illumina libraries were generated by the Georgia Genomic and Bioinformatics Core and Admera Health using the purePlex DNA Library Preparation Kit (GGBC Project #5256) or the Nextera DNA-Seq Library Protocol (also known as Illumina DNA Prep; GGBC Project #5881; Admera Health Project 24118-01; Table S2). Paired-end sequencing was performed on the Illumina NextSeq2000 platform (2 x 150 bp). Library preparation for one strain was completed by Azenta (formerly GeneWiz) via the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina and paired-end sequencing was performed on the Illumina MiSeq platform (2 x 250 bp; Table 2.S2). Genomic data will be made available on NCBI-SRA before publication.

Genome assembly and base calling

Paired-end genomic Illumina reads were downloaded from (i) the European Bioinformatics Institute (https://www.ebi.ac.uk/) or (ii) the NCBI using the SRA Toolkit (v3.0.3; https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software), or generated by this study. Reads were mapped to the *L. thermotolerans* reference genome, CBS6340 (assembly ASM14280v1; The Genolevures Consortium et al., 2009), using Burrows-Wheeler Aligner bwa-mem (v0.7.17; Li & Durbin, 2009). We used SAMtools (v1.16.1; Li et al., 2009) to sort, index, and compress bam files and generated a consensus sequence excluding indels. We then used the BCFtools (v1.15.1; Li et al., 2009) call function to generate a consensus sequence and converted VCF files to the FASTQ format in SAMtools using the vcfutils.pl vcf2fq command. Base calls with a phred-scaled quality score less than 40 were treated as missing data (seqtk v1.3; https://github.com/lh3/seqtk) when converting each consensus sequence to FASTA format. Whole genome coverage was calculated using SAMtools depth and visualized in R (v4.2.1; https://www.R-project.org/) (Table 2.S1). Strains were inspected for evidence of heterozygosity and intraspecies contamination using B-allele frequency visualized using the script vcf2alleleplot.pl (Bensasson et al., 2019; Scopel et al., 2021).

Population genomics: species tree and genetic admixture

Consensus genome sequences (based on mapped reads) from all strains were concatenated to create a single multiple alignment file for each chromosome in FASTA format. Ambiguity codes or lowercase base calls were converted to Ns, and ends were filled to make alignments the same length. Chromosomal multiple alignment files were concatenated to produce one whole-genome multiple alignment file.

Neighbor-joining (NJ) trees genetic distance trees were generated for whole-genome and individual gene alignments using MEGA-CC (v10.0.5; Kumar et al., 2012, 2018) and the Tamura-Nei substitution model (Tamura & Nei, 1993) with a gamma distribution and 100 bootstrap replicates. We discarded gaps or missing data from pairwise strain sequence comparisons. The *L. quebecensis* strain LL2012_118 (Freel et al., 2016) served as an outgroup. Resulting NJ trees were rotated using 'ape' (v5.7.1; Paradis & Schliep, 2019) and visualized using the 'ggtree' package (v3.6.2; Yu, 2020; Yu et al., 2017, 2018) in R (v4.2.1). Information about substrate and country of origin was plotted beside the tree (Figure 2.1). Maps were drawn using the 'maps' (v3.4.2.1; https://CRAN.R-project.org/package=maps) and 'ggplot2' (v3.5.1; Wickham, 2016) packages in R.

Population structure and strain ancestry of *L. thermotolerans* were determined using SNP allele frequencies via ADMIXTURE (v1.3.0; Alexander et al., 2009). Genome data from each strain were merged into a single alignment (in a VCF file) using BCFtools. Variant sites were thinned in BCFtools to obtain the minimum allele count of sites to specify the first alternative allele. Low-quality reads (Phred score < 40) were removed using the minQ option in VCFtools (v0.1.16; Danecek et al., 2011). Single alignment VCF files were converted to text-formatted and binary files using PLINK (v1.9b_6.21; Purcell et al., 2007) for ADMIXTURE analyses. Multiple runs of ADMIXTURE, with different numbers of populations or genetic clusters (K) from 2 to 20 with five replicates per K, assigned each strain to one or more clusters. We selected the run with the highest loglikelihood value for each K to visualize population structure (Figure 2.S1). Ancestry proportions were aligned across K values using the CLUMPAK 'Distruct for many K's' pipeline (v1.1; Kopelman et al., 2015) and visualized using the R package 'pophelper'

(v2.3.1; Francis, 2017) (Figure 2.S2). Genetic clusters were verified using monophyletic clades with 100% bootstrap support in a NJ tree (Figure 2.1a).

Phylogenomic relationships were further examined using a maximum likelihood (ML) tree after excluding strains showing recent genetic admixture when K = 15 (Figure 2.S3 and Table 2.S3). Admixed strains were defined as individuals with percent ancestry less than 90% from a single population in ADMIXTURE results (Ward et al., 2025). Strains passing this quality threshold were incorporated into a whole-genome multiple alignment file. The tree was generated using a general time-reversible model with a gamma distribution and ultrafast bootstrapping in IQ-TREE (v1.6.12; Minh et al., 2013; Nguyen et al., 2015) and visualized using the 'phytools' (v2.3.0; Revell, 2024), 'ape,' and 'ggtree' packages in R (v4.2.1).

Population substructure within Europe and North America

To identify fine-scale population substructure within European and North American populations, we used a neighbor-joining tree and ADMIXTURE using strains from the Asian population as an outgroup for both analyses (see 'Population genomics' above; Table 2.S3). ADMIXTURE was run by varying K from 3 to 12 (Europe) or from 3 to 6 (North America) with five replicates per value of K and visualized results from the run with the highest loglikelihood for each K value (Figure 2.S4; Europe: Figures 2.S5, 2.S6; North America: Figures 2.S7, 2.S8).

Estimating historical gene flow events

Historical relationships between populations were examined using TreeMix (v1.13; Pickrell & Pritchard, 2012) after excluding (i) population groups with less than 3 strains and (ii) strains that were admixed when K = 15. Populations were assigned to the remaining strains as in

the most likely ADMIXTURE run (also K = 15; Table 2.S3). Genome data from strains (VCF files) were merged into a single alignment, thinned, and filtered using the same methods for preparation of ADMIXTURE files (see 'Population genomics' above). Allele frequency per population was calculated with PLINK from the merged VCF file and converted into a TreeMix-compatible file using the TreeMix-provided 'plink2treemix.py' script. SNPs were grouped in blocks of 500 SNPs for jackknife standard errors and trees were rooted with the *L. quebecensis* outgroup. Runs assumed different numbers of migration events (M) from 0 to 10 with five replicates per migration event. We selected the run with the highest log-likelihood value for each value of M (Figure 2.S10). Resulting graphs and residual covariance plots were visualized in R (v4.2.1) using the 'plotting_funcs.R' (Pickrell & Pritchard, 2012) and 'popcorn' (v0.02; https://github.com/andrewparkermorgan/popcorn) packages (Figure 2.3, Figure 2.S10). The model including the most significant migration events ($P < 0.01$) (i) supported by lower values of M that persisted at higher levels of M and (ii) validated by more than 50% of significant four-population (f4) tests (Reich et al., 2009; Z-scores ≤ -3.00 or ≥ 3.00) invoked 9 migration events; 5 of these were validated using f4 tests (Figure 2.3, Table 2.S4).

Copy number variation among lineages and ecological origins

De novo assemblies were generated from short read Illumina data for all strains using paired-end libraries and SPAdes (v3.15.5; Prjibelski et al., 2020). Additional quality checks using QUAST (Quality Assessment Tool for Genome Assemblies; Gurevich et al., 2013) and BUSCO scores (Benchmarking Universal Single-Copy Orthologs; Simão et al., 2015) will allow us to determine the completeness of the assembly and are necessary for a quantitative comparison of studies. This is especially important because we use short-read genome data

which can produce incomplete assemblies. The genome assembly for each *L. thermotolerans* strain was compared to 16 *S. cerevisiae* gene sequences with known homologs and copy number variation in *L. thermotolerans* (Vicente et al., 2025). More specifically, each *L. thermotolerans* was treated as a query and compared to a database of *S. cerevisiae* genes using using 'blastn' (BLAST+ v2.14.1; Camacho et al., 2009). We defined hits as homologous genes if they showed significant similarity (e-value < 1e-4; Table 2.S5). This 'blastn' approach only matched the most conserved regions between *S. cerevisiae* and *L. thermotolerans*. Vicente *et al.* (2025) also used a BLAST-based approach to identify copy number variation; however, they also implemented the Control-FREEC tool (Boeva et al., 2012) to assess copy number using read depth.

For visualization, we generated a plot showing the location of the homologous genes in *L. thermotolerans* (Figure 2.S11) using the R (v4.2.1) package 'chromoMap' (v4.1.1; Anand & Rodriguez Lopez, 2022). The number of hits in the *L. thermotolerans* assemblies per *S. cerevisiae* gene of interest was tabulated and used as a measure of copy number variation (CNV; Table 2.S6). Strains were categorized into subgroups based on their ecological origin by substrate and further classified as 'domestic' or 'wild' (Table 2.S6) according to whether they were isolated from an environment with exposure to human activity (e.g., vine or crop tree bark, grape must, wine strains) or from the wild (e.g., from trees or soil in forests, insects).

Copy number for each gene was mapped onto the species neighbor-joining tree with *L. quebecensis* as an outgroup, along with information about substrate, domestic status, and clade (Figure 2.S12). To further investigate the relationship between copy number variation and genetic relationship within *L. thermotolerans*, we performed ancestral character estimation using a probabilistic graphical model (Höhna et al., 2014) in RevBayes (v1.2.5; Höhna et al., 2016) using the 'Revticulate' package (v1.0.0; Charpentier & Wright, 2022) in R (v4.4.2). The

neighbor joining tree of all strains (see above) was transformed from NEWICK to NEXUS format for use with RevBayes using the 'ape' package in R (v4.4.2). We selected the free rates model by comparing model fit using Bayes factors computed from stepping-stone and path sampling simulations of marginal likelihood for the equal and free rates models and comparing rates of evolutionary trait evolution (Table 2.S7). All models were checked for convergence (Table 2.S7) using the R package 'convenience' (v1.0.0; https://github.com/lfabreti/convenience). Posterior distribution of the rates of morphological trait evolution (Figure 2.S13) and phylogenies (Figures 2.4 and 2.S14) for the selected model were visualized using the 'ggplot2' (v3.5.1), 'ggtree' (v3.14.0), 'RevGadgets' (v1.2.1; Tribble et al., 2022), and 'treeio' (v1.30.0; Wang et al., 2020) packages in R (v4.4.2).

Comparing L. thermotolerans nucleotide diversity with Saccharomyces species

Estimates of lineage divergence or nucleotide diversity were calculated using the desktop MEGA 11 application for macOS (Stecher et al., 2020; Tamura et al., 2021) and the Tamura-Nei substitution model (Tamura & Nei, 1993) with 100 bootstrap replicates. All ambiguous positions were removed for each sequence pair using the pairwise deletion option in MEGA. For *L. thermotolerans*, one strain from each of the 12 genetic lineages confirmed in ML analyses was selected at random to represent the clade (Table 2.S8). *Saccharomyces cerevisiae* and *S. paradoxus* sequences were taken from 25 (Almeida et al., 2015; Duan et al., 2018; Peña et al., 2025; Peter et al., 2018) and 10 genetically distinct lineages (Eberlein et al., 2019; He et al., 2022; Hénault et al., 2017; Koufopanou et al., 2020; Liti et al., 2009; Xia et al., 2017; Yue et al., 2017), respectively (Table 2.S4). Selected strain sequence information was downloaded and mapped as seen in 'Genome assembly and base calling,' but mapped to the *S. cerevisiae* (S288c;

SacCer_Apr2011/sacCer3 from UCSC) or *S. paradoxus* (CBS432; assembly ASM207905v1; Yue et al., 2017) reference genomes depending on strain species. Results were visualized using the 'ggplot2' package (v3.5.1) in R (v4.2.1) (Figure 2.S15).

RESULTS

We examined population structure in *L. thermotolerans* using genome data for 239 strains from wild and domestic environments. 'Wild' strains were those from animals, flowers, fruit from uncultivated trees, soil and tree bark or exudate from forests, and 'domestic' isolates were from agricultural environments, bark and fruit from common cultivars (grapes and grapevines, olives and olive trees), foodstuffs, and industrial fermentation strains. Phylogenetic analyses of strains revealed several genetically distinct lineages that occur within Asia, Europe, and North and South America (Figures 2.1, 2.2, 2.S2, and Table 2.S3). These include previously studied wild ('Asia,' 'Americas') and human-associated lineages ('Europe Domestic 1', 'Europe Domestic 2'; Freel et al., 2014; Hranilovic et al., 2017; Vicente et al., 2025). The addition of 83 new genomes from forest habitats in this analysis revealed 5 new subpopulations by phylogeographic (lineages labelled with black stars in Figures 2.1b and 2.2) and allele frequency (ADMIXTURE) analyses (Figures 2.1b, 2.2, 2.S2). The yeast in these subpopulations were primarily from trees in Europe and North America.

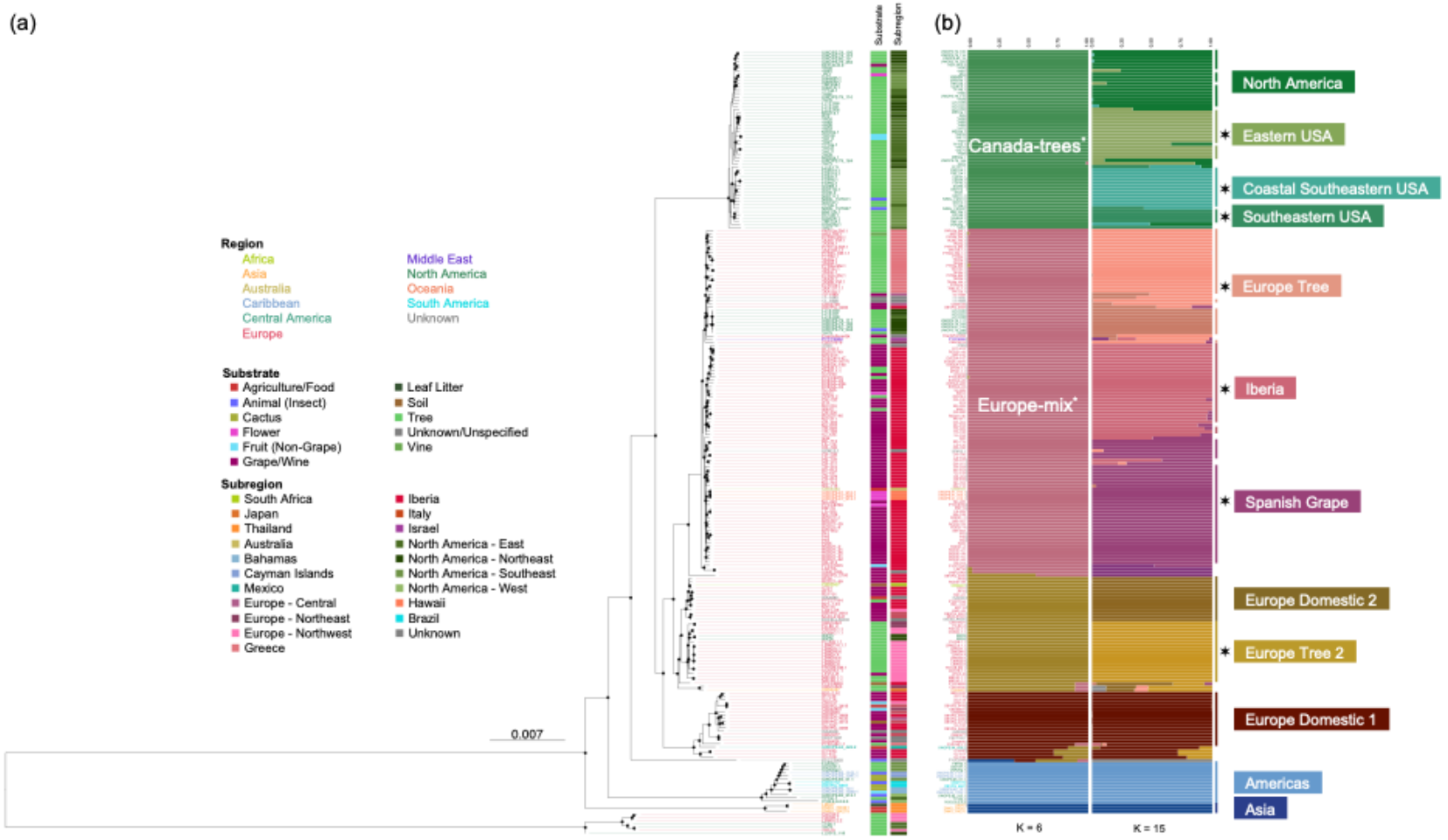


Figure 2.1: Population structure of *Lachancea thermotolerans*. (a) Whole-genome neighbor-joining (NJ) tree of 239 *L. thermotolerans* strains rooted with *Lachancea quebecensis* outgroup (Table S1; dashed line) using the Tamura-Nei model with a gamma distribution (Tamura & Nei, 1993) and 100 bootstraps. Black circles at nodes indicate 100% bootstrap support. Strain names are colored by region of origin; colored bars next to strain names indicate (i) substrate and (ii) country of origin. (b) ADMIXTURE plot with population cluster values K = 6 and K = 15 showing percent ancestry per strain. Strain order corresponds with NJ tree tip order. New clade names are highlighted with black stars.

Population substructure of domestic and wild L. thermotolerans from Europe

Within Europe, there are at least seven distinct populations separated by ecology and geography (Figures 2.1, 2.2d and 2.S4). These include the previously established domestic lineages that are broadly distributed ('Europe Domestic 1') or are mostly in western Europe ('Europe Domestic 2'; Freel et al., 2014; Hranilovic et al., 2017; Vicente et al., 2025), which only rarely occur in wild environments - only one tree isolate from each clade (Figures 2.1 and 2.S4, Table 2.S1). Oak tree strains from northern and western Europe (Finland, France, United Kingdom) form a lineage closely related to the Europe Domestic 2 lineage; we have termed this lineage 'Europe Tree 2' (Figure 2.S4, Tables 2.S1 and 2.S3). This Europe Tree 2 lineage includes two non-European strains; these were from oaks in Ontario, Canada (Tables 2.S1 and 2.S3).

Previous work identified a second lineage primarily from Europe and wine-producing environments, which also occurs outside of Europe; this lineage of heterogeneous strains was named 'Europe-mix' (Hranilovic et al., 2017; Vicente et al., 2025). Our analysis of strains showed evidence for at least three subpopulations within the 'Europe-mix' lineage (Figures 2.2d, 2.S4, 2.S5). These include (i) a sublineage primarily consisting of strains from Spanish vineyards we have named 'Spanish Grape,' (ii) wild and domestic strains from Portugal and Spain termed 'Iberia,' and (iii) tree-associated strains primarily from Greece and Ukraine we are referring to as 'Europe Tree.' A subpopulation of wild strains within the Europe Tree lineage exclusively appears in northeastern North America (Figure 2.2, Tables 2.S1 and 2.S3).

Population substructure in North America

There are at least four wild North American *L. thermotolerans* lineages in the eastern United States and Canada (Figures 2.2a, 2.2b, 2.2c) supported across phylogenetic and

ADMIXTURE analyses (Figures 2.1, 2.S3, and 2.S6). Seven strains from Canada and a strain from Missouri were previously described as part of the 'Canada-trees' lineage (Hranilovic et al., 2017; Vicente et al., 2025). After incorporating new genome data, this lineage includes strains from Georgia, Indiana, New Jersey, and Pennsylvania; we refer to this subclade as 'North America' to reflect its presence in the USA (Tables 2.S1 and 2.S3). There are three additional wild lineages in the USA: 'Eastern USA,' 'Coastal Southeastern USA,' and 'Southeastern USA' (Figure 2.2c). Eastern USA strains are from Indiana, Pennsylvania, West Virginia, and a strain is from Ontario, Canada (Tables 2.S1 and 2.S3). The Coastal Southeastern USA lineage is primarily from the southeastern United States (Florida, Georgia, and Louisiana) and includes a strain from Pennsylvania; the Southeastern USA lineage is sister to Coastal Southeastern USA, with strains from inland areas of Alabama, Georgia, and North Carolina.

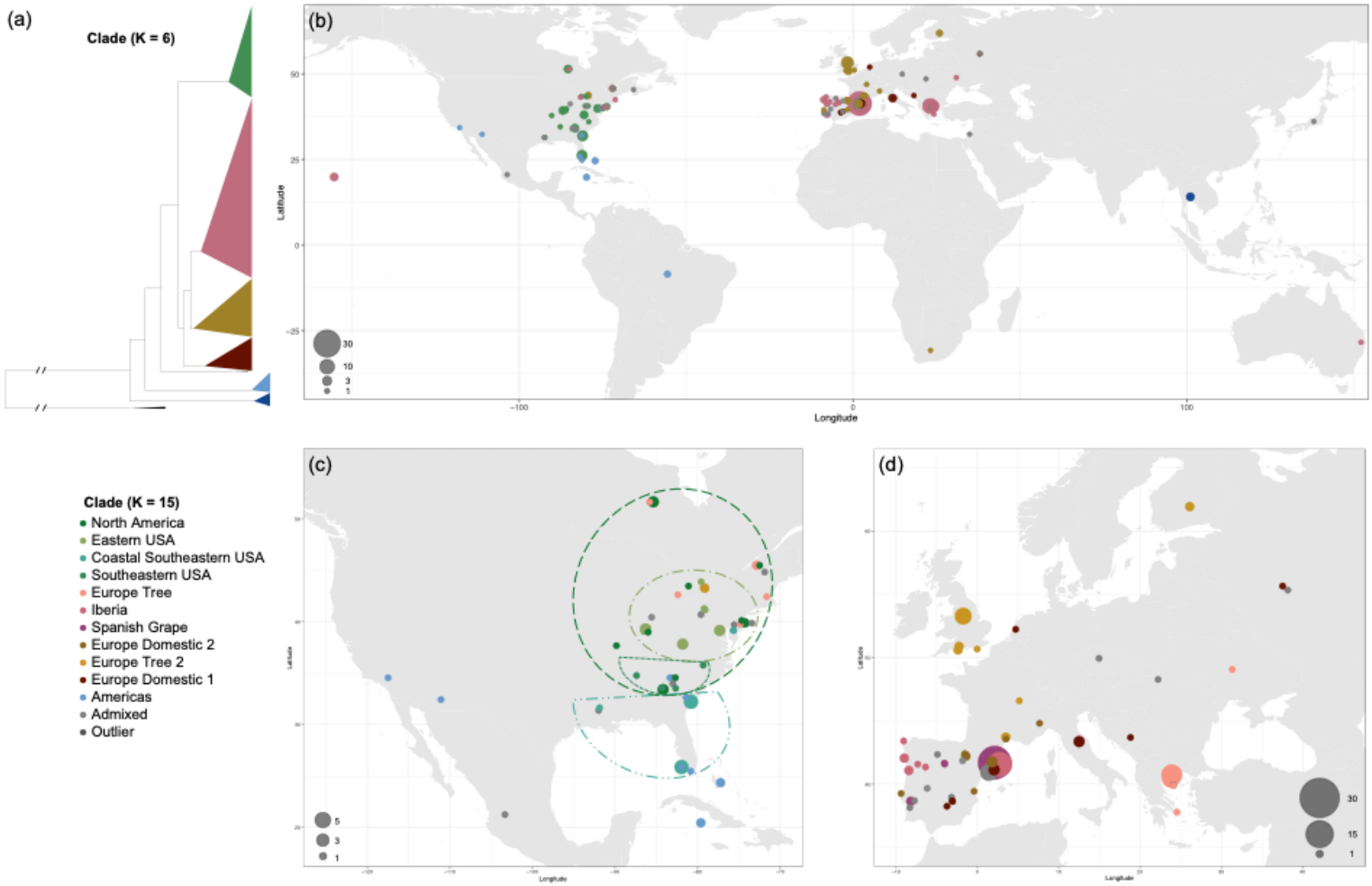
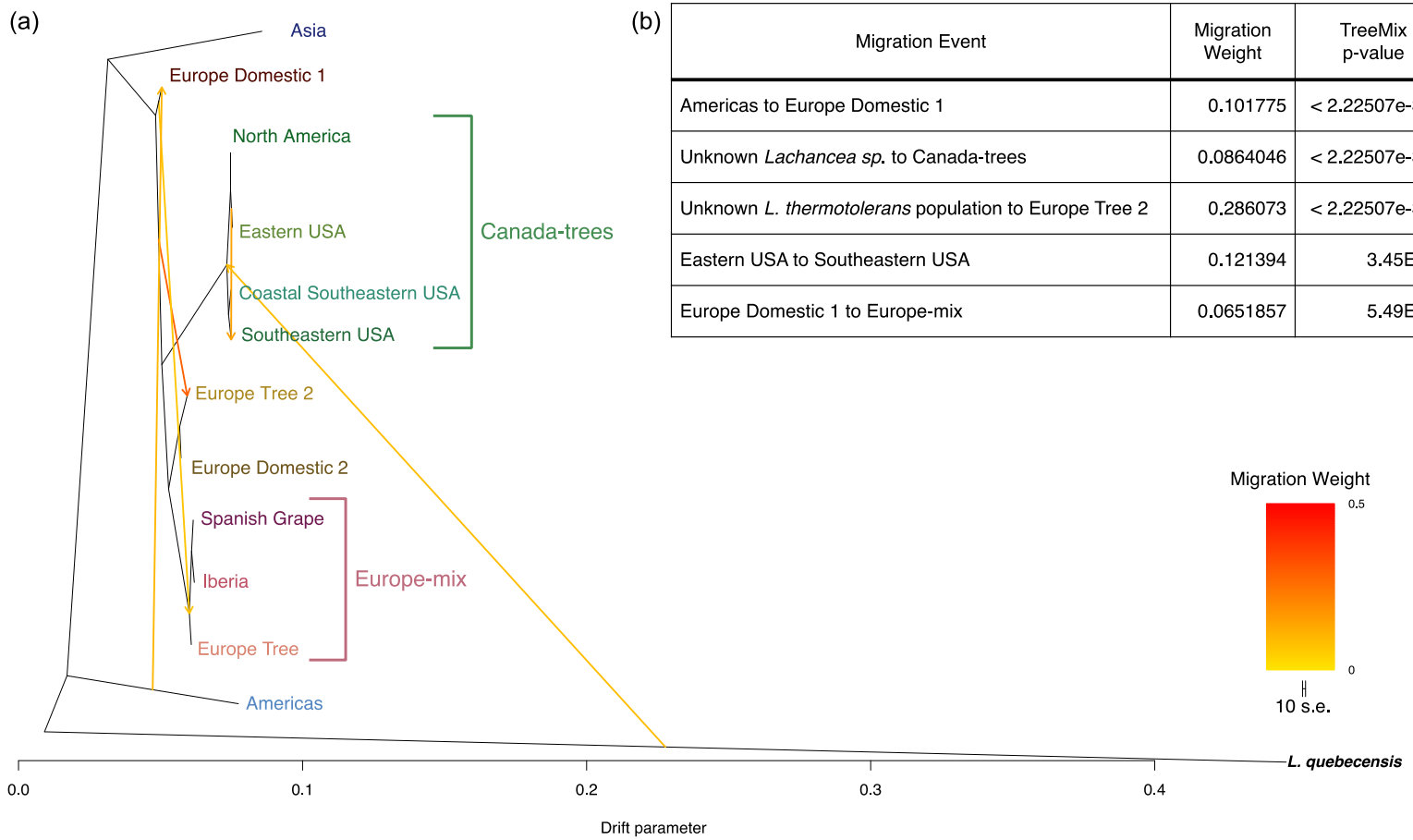


Figure 2.2: Global distribution of *L. thermotolerans*. (a) Collapsed clade tree colored according to ancestry indicated by ADMIXTURE when K = 6 (Figure 2.1b). Map showing the (b) global, (c) North American, and (d) European distribution of *L. thermotolerans* used in this study. Dashed outlines are colored as described in the legend and highlight the distributions of North American, Eastern USA and Southeastern USA. Circle sizes are scaled according to square-root-transformed sample sizes. Admixed strains (< 90% single population ancestry) are colored in light grey; strains without location information are not shown.

Historical gene flow between distinct populations

Europe Tree and Europe Tree 2 lineage strains were found in both Europe and North America (Figure 2.2), raising questions about the potential for ancient gene flow within *L. thermotolerans*. Could the many lineages we observe have arisen from older admixture events? After excluding strains showing recent admixture, we used TreeMix (Pickrell & Pritchard, 2012) to estimate historical migration events impacting genome content of modern *L. thermotolerans* populations (Figure 2.3a). Five events were supported by f4 tests with Z-scores $\leq$ -3.00 or $\geq$ 3.00 (Reich et al., 2009) (Figure 2.3b, Table 2.S5): (i) an unidentified population of *L. thermotolerans* into Europe Tree, (ii) Eastern USA into Southeastern USA, (iii) Americas into Europe Domestic 1, (iv) Europe Domestic 1 into the 'Europe-mix' or southern European lineages (Figure 2.1b, K = 6; Spanish Grape, Iberia, and Europe Tree), and (v) an unknown *Lachancea sp.* into 'Canada-tree' lineages (Figure 2.1b, K = 6; North America, Eastern USA, Southeastern USA, and Coastal Southeastern USA).

(a)



(b)

Migration Event	Migration Weight	TreeMix p-value	Percent Significant f4 Tests
Americas to Europe Domestic 1	0.101775	< 2.22507e-308	60.61
Unknown *Lachancea sp.* to Canada-trees	0.0864046	< 2.22507e-308	75.61
Unknown *L. thermotolerans* population to Europe Tree 2	0.286073	< 2.22507e-308	80.00
Eastern USA to Southeastern USA	0.121394	3.45E-12	100.00
Europe Domestic 1 to Europe-mix	0.0651857	5.49E-03	88.89

Figure 2.3: Historical patterns of migration between *Lachancea thermotolerans* populations. (a) TreeMix admixture graph model rooted with *Lachancea quebecensis* showing relationship between lineages with five migration events confirmed by four-population (f4) tests (Reich et al., 2009) testing for introgression. (b) Table showing migration weights, p-values, and the percent of f4 tests that returned a significant result (Z-scores ≤ -3.00 or ≥ 3.00). Clades are colored as in Figure 2.1b. Branch lengths were estimated by maximum likelihood.

Preliminary results suggest gene copy number changes may not correlate with domestication

Lachancea thermotolerans is widely regarded as an oenologically significant species especially well-suited to wine-making environments (Hranilovic et al., 2017; Vicente et al., 2025). Previous work highlighted at least sixteen genes in *L. thermotolerans* with homologs in *Saccharomyces cerevisiae* that vary in presence and copy number based on wild versus domestic origin (Vicente et al., 2025). These genes, such as *DAL5* and *MAL31* homologs, are largely involved in metabolism of alternative carbon and nitrogen sources (Vicente et al., 2025). It appeared that increases in the copy number of genes important for fermentation occurred as they adapted to vineyards (Vicente et al., 2025). After *de novo* assembly to prevent non-reference sequence from being excluded, we compared whole *L. thermotolerans* genome sequences to *S. cerevisiae* sequences with BLAST and tallied the number of hits for each strain and gene of interest (Figure 2.S14 and Table 2.S7). Our preliminary analysis was less thorough than that of Vicente et al. (2025), which included an analysis of read depth to capture copy number variation that is likely to be missed in an assembly of short read data and is especially good for identifying recent gains and losses for these genes which were all present in the reference. Additionally, our use of the 'blastn' algorithm only matched highly conserved genes, and therefore may exclude copies that are diverged from the *S. cerevisiae* nucleotide sequences. Given the large divergence time between *S. cerevisiae* and *L. thermotolerans*, it is likely that we underestimated copy numbers for these genes.

Using our approach, the European Domestic 1 lineage did not appear to be the only lineage to show an increase in *DAL5* gene copy number; the wild Canada-trees lineage (Figure 2.1b, K = 6) also showed an increase and strains within that clade showed further increases (Figures 2.4a, 2.S11). Similarly, *MAL31* showed high copy number in strains from both domestic

and wild backgrounds (Figures 2.4b, 2.S11). European Domestic 1 lineage unexpectedly showed fewer copies than wild Canada-trees (Figure 2.1b, K = 6) or any other European lineage (Figures 2.4b, 2.S11). Ancestral character estimation models including sister species *L. quebecensis* as an outgroup suggest several genes (7/16) have increased copy number at the shared ancestor of *L. thermotolerans* and *L. quebecensis* (Figures 2.4, 2.S14).

While we generally recapitulate clade-wide patterns seen in Vicente *et al* (2015) for genes with only one locus in the *L. thermotolerans* reference genome, we were not able to recapitulate the gains or losses seen across lineages for multilocus genes (data not shown). In addition to the *de novo* assembly and BLAST method we show here, Vicente *et al* (2015) used the Control-FREEC tool (Boeva et al., 2012) to assess copy number using read depth. Our method returned lower copy numbers across strains for most genes investigated here, as we might have expected because BLAST can only identify gene copies with distinct copies in the assembly (Figure 2.S11). Gene copies that did not assemble correctly, as is more common for tandem repeats, will be missed by BLAST. For 3 out of 16 genes, we found higher copy numbers than Vicente *et al* (2015) in a small number of genomes; homologs of *S. cerevisiae* genes EBP2 (admixed strain CBS6467), GAL1, and GAL3 (Americas clade). Additional analyses, including quality control of our genome assemblies, comparisons of differences between strains tested here and in Vicente *et al* (2015) and Control-FREEC analyses of strains are needed before we can make conclusive statements about the frequency of copy number variation in genes associated with adaptation to the winemaking environment in wild strains of *L. thermotolerans*.

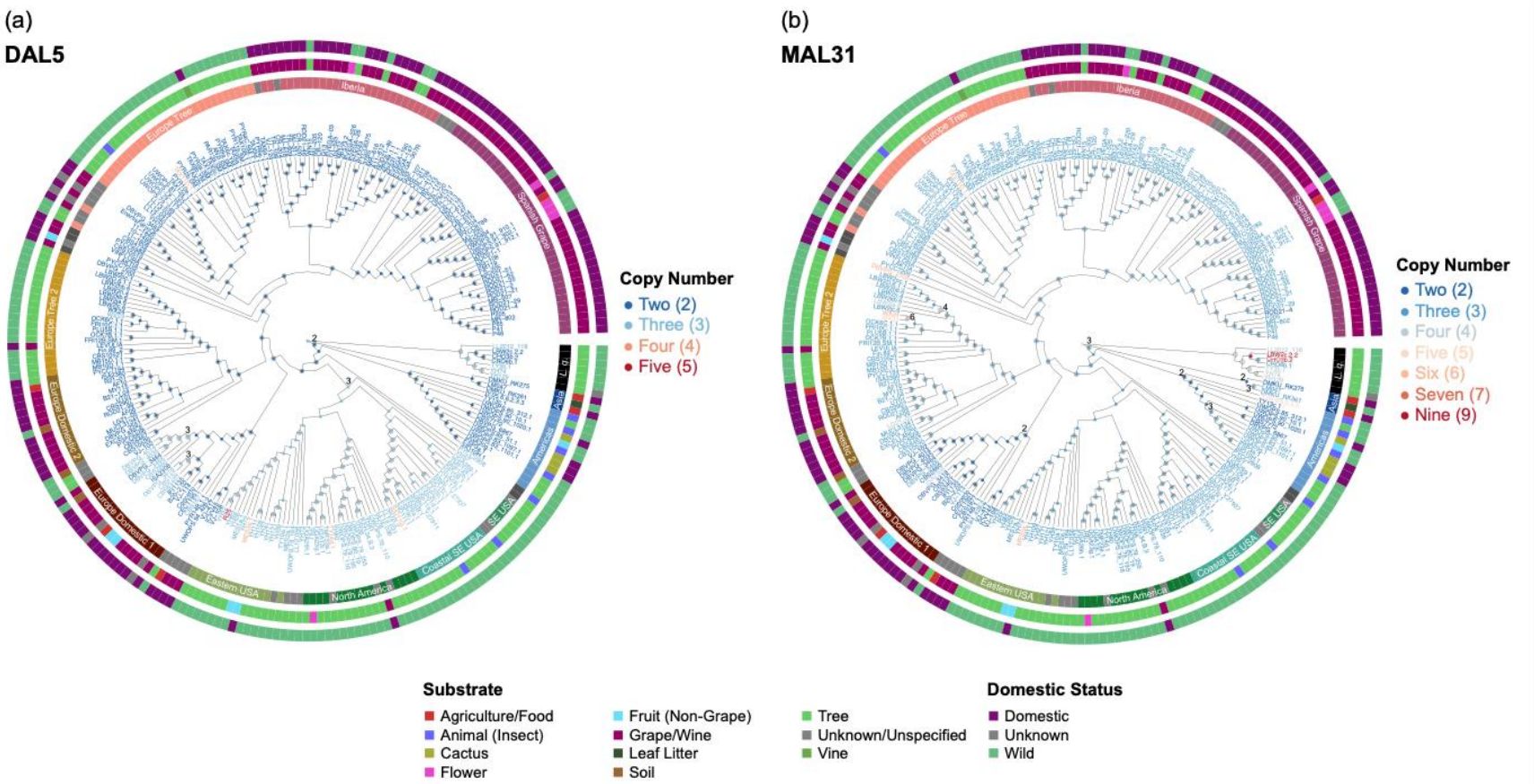


Figure 2.4: Copy number variation of metabolic genes previously associated with the winemaking environment across species tree. Cladograms depicting ancestral copy number estimations for (b) DAL5 and (c) MAL31 have individual color scales; node and tip label are colored according to estimated or observed copy number, respectively. Transitions in copy number at nodes within *L. thermotolerans* are highlighted with the number indicated by node color. Nodes with a bootstrap value of 100 have been circled in black. Colored rings around each cladogram indicate clade, substrate of origin, and domestic status of habitat. Clade color is assigned when K = 15 and includes admixed (light grey) and outlier (dark grey) strains as described in Figure 2.1b.

DISCUSSION

Genetic isolation by environmental niche and geographical origin

Although species in the genus *Lachancea* are some of the most frequently isolated yeasts worldwide, investigation into how and why *Lachancea* species came to inhabit so many ecological niches, including those within woodlands, has been limited (Mozzachiodi et al., 2022; Porter et al., 2019). *Lachancea thermotolerans*, the type species and one of the two most common species within the genus, has been previously associated with foodstuffs; especially the winemaking environment (Benito, 2018; Jolly et al., 2014; Mateus et al., 2020). In this study, we expand the number of available *L. thermotolerans* genome sequences from strains occurring in arboretums and forests and characterize lineages of woodland *L. thermotolerans* in Europe and North America. Phylogeographic analysis of new and previously published *L. thermotolerans* sequences shows that strains are recognizably from northwestern (Europe Tree 2) or southern European forests (Europe Tree), while others are distinguishable as being from the Iberian Peninsula (Figures 2.1, 2.2, and 2.S4). At least four wild lineages separated by geographical region make up the previously described 'Canada-trees' lineage (Hranilovic et al., 2017; Vicente et al., 2025): North America, Eastern USA, Coastal Southeastern USA, and Southeastern USA (Figures 2.1, 2.2, and 2.S6). These results are consistent with previous analyses of *L. thermotolerans* that showed subpopulation differentiation along axes of geographical and ecological origin (Freel et al., 2014; Hranilovic et al., 2017; Masneuf-Pomarede et al., 2016; Vicente et al., 2025). Our discovery of many further subpopulations within the woodland niche suggests that much population diversity and structure remains to be sampled in other continents and habitats.

Although *L. thermotolerans* has been isolated from insects (Babcock et al., 2018; Kogan et al., 2023; Lachance & Kurtzman, 2011) the high level of population structure within these populations is consistent with previous work suggesting that long-distance migration is rare in forest environments (Tilakaratna & Bensasson, 2017). *Lachancea thermotolerans* has not been isolated from atmospheric samples to date, although that does not necessarily preclude air dispersal as a migratory mechanism (Péter et al., 2017). *Saccharomyces cerevisiae* and *S. paradoxus* have also shown high population structure in woodland environments (Leducq et al., 2014; Peña et al., 2025), suggesting that this pattern may be consistent across *Saccharomycotina*.

Intercontinental gene flow of L. thermotolerans

While we have found evidence of European lineages in North America (Figure 2.2), we do not see examples of 'Canada-trees' lineages in Europe. The presence of these European tree lineages in North American forests may be due to human migration. Given *L. thermotolerans* use in wine and food, it is possible that, like *S. cerevisiae*, humans may have unknowingly carried *L. thermotolerans* via victuals or insects traveling with them (Peña et al., 2025). There is also evidence for transoceanic migration in woodland lineages of the wild yeast *S. paradoxus* (Kuehne et al., 2007). Either of these processes could explain recent migrants and gene flow in *L. thermotolerans* (Figures 2.2 and 2.3, Table 2.S5). Current observations of recent migrants are consistent with estimates of gene flow suggesting migration events within and between Europe and North America; our results also suggest that there may be more unsampled lineages of *L. thermotolerans* or another *Lachancea* species closely related to *L. thermotolerans* and *L. quebecensis* (Figure 2.3, Table 2.S5). There appears to be an intercontinental example of

historical gene flow from yeast related to the Americas lineage into the modern Europe-Domestic 1 lineage.

This study shows slightly more lineage diversity within North America than within Europe (Figures 2.1, 2.3, and 2.S8), however there is too little sampling in other continents to determine the geographic origin of the species. Analyses using mutation and recombination rates from *S. cerevisiae* laboratory estimates in the species *Lachancea cidri* dated its lineage divergences to the late Pleistocene period, near the last glacial maximum (Villarreal et al., 2022). Spread and divergence of *S. cerevisiae* has also been dated to the last glacial maximum (Peña et al., 2025). It seems likely that the reduction in ice cover globally also had an impact on the evolution of *L. thermotolerans*, although the much greater genetic divergence of its lineages compared with *S. cerevisiae* (Table 2.S15) suggest either a much older global distribution, or a faster generation time or mutation rate. In both *L. cidri* and *S. cerevisiae*, European populations have shown increased genetic admixture that appears to be related in some capacity to patterns of human migration (Peña et al., 2025; Villarreal et al., 2022).

Possible genomic signatures of domestication in domestic and wild strains of L. thermotolerans

Previous work identified increased copy number of genes associated with the domestication of *S. cerevisiae,* such as *DAL5* and *MAL31,* in *L. thermotolerans* strains from vineyards (Vicente et al., 2025). In contrast, we found that (i) both the Europe Domestic 1 and wild Canada-trees (Figure 2.1b, K = 6) lineages showed increase in *DAL5* gene copy number and (ii) *MAL31* showed high copy number in strains from both domestic and wild origins, but fewer copies in the Europe Domestic 1 lineage than in Canada-trees (Figure 2.1b, K = 6) or other European lineages (Figures 2.4 and 2.S12, Table 2.S6). Tandem repeats in *S. cerevisiae* have

been correlated with the species success in rapid adaptation to industrial environments (Guillamón & Barrio, 2017). If tandem duplications are more likely than other duplications in domestic lineages relative to wild yeast, our methods might explain why we observed high copy numbers of domestication-associated genes in wild lineages.

Lachancea thermotolerans as a model for population genomics

Nucleotide divergence between clades within *L. thermotolerans* is comparable to that seen in *S. paradoxus*, with both species showing nucleotide divergence up to approximately 4% between the most diverged lineages (Figure 2.S15). *Saccharomyces paradoxus* shows some degree of reproductive isolation among its most diverged lineage and has been used as a model for speciation (Delneri et al., 2003); exploring this possibility in *L. thermotolerans* is an interesting direction for future research. The genetic distance seen within *L. thermotolerans* and *S. paradoxus* is double the genetic distance between the most diverged lineages within *S. cerevisiae* (Figure 2.S15). *Lachancea thermotolerans* co-occurs with both species in woodland environments (Robinson et al., 2016; Spurley et al., 2022) and is frequently used alongside *S. cerevisiae* in winemaking (Benito, 2018; Binati et al., 2020; Gobbi et al., 2013) and beer production (Osburn et al., 2018; Postigo et al., 2022). In addition to this overlap within fermentative industry, *L. thermotolerans* and *S. cerevisiae* are both thought to have domesticated lineages associated with expansion of viticulture in the Mediterranean Basin (Almeida et al., 2015; Banilas et al., 2016; Hranilovic et al., 2018); further, genes associated with the suite of domestication traits in both species show significant homology (Liti et al., 2009; Vicente et al., 2025).

The large genetic distance between the *Lachancea* and *Saccharomyces* genera (approximately 125-150 mya, equivalent to that between humans and starfish; Shen et al., 2018), combined with their ecological and genetic similarities, make *L. thermotolerans* a compelling addition to *S. cerevisiae* and *S. paradoxus* as a model system for understanding convergent and parallel evolution dynamics. Continued development of *L. thermotolerans* as a model for population genomics, including additional investigation into the genetic architecture and phenotypic profiles of the species, is needed to develop investigations into processes of repeated evolution between fungal microbes.

SUPPLEMENTARY TABLE INFORMATION.

Due to their size, the following supplemental tables for this chapter are available on GitHub here: https://github.com/bensassonlab/data/tree/master/Ward2025_Dissertation

Table 2.S1. Overview of *Lachancea thermotolerans* and *Lachancea quebecensis* strains used in the study, including compiled metadata from publicly available and new whole-genome data. "Published Clade Assignment" includes the original name of the clade followed by the assignment source; "Study" lists the source of the isolated strain followed by the genome source, if different.

Table 2.S2. Library preparation methods and sequencing platforms used during this study.

Table 2.S3. Strains used in population structure analyses. We defined a strain as admixed if percent ancestry to a single population is < 90% from ADMIXTURE analysis when K = 15. "Published Clade Assignment" includes the original name of the clade followed by the assignment source; "Study" lists the source of the isolated strain followed by the genome source, if different.

Table 2.S4. f4 test results for significant migration events. Table shows f4 test ((A,B);(C,D)) results for significant ($P < 0.01$) edges across TreeMix graphs for all values of M (M0-M10).

Table 2.S6. Copy number of genes of interest per strain. Copy number of 16 *Saccharomyces cerevisiae* homologs in *Lachancea thermotolerans* strains and description of strain clade and ecological origin.

AUTHORS' CONTRIBUTIONS

Research was conceptualized and designed by AKW and DB. AKW, JJP, DRA, and BNC performed DNA extractions and AKW, DRA, and BNC prepared samples for genome sequencing. AKW, JJP, and MMM developed bioinformatic pipelines. AKW obtained and curated public data. AKW performed analyses and data visualization. AKW wrote this chapter with feedback from DB.

REFERENCES

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

Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., Albertin, W., Durrens, P., Sherman, D. J., Marullo, P., Todd Hittinger, C., Gonçalves, P., & Sampaio, J. P. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nature Communications*, *5*(1), 4044. https://doi.org/10.1038/ncomms5044

Anand, L., & Rodriguez Lopez, C. M. (2022). ChromoMap: An R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinformatics*, *23*(1), 33. https://doi.org/10.1186/s12859-021-04556-z

Babcock, T., Borden, J., Gries, R., Carroll, C., Moore, M., & Gries, G. (2018). *Lachancea thermotolerans*, a Yeast Symbiont of Yellowjackets, Enhances Attraction of Three Yellowjacket Species (*Hymenoptera: Vespidae*) to Fruit Powder. *Environmental Entomology*, *47*(6), 1553-1559. https://doi.org/10.1093/ee/nvy139

Banilas, G., Sgouros, G., & Nisiotou, A. (2016). Development of microsatellite markers for *Lachancea thermotolerans* typing and population structure of wine-associated isolates. *Microbiological Research*, *193*, 1–10. https://doi.org/10.1016/j.micres.2016.08.010

Barnett, J. A., Payne, R. W., & Yarrow, D. (2000). *Yeasts: Characteristics and Identification* (3rd ed.). Cambridge University Press.

Benito, S. (2018). The impacts of *Lachancea thermotolerans* yeast strains on winemaking. *Applied Microbiology and Biotechnology*, *102*(16), 6775–6790. https://doi.org/10.1007/s00253-018-9117-z

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

Binati, R. L., Lemos Junior, W. J. F., Luzzini, G., Slaghenaufi, D., Ugliano, M., & Torriani, S. (2020). Contribution of non-Saccharomyces yeasts to wine volatile and sensory diversity: A study on *Lachancea thermotolerans*, *Metschnikowia spp*. and *Starmerella bacillaris* strains isolated in Italy. *International Journal of Food Microbiology*, *318*, 108470. https://doi.org/10.1016/j.ijfoodmicro.2019.108470

Boeva, V., Popova, T., Bleakley, K., Chiche, P., Cappo, J., Schleiermacher, G., Janoueix-Lerosey, I., Delattre, O., & Barillot, E. (2012). Control-FREEC: A tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, *28*(3), 423–425. https://doi.org/10.1093/bioinformatics/btr670

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Canonico, L., Galli, E., Ciani, E., Comitini, F., & Ciani, M. (2019). Exploitation of Three Non-Conventional Yeast Species in the Brewing Process. *Microorganisms*, *7*(1). https://doi.org/10.3390/microorganisms7010011

Capece, A., Romaniello, R., Siesto, G., & Romano, P. (2018). Conventional and Non-Conventional Yeasts in Beer Production. *Fermentation*, *4*(2), 38. https://doi.org/10.3390/fermentation4020038

Charpentier, C. P., & Wright, A. M. (2022). Revticulate: An R framework for interaction with RevBayes. *Methods in Ecology and Evolution*, *13*(6), 1177–1184. https://doi.org/10.1111/2041-210X.13852

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

Dauphin, B., Rellstab, C., Wüest, R. O., Karger, D. N., Holderegger, R., Gugerli, F., & Manel, S. (2023). Re-thinking the environment in landscape genomics. *Trends in Ecology & Evolution*, *38*(3), 261–274. https://doi.org/10.1016/j.tree.2022.10.010

Delneri, D., Colson, I., Grammenoudi, S., Roberts, I. N., Louis, E. J., & Oliver, S. G. (2003). Engineering evolution to study speciation in yeasts. *Nature*, *422*(6927), 68–72. https://doi.org/10.1038/nature01418

Domizio, P., House, J. F., Joseph, C. M. L., Bisson, L. F., & Bamforth, C. W. (2016). *Lachancea thermotolerans* as an alternative yeast for the production of beer. *Journal of the Institute of Brewing*, *122*(4), 599–604. https://doi.org/10.1002/jib.362

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

Eberlein, C., Hénault, M., Fijarczyk, A., Charron, G., Bouvier, M., Kohn, L. M., Anderson, J. B., & Landry, C. R. (2019). Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. *Nature Communications*, *10*(1), 923. https://doi.org/10.1038/s41467-019-08809-7

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

Freel, K. C., Friedrich, A., Hou, J., & Schacherer, J. (2014). Population Genomic Analysis Reveals Highly Conserved Mitochondrial Genomes in the Yeast Species *Lachancea thermotolerans*. *Genome Biology and Evolution*, *6*(10), 2586–2594. https://doi.org/10.1093/gbe/evu203

Freel, K. C., Friedrich, A., Sarilar, V., Devillers, H., Neuvéglise, C., & Schacherer, J. (2016). Whole-Genome Sequencing and Intraspecific Analysis of the Yeast Species *Lachancea quebecensis*. *Genome Biology and Evolution*, *8*(3), 733–741. https://doi.org/10.1093/gbe/evv262

Gatto, V., Binati, R. L., Lemos Junior, W. J. F., Basile, A., Treu, L., de Almeida, O. G. G., Innocente, G., Campanaro, S., & Torriani, S. (2020). New insights into the variability of lactic acid production in *Lachancea thermotolerans* at the phenotypic and genomic level. *Microbiological Research*, *238*, 126525. https://doi.org/10.1016/j.micres.2020.126525

Gobbi, M., Comitini, F., Domizio, P., Romani, C., Lencioni, L., Mannazzu, I., & Ciani, M. (2013). *Lachancea thermotolerans* and *Saccharomyces cerevisiae* in simultaneous and sequential co-fermentation: A strategy to enhance acidity and improve the overall quality of wine. *Food Microbiology*, *33*(2), 271–281. https://doi.org/10.1016/j.fm.2012.10.004

Guillamón, J. M., & Barrio, E. (2017). Genetic Polymorphism in Wine Yeasts: Mechanisms and Methods for Its Detection. *Frontiers in Microbiology*, *8*, 806. https://doi.org/10.3389/fmicb.2017.00806

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

He, P., Shao, X., Duan, S., Han, D., Li, K., Shi, J., Zhang, R., Han, P., Wang, Q., & Bai, F. (2022). Highly diverged lineages of *Saccharomyces paradoxus* in temperate to subtropical climate zones in China. *Yeast*, *39*(1–2), 69–82. https://doi.org/10.1002/yea.3688

Hénault, M., Eberlein, C., Charron, G., Durand, É., Nielly-Thibault, L., Martin, H., & Landry, C. R. (2017). Yeast Population Genomics Goes Wild: The Case of *Saccharomyces paradoxus*. In M. F. Polz & O. P. Rajora (Eds.), *Population Genomics: Microorganisms* (pp. 207–230). Springer International Publishing. https://doi.org/10.1007/978-3-030-04756-6

Höhna, S., Heath, T. A., Boussau, B., Landis, M. J., Ronquist, F., & Huelsenbeck, J. P. (2014). Probabilistic Graphical Model Representation in Phylogenetics. *Systematic Biology*, *63*(5), 753–771. https://doi.org/10.1093/sysbio/syu039

Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., & Ronquist, F. (2016). RevBayes: Bayesian Phylogenetic Inference Using Graphical Models and an Interactive Model-Specification Language. *Systematic Biology*, *65*(4), 726–736. https://doi.org/10.1093/sysbio/syw021

Hranilovic, A., Bely, M., Masneuf-Pomarede, I., Jiranek, V., & Albertin, W. (2017). The evolution of *Lachancea thermotolerans* is driven by geographical determination, anthropisation and flux between different ecosystems. *PLoS ONE*, *12*(9). https://doi.org/10.1371/journal.pone.0184652

Hranilovic, A., Gambetta, J. M., Schmidtke, L., Boss, P. K., Grbin, P. R., Masneuf-Pomarede, I., Bely, M., Albertin, W., & Jiranek, V. (2018). Oenological traits of *Lachancea*

thermotolerans show signs of domestication and allopatric differentiation. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-33105-7

Jolly, N. P., Varela, C., & Pretorius, I. S. (2014). Not your ordinary yeast: Non- *Saccharomyces* yeasts in wine production uncovered. *FEMS Yeast Research*, *14*(2), 215–237. https://doi.org/10.1111/1567-1364.12111

Kogan, H. V., Elikan, A. B., Glaser, K. F., Bergmann, J. M., Raymond, L. M., Prado-Irwin, S. R., & Snow, J. W. (2023). Colonization of Honey Bee Digestive Tracts by Environmental Yeast *Lachancea thermotolerans* Is Naturally Occurring, Temperature Dependent, and Impacts the Microbiome of Newly Emerged Bees. *Microbiology Spectrum*, *11*(2), e05194-22. https://doi.org/10.1128/spectrum.05194-22

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

Koufopanou, V., Lomas, S., Pronina, O., Almeida, P., Sampaio, J. P., Mousseau, T., Liti, G., & Burt, A. (2020). Population Size, Sex and Purifying Selection: Comparative Genomics of Two Sister Taxa of the Wild Yeast *Saccharomyces paradoxus*. *Genome Biology and Evolution*, *12*(9), 1636–1645. https://doi.org/10.1093/gbe/evaa141

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

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

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

Lachance, M.-A., & Kurtzman, C. P. (2011). *Lachancea* Kurtzman (2003). In *Yeasts: A Taxonomic Study* (5th ed., Vols. 1–3, pp. 511–519). Elsevier. doi.org/10.1016/B978-0-444-52149-1.00041-0

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

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., & 1000 Genome Project Data Processing Subgroup. (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, 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

Masneuf-Pomarede, I., Bely, M., Marullo, P., & Albertin, W. (2016). The Genetics of Non-conventional Wine Yeasts: Current Knowledge and Future Challenges. *Frontiers in Microbiology*, *6*. https://doi.org/10.3389/fmicb.2015.01563

Mateus, D., Sousa, S., Coimbra, C., S. Rogerson, F., & Simões, J. (2020). Identification and Characterization of Non-*Saccharomyces* Species Isolated from Port Wine Spontaneous Fermentations. *Foods*, *9*(2), 120. https://doi.org/10.3390/foods9020120

Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195. https://doi.org/10.1093/molbev/mst024

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*, *39*(1–2), 4–24. https://doi.org/10.1002/yea.3699

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & 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

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. (2019). 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

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. https://doi.org/10.1111/mec.17669

Péter, G., Takashima, M., & Čadež, N. (2017). Yeast Habitats: Different but Global. In P. Buzzini, M.-A. Lachance, & A. Yurkov (Eds.), *Yeasts in Natural Ecosystems: Ecology* (pp. 39–71). Springer International Publishing. https://doi.org/10.1007/978-3-319-61575-2_2

Peter, J., De Chiara, M., 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), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genetics*, *8*(11), e1002967. https://doi.org/10.1371/journal.pgen.1002967

Porter, T. J., Divol, B., & Setati, M. E. (2019). *Lachancea* yeast species: Origin, biochemical characteristics and oenological significance. *Food Research International*, *119*, 378–389. https://doi.org/10.1016/j.foodres.2019.02.003

Postigo, V., Sánchez, A., Cabellos, J. M., & Arroyo, T. (2022). New Approaches for the Fermentation of Beer: Non-*Saccharomyces* Yeasts from Wine. *Fermentation*, *8*(6), 280. https://doi.org/10.3390/fermentation8060280

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics*, *70*(1), e102. https://doi.org/10.1002/cpbi.102

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., 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

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, *461*(7263), 489–494. https://doi.org/10.1038/nature08365

Revell, L. J. (2024). phytools 2.0: An updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ*, *12*, e16505. https://doi.org/10.7717/peerj.16505

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

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

Scopel, E. F. C., Hose, J., Bensasson, D., & Gasch, A. P. (2021). Genetic variation in aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages. *Genetics*, *217*(4), 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

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

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

Stecher, G., Tamura, K., & Kumar, S. (2020). Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution*, *37*(4), 1237–1239. https://doi.org/10.1093/molbev/msz312

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*. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

The Genolevures Consortium, Souciet, J.-L., Dujon, B., Gaillardin, C., Johnston, M., Baret, P. V., Cliften, P., Sherman, D. J., Weissenbach, J., Westhof, E., Wincker, P., Jubin, C., Poulain, J., Barbe, V., Segurens, B., Artiguenave, F., Anthouard, V., Vacherie, B., Val, M.-E., … Talla, E. (2009). Comparative genomics of protoploid *Saccharomycetaceae*. *Genome Research*, *19*(10), 1696–1709. https://doi.org/10.1101/gr.091546.109

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

Tribble, C. M., Freyman, W. A., Landis, M. J., Lim, J. Y., Barido-Sottani, J., Kopperud, B. T., Höhna, S., & May, M. R. (2022). RevGadgets: An R package for visualizing Bayesian phylogenetic analyses from RevBayes. *Methods in Ecology and Evolution*, *13*(2), 314–323. https://doi.org/10.1111/2041-210X.13750

Vicente, J., Friedrich, A., Schacherer, J., Freel, K., Marquina, D., & Santos, A. (2025). Whole-Genome Sequencing and Phenotyping Reveal Specific Adaptations of *Lachancea thermotolerans* to the Winemaking Environment. *Molecular Ecology*, e17667. https://doi.org/10.1111/mec.17667

Vicente, J., Kelanne, N., Rodrigo-Burgos, L., Navascués, E., Calderón, F., Santos, A., Marquina, D., Yang, B., & Benito, S. (2023). Influence of different *Lachancea thermotolerans* strains in the wine profile in the era of climate challenge. *FEMS Yeast Research*, *23*, foac062. https://doi.org/10.1093/femsyr/foac062

Vilela, A. (2018). *Lachancea thermotolerans*, the Non-*Saccharomyces* Yeast that Reduces the Volatile Acidity of Wines. *Fermentation*, *4*(3), 56. https://doi.org/10.3390/fermentation4030056

Villarreal, P., O'Donnell, S., Agier, N., Muñoz-Guzman, F., Benavides-Parra, J., Urbina, K., Peña, T. A., Solomon, M., Nespolo, R. F., Fischer, G., Varela, C., & Cubillos, F. A. (2024). Domestication signatures in the non-conventional yeast *Lachancea cidri*. *mSystems*, *9*(1), e0105823. https://doi.org/10.1128/msystems.01058-23

Villarreal, P., Villarroel, C. A., O'Donnell, S., Agier, N., Quintero-Galvis, J. F., Peña, T. A., Nespolo, R. F., Fischer, G., Varela, C., & Cubillos, F. A. (2022). Late Pleistocene-dated divergence between South Hemisphere populations of the non-conventional yeast *L. cidri*. *Environmental Microbiology*, *24*(12), 5615–5629. https://doi.org/10.1111/1462-2920.16103

Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2020). Treeio: An R Package for Phylogenetic Tree Input and Output with Richly Annotated and Associated Data. *Molecular Biology and Evolution*, *37*(2), 599–603. https://doi.org/10.1093/molbev/msz240

Ward, A. K., Scopel, E. F. C., Shuman, B., Momany, M., & Bensasson, D. (2025). Low level contamination confounds population genomic analysis. *bioRxiv*. https://doi.org/10.1101/2025.01.17.633387

Wickham, H. (2016). *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Xia, W., Nielly-Thibault, L., Charron, G., Landry, C. R., Kasimer, D., Anderson, J. B., & Kohn, L. M. (2017). Population genomics reveals structure at the individual, host-tree scale and persistence of genotypic variants of the undomesticated yeast *Saccharomyces paradoxus* in a natural woodland. *Molecular Ecology*, *26*(4), 995–1007. https://doi.org/10.1111/mec.13954

Yu, G. (2020). Using ggtree to Visualize Data on Tree-Like Structures. *Current Protocols in Bioinformatics*, *69*(1), e96. https://doi.org/10.1002/cpbi.96

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

Zeidan, R., Ul-Hassan, Z., Al-Thani, R., Balmas, V., & Jaoua, S. (2018). Application of Low-Fermenting Yeast *Lachancea thermotolerans* for the Control of Toxigenic Fungi *Aspergillus parasiticus*, *Penicillium verrucosum* and *Fusarium graminearum* and Their Mycotoxins. *Toxins*, *10*(6), 242. https://doi.org/10.3390/toxins10060242

Table 2.S5. Genes used in copy-number variation analyses. List of gene names and positions within the *L. thermotolerans* reference genome with names of corresponding *Saccharomyces cerevisiae* homologs.

L. thermotolerans Gene Name	*S. cerevisiae* Standard Gene Name	*S. cerevisiae* Systemic Gene Name	Reference	Chromosome (Letter)	Chromosome (Number)	Gene Start (bp)	Gene End (bp)
KLTH0A00308g	DIP5	YPL265W	Vicente et al., 2025	chrA	chr1	23428	25053
KLTH0A00550g	FLO11	YIR019C	Vicente et al., 2025	chrA	chr1	49572	53438
KLTH0A01166g	MCH4	YOL119C	Vicente et al., 2025	chrA	chr1	108732	110162
KLTH0A01936g	GAL1, GAL3	YBR020W, YDR009W	Vicente et al., 2025	chrA	chr1	171070	172647
KLTH0A01958g	GAL10	YBR019C	Vicente et al., 2025	chrA	chr1	172931	174997
KLTH0A07832g	DAL5	YJR152W	Vicente et al., 2025	chrA	chr1	660003	661517
KLTH0B00176g	YER152C	YER152C	Vicente et al., 2025	chrB	chr2	10791	12137
KLTH0B01166g	DIP5	YPL265W	Vicente et al., 2025	chrB	chr2	102227	103960
KLTH0B09284g	PUG1, RTA1	YER185W, YGR213C	Vicente et al., 2025	chrB	chr2	761694	762773
KLTH0B10384g	DAL5	YJR152W	Vicente et al., 2025	chrB	chr2	877070	878725
KLTH0D00396g	MCH4	YOL119C	Vicente et al., 2025	chrD	chr4	35809	37278
KLTH0D03894g	FLO11	YIR019C	Vicente et al., 2025	chrD	chr4	376254	378374
KLTH0D15378g	DAL5	YJR152W	Vicente et al., 2025	chrD	chr4	1270232	1271893
KLTH0E03234g	DAL5	YJR152W	Vicente et al., 2025	chrE	chr5	293162	294727
KLTH0E05808g	PUG1	YER185W	Vicente et al., 2025	chrE	chr5	520621	521553
KLTH0E13508g	MCH4	YOL119C	Vicente et al., 2025	chrE	chr5	1197940	1199244
KLTH0E16588g	DAL5	YJR152W	Vicente et al., 2025	chrE	chr5	1470414	1472027
KLTH0E16984g	MAL31	YBR298C	Vicente et al., 2025	chrE	chr5	1508741	1510360
KLTH0F00440g	CHA1	YCL064C	Vicente et al., 2025	chrF	chr6	34215	35240
KLTH0F04620g	EBP2	YKL172W	Vicente et al., 2025	chrF	chr6	410119	411339
KLTH0F15268g	ATO3	YDR384C	Vicente et al., 2025	chrF	chr6	1251729	1252580
KLTH0G19360g	YKL107W	YKL107W	Vicente et al., 2025	chrG	chr7	1667351	1668259
KLTH0G19624g	YER152C	YER152C	Vicente et al., 2025	chrG	chr7	1697587	1698993
KLTH0G19756g	YKL107W	YKL107W	Vicente et al., 2025	chrG	chr7	1713024	1713935
KLTH0H05280g	MAL31	YBR298C	Vicente et al., 2025	chrH	chr8	469755	471596
KLTH0H15928g	YIM1	YMR152W	Vicente et al., 2025	chrH	chr8	1367388	1368470
KLTH0H15950g	YIM1	YMR152W	Vicente et al., 2025	chrH	chr8	1369786	1370871

Table 2.S7. Marginal likelihood summary of ancestral character estimation model selection. Summary of marginal likelihood (ml) results from the equal rates (ERM) and free rates (freeK) models of evolution for ancestral character estimation in RevBayes calculated via stepping-stone (ss) and path sampling (ps) methods. Includes calculated Bayes Factors (BF).

Gene	ERM Convergence	freeK Convergence	ERM ml, ss	ERM ml, ps	freeK ml, ss	freeK ml, ps	BF ml, ss	BF ml, ps	BF Model Preference	Rates Graph Model Preference
ATO3	yes	yes	-50.41187	-50.41305	-49.99625	-50.00526	0.00827866	0.00812187	noPref	freeK
CHA1	yes	yes	-44.51652	-44.51188	-43.32954	-43.33495	0.02702574	0.02679665	noPref	freeK
DAL5	yes	yes	-85.80795	-85.80756	-82.03109	-82.02999	0.04501334	0.0450222	noPref	freeK
DIP5	yes	yes	-262.3302	-262.3326	-236.7455	-236.7446	0.10261829	0.10263124	noPref	freeK
EBP2	yes	yes	-68.27846	-68.27872	-65.15131	-65.15618	0.04688193	0.04681099	noPref	freeK
FLO11	yes	yes	-377.2594	-377.2634	-340.3798	-340.4497	0.10287096	0.10267623	noPref	freeK
GAL1	yes	yes	-106.8265	-106.8254	-101.2474	-101.243	0.053639	0.05367216	noPref	freeK
GAL3	yes	yes	-51.13686	-51.13253	-48.47831	-48.48671	0.05338909	0.05313115	noPref	freeK
GAL10	yes	yes	-64.25402	-64.25009	-58.35111	-58.35826	0.09636591	0.09618222	noPref	freeK
MAL31	yes	yes	-126.3251	-126.3215	-121.3723	-121.3737	0.03999606	0.03995603	noPref	freeK
MCH4	yes	yes	-54.26629	-54.26611	-53.47012	-53.4734	0.01478023	0.01471557	noPref	freeK
PUG1	yes	yes	-31.83669	-31.84132	-31.59398	-31.5987	0.0076528	0.00764884	noPref	freeK
RTA1	yes	yes	-14.62019	-14.62078	-13.96624	-13.96457	0.04576046	0.0459204	noPref	freeK
YER152C	yes	yes	-68.12975	-68.12349	-69.75685	-69.74545	-0.0236016	-0.0235301	noPref	freeK
YIM1	yes	yes	-38.15461	-38.15294	-35.27338	-35.2737	0.07851802	0.07846518	noPref	freeK
YKL107W	yes	yes	-9.458048	-9.460955	-8.669572	-8.666515	0.0870466	0.08770658	noPref	freeK

Table 2.S8. Strains used in lineage divergence or nucleotide diversity estimates.

Strain	Species	Clade	Clade Source	Genome Source
UWOPS.85_312.1	*L. thermotolerans*	Americas	Vicente et al., 2025	Vicente et al., 2025
LM047	*L. thermotolerans*	Asia	Vicente et al., 2025	Vicente et al., 2025
TY10b.1	*L. thermotolerans*	Coastal Southeastern USA	This Study	This Study
YH171	*L. thermotolerans*	Eastern USA	This Study	This Study
DBVPG_4035	*L. thermotolerans*	Europe Domestic 1	Vicente et al., 2025	Vicente et al., 2025
UT01	*L. thermotolerans*	Europe Domestic 2	Vicente et al., 2025	Vicente et al., 2025
TAX8b.YM.1	*L. thermotolerans*	Europe Tree	This Study	This Study
Fin.89_2	*L. thermotolerans*	Europe Tree2	This Study	Vicente et al., 2025
TR-108	*L. thermotolerans*	Iberia	This Study	Vicente et al., 2025
LL12_036	*L. thermotolerans*	North America	This Study	Hranilovic et al., 2017
TMF12k.1	*L. thermotolerans*	Southeastern USA	This Study	This Study
ROD21-4	*L. thermotolerans*	Spanish Grape	This Study	Vicente et al., 2025
BJ13	*S. cerevisiae*	CHNVIII	Duan et al., 2018	Duan et al., 2018
BJ22	*S. cerevisiae*	CHNIV/Far_East_Asia	Peña et al., 2025	Duan et al., 2018
CBS2910	*S. cerevisiae*	Alpechin	Peter et al., 2018	Peter et al., 2018
CLIB553	*S. cerevisiae*	French dairy	Peter et al., 2018	Peter et al., 2018
CLQCA_24SC-235	*S. cerevisiae*	Asian fermentation, Sake, Huangjiu, Manou7	Peña et al., 2025	Peter et al., 2018
DBVPG1621-5A	*S. cerevisiae*	Wine/European	Peña et al., 2025	Peter et al., 2018
EN14S01	*S. cerevisiae*	CHNIX/Taiwanese	Peña et al., 2025	Peter et al., 2018
FJ8	*S. cerevisiae*	CHNI	Duan et al., 2018	Duan et al., 2018
FSP15b.1	*S. cerevisiae*	North American C	Peña et al., 2025	Peña et al., 2025
HE009	*S. cerevisiae*	French Guiana human	Peter et al., 2018	Peter et al., 2018
HN15	*S. cerevisiae*	CHNV	Duan et al., 2018	Duan et al., 2018
HN19	*S. cerevisiae*	CHNIII	Duan et al., 2018	Duan et al., 2018
JHBMY20.1	*S. cerevisiae*	CHNIX	Duan et al., 2018	Duan et al., 2018
MAJ_G	*S. cerevisiae*	African beer	Peter et al., 2018	Peter et al., 2018
N39-7A	*S. cerevisiae*	Far East, Russian	Peter et al., 2018	Peter et al., 2018
PYR4b.1.1	*S. cerevisiae*	European oak	Peña et al., 2025	Peña et al., 2025
S8BM-30-2D	*S. cerevisiae*	Asian islands	Peter et al., 2018	Peter et al., 2018
SDO9s1	*S. cerevisiae*	North American B	Peña et al., 2025	Peña et al., 2025
SX5	*S. cerevisiae*	CHNII	Duan et al., 2018	Duan et al., 2018
TY14b.1	*S. cerevisiae*	American wild	Peña et al., 2025	Peña et al., 2025
UWOPS03-459.1	*S. cerevisiae*	Malaysian	Peter et al., 2018	Peter et al., 2018
YPS1009_b	*S. cerevisiae*	African palm wine	Peter et al., 2018	Peter et al., 2018
YPS604	*S. cerevisiae*	North American A	Peña et al., 2025	Peña et al., 2025
ZP674	*S. cerevisiae*	Japan B	Peña et al., 2025	Almeida_2015
ZP785	*S. cerevisiae*	Japan A	Peña et al., 2025	Almeida_2015
SN-TB12-1	*S. paradoxus*	China	He et al., 2022	He et al., 2022
Z3	*S. paradoxus*	Europe	He et al., 2022	Koufopanou et al., 2020
HB-XS1-1	*S. paradoxus*	Far East	He et al., 2022	He et al., 2022
UWOPS91-917.1	*S. paradoxus*	Hawaii	Liti et al., 2009	Yue et al., 2017
B2	*S. paradoxus*	SpA	Eberlein et al., 2019	Xia et al., 2017
16_253B	*S. paradoxus*	SpB	He et al., 2022	Eberlein et al., 2019
14_164C	*S. paradoxus*	SpC	Hénault et al., 2017	Eberlein et al., 2019
16_199C	*S. paradoxus*	SpC*	Hénault et al., 2017	Eberlein et al., 2019
R24	*S. paradoxus*	SpD1	He et al., 2022	Xia et al., 2017
WX20	*S. paradoxus*	SpD2	He et al., 2022	Xia et al., 2017

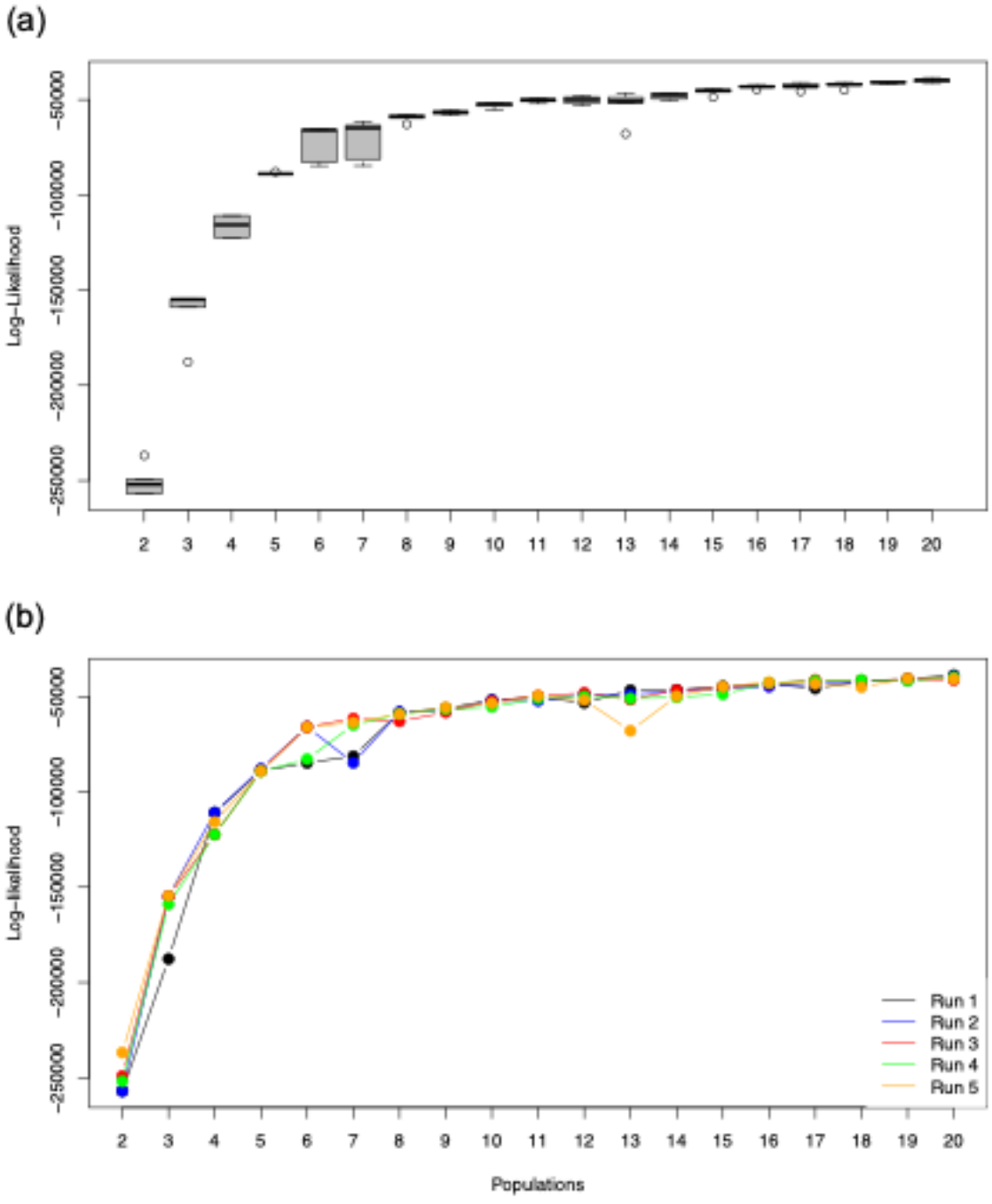


Figure 2.S1. ADMIXTURE selection criteria. Loglikelihood values from ADMIXTURE analyses from five replicates of population cluster values K = 2 to K = 20 for 239 strains of *Lachancea thermotolerans*. (a) Boxplot representing loglikelihood values as a function of five replicate runs. (b) Line graph showing the change in loglikelihood as value of K increases. Lines are colored by replicate.

Figure 2.S2 Population structure of *Lachancea thermotolerans*. Plots showing percent ancestry per strain for population cluster values K = 2 to K = 20 for 239 strains of *L. thermotolerans*. Plotted runs represent the highest loglikelihood value from each value of K; strain order corresponds with neighbor-joining (NJ) tree tip order. The plot highlighted by the red arrow (K = 15) is the model with distinct populations that best correlated with monophyletic clades in the NJ tree. Strain names are colored according to region of origin.

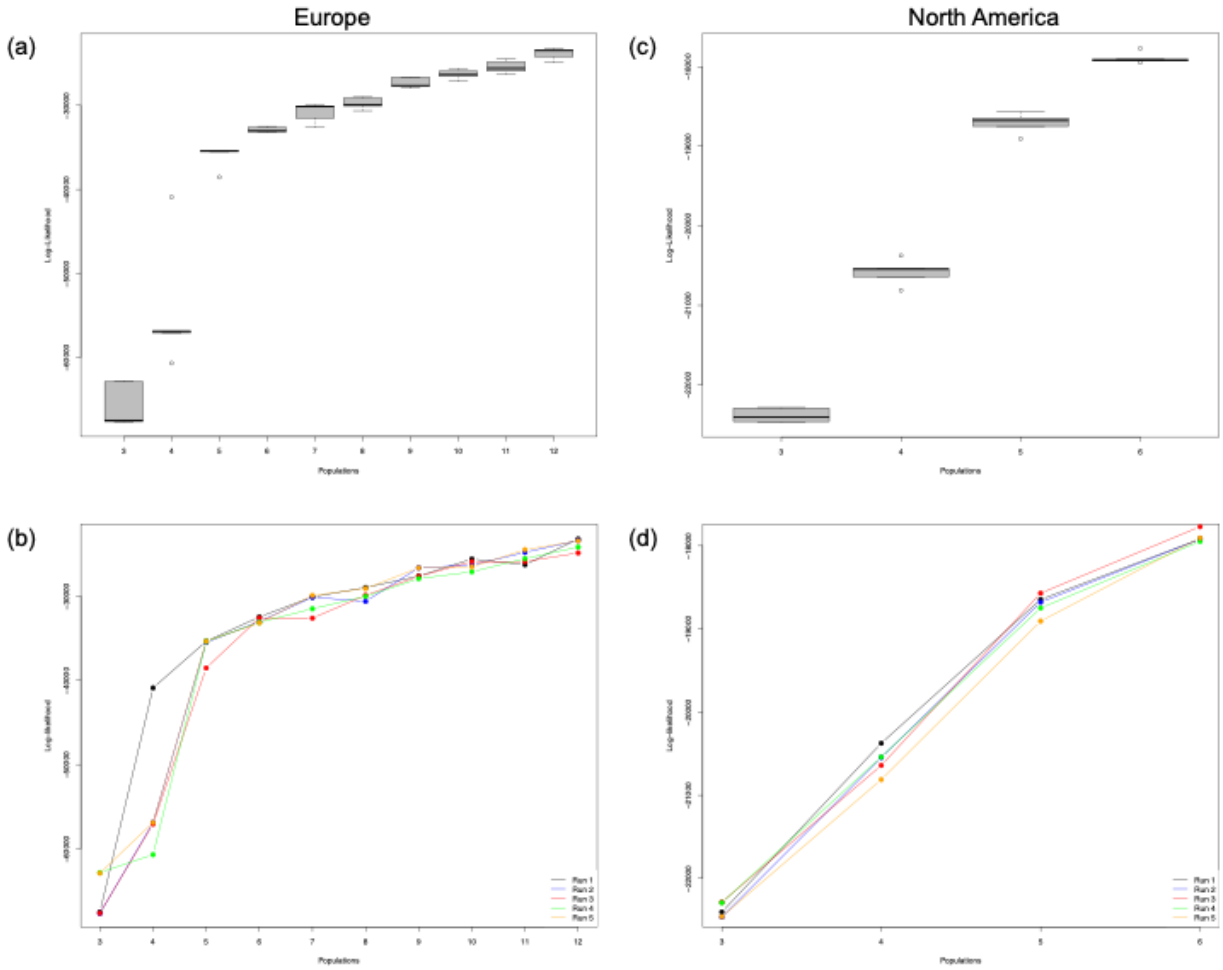


Figure 2.S3: ADMIXTURE selection criteria for European and North American substructure. Loglikelihood values from ADMIXTURE analyses from five replicates of population cluster values (a, b) K = 3 to K = 12 for Europe and (c, d) K = 3 to K = 6 for North America. The Asian clade was included in both analyses as an outgroup. (a, c) Boxplots representing loglikelihood values as a function of five replicate runs. (b, d) Line graphs showing change in loglikelihood as value of K increases. Lines are colored by replicate.

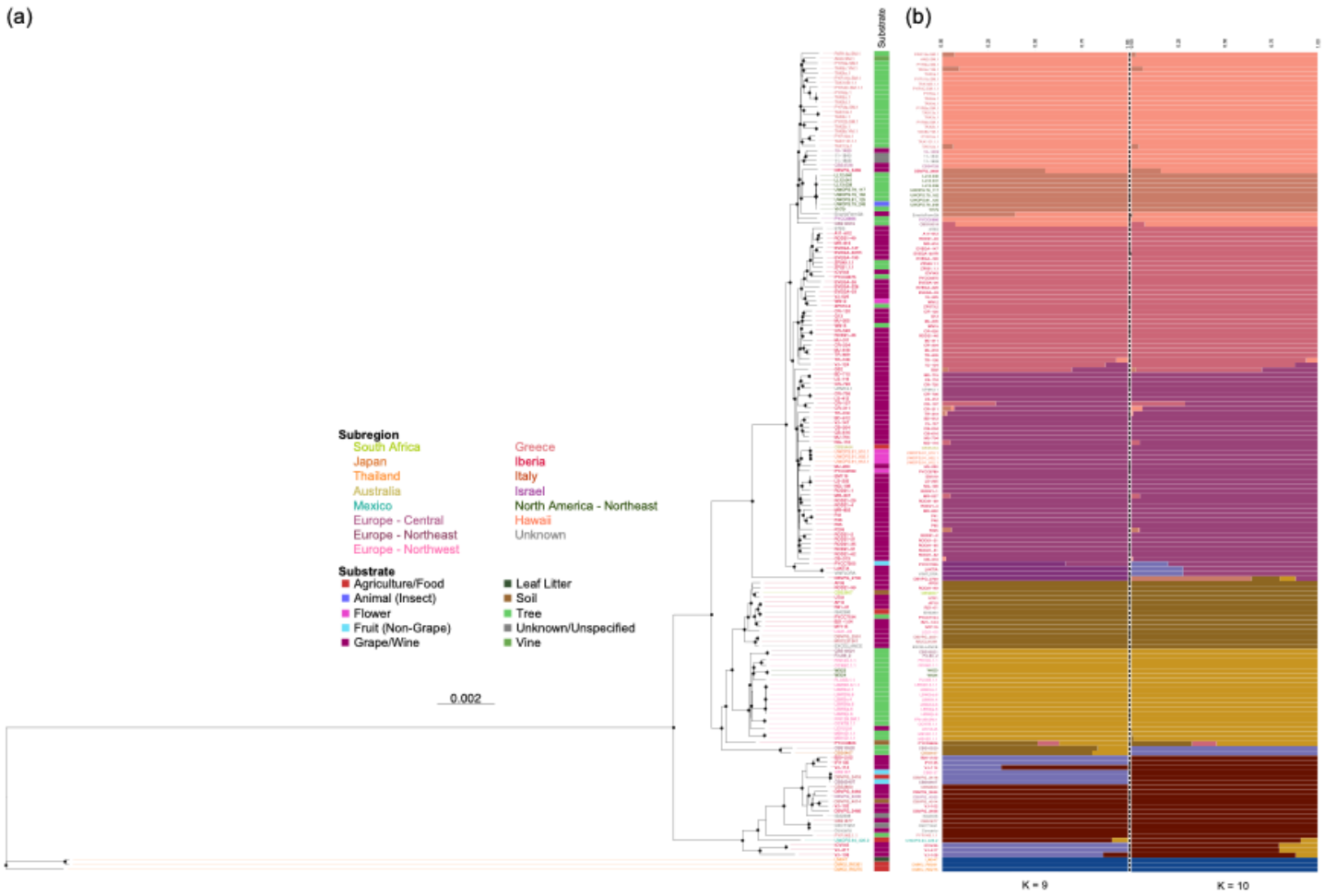


Figure 2.S4: Substructure within European populations of *Lachancea thermotolerans*. (a) Whole-genome neighbor-joining (NJ) tree of 169 *L. thermotolerans* strains rooted at the Asian clade (Table S1) using the Tamura-Nei model with a gamma distribution (Tamura & Nei, 1993) and 100 bootstraps. Black circles at nodes indicate 100% bootstrap support. Strain names are colored by country; colored bars next to strain names indicate substrate of origin. (b) ADMIXTURE plots with population cluster values K = 9 and K = 10 showing percent ancestry per strain. Strain order corresponds with NJ tree tip order.

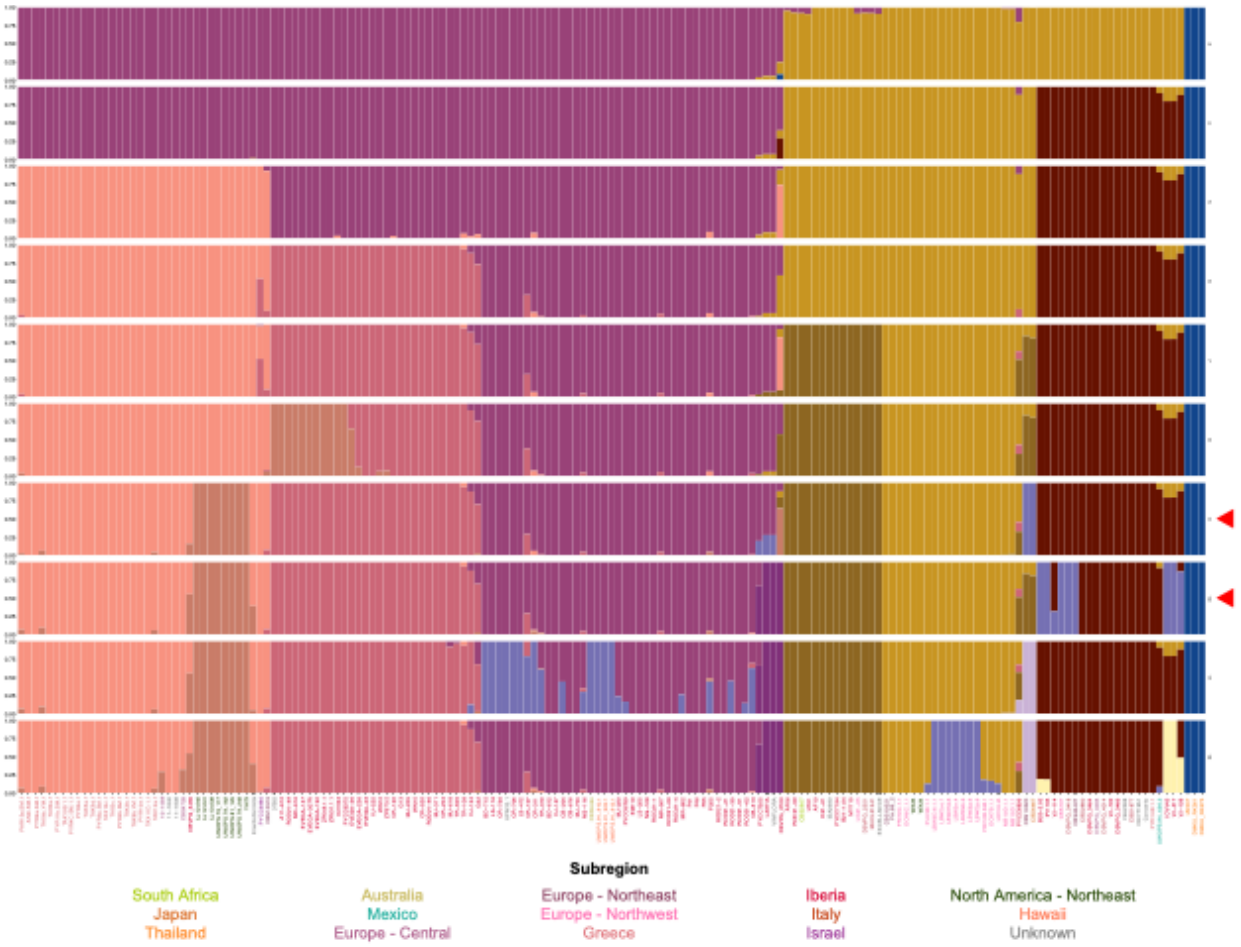


Figure 2.S5: Population substructure of European *Lachancea thermotolerans*. Plots showing percent ancestry per strain for population cluster values K = 3 to K = 12 for strains of *L. thermotolerans* from European clades (N = 166). The Asian clade (N=3) was included as an outgroup. Plotted runs represent the highest loglikelihood value from each value of K; strain order corresponds with neighbor-joining (NJ) tree tip order. The plot highlighted by the red arrow (K = 9) is the model with distinct populations that best correlated with monophyletic clades in the NJ tree. Strain names are colored according to country.

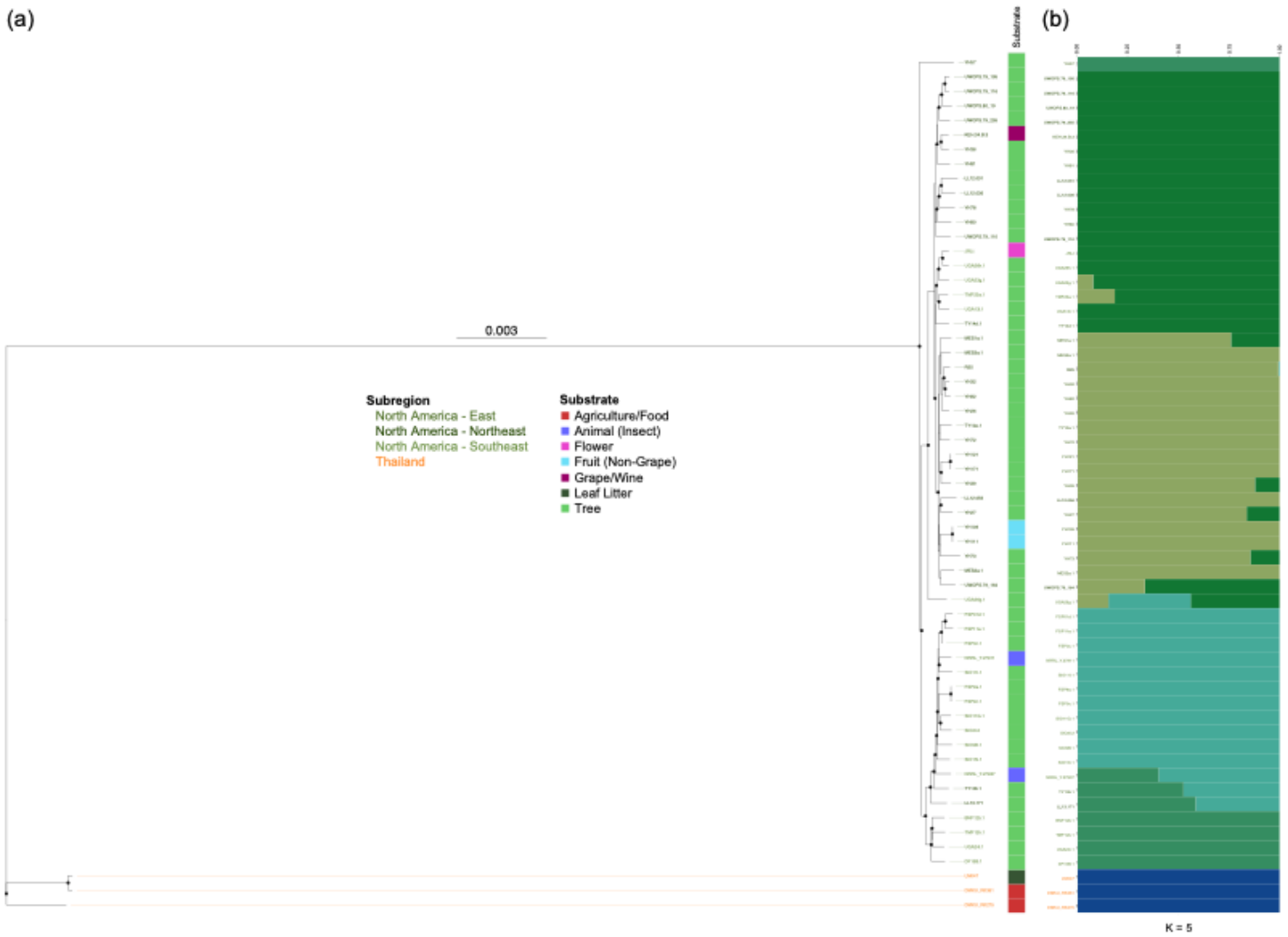


Figure 2.S6: Substructure within North American populations of *Lachancea thermotolerans*. (a) Whole-genome neighbor-joining (NJ) tree of 59 *L. thermotolerans* strains rooted at the Asian clade (Table S1) using the Tamura-Nei model with a gamma distribution (Tamura & Nei, 1993) and 100 bootstraps. Black circles at nodes indicate 100% bootstrap support. Strain names are colored by country; colored bars next to strain names indicate substrate of origin. (b) ADMIXTURE plot with population cluster value K = 5 showing percent ancestry per strain. Strain order corresponds with NJ tree tip order.

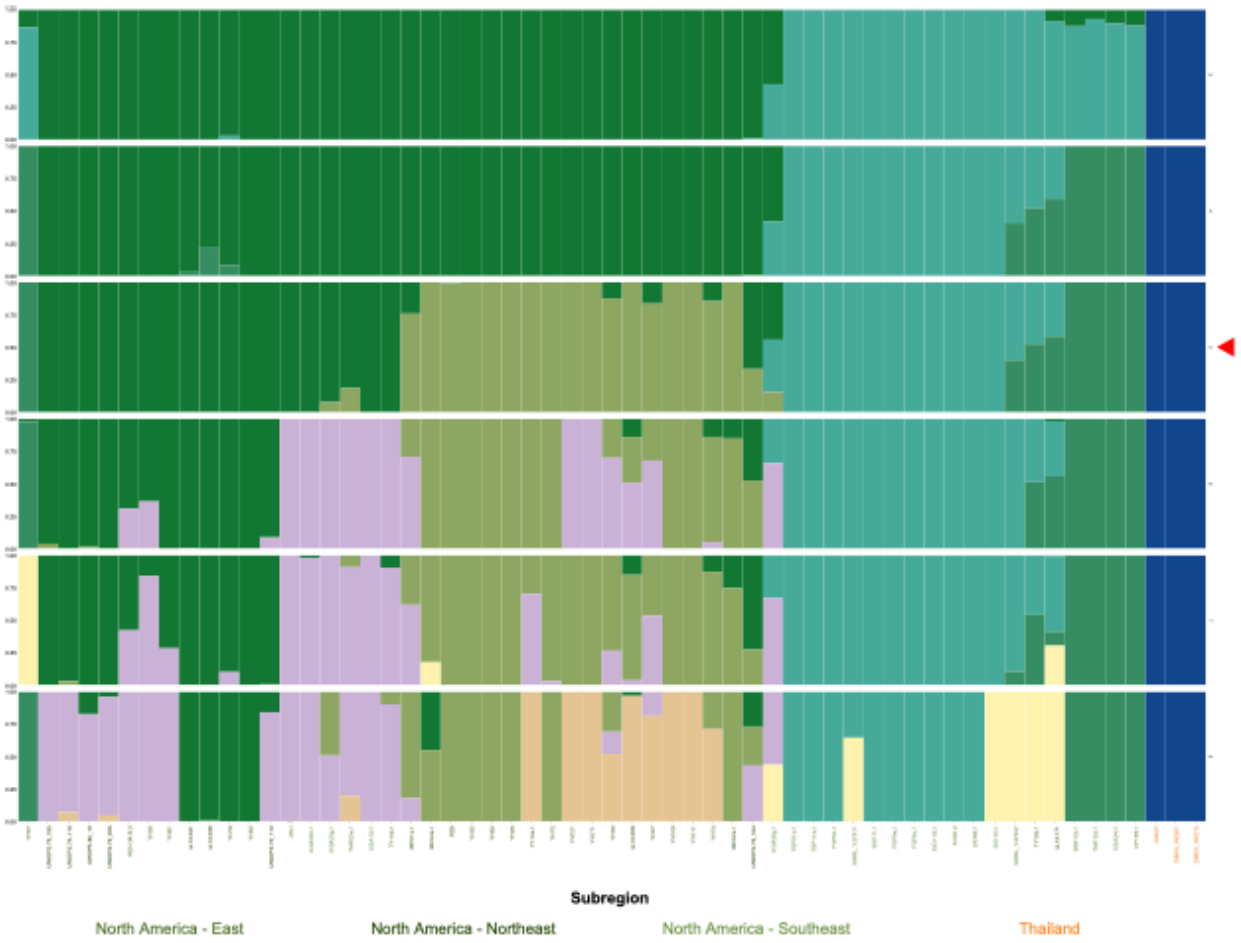


Figure 2.S7: Population substructure of North American *Lachancea thermotolerans*. Plots showing percent ancestry per strain for population cluster values K = 2 to K = 20 for strains of *L. thermotolerans* from North American clades (N = 56). The Asian clade (N=3) was included as an outgroup. Plotted runs represent the highest loglikelihood value from each value of K; strain order corresponds with neighbor-joining (NJ) tree tip order. The plot highlighted by the red arrow (K = 5) is the model with distinct populations that best correlated with monophyletic clades in the NJ tree. Strain names are colored by country.

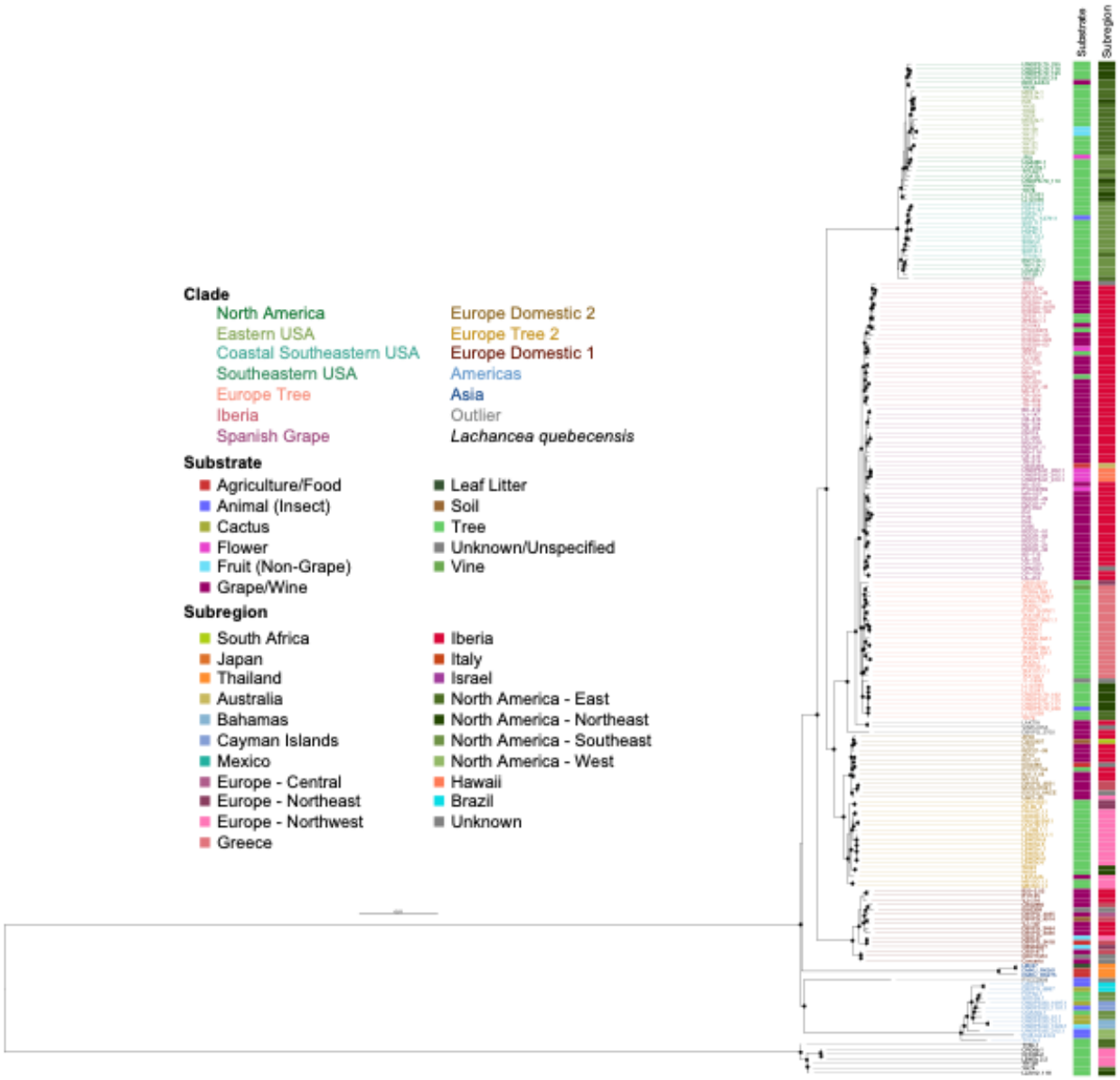


Figure 2.S8: Phylogenetic relationships of non-admixed *Lachancea thermotolerans*. Whole-genome maximum-likelihood (ML) phylogeny of 210 *L. thermotolerans* strains using a general time-reversible model with a gamma distribution after excluding admixed strains (percent ancestry from a single lineage < 90% when K = 15; Figure 1b) rooted with *Lachancea quebecensis* outgroup. Tree was constructed using IQ-TREE ultrafast bootstrapping (1000 bootstraps); black circles at nodes indicate 100% bootstrap support. Strain names are colored by lineage when K = 15 and colored bars next to strain name indicate (i) substrate and (ii) country of origin.

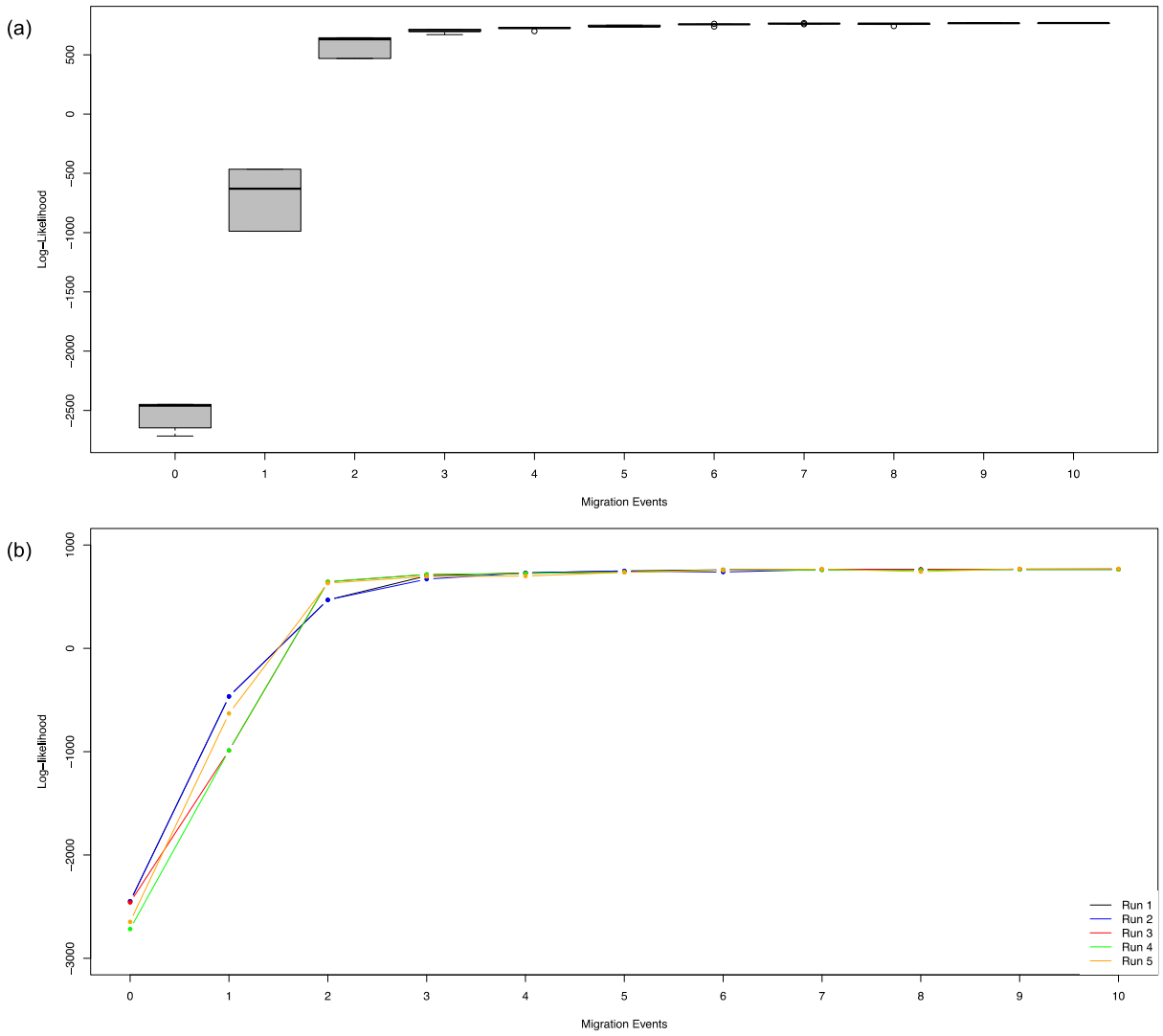


Figure 2.S9: TreeMix run selection criteria. Loglikelihood values from TreeMix analyses from five replicates of migration values M = 0 to M = 10 for 208 strains of *Lachancea thermotolerans* after excluding (i) populations with less than 3 strains, (ii) admixed strains, and (iii) outlier or singleton strains and grouping strains based on ancestry (Fig. 1a, 1b). (a) Boxplot representing loglikelihood values as a function of five replicate runs. (b) Line graph showing the change in loglikelihood as value of M increases. Lines are colored by replicate.

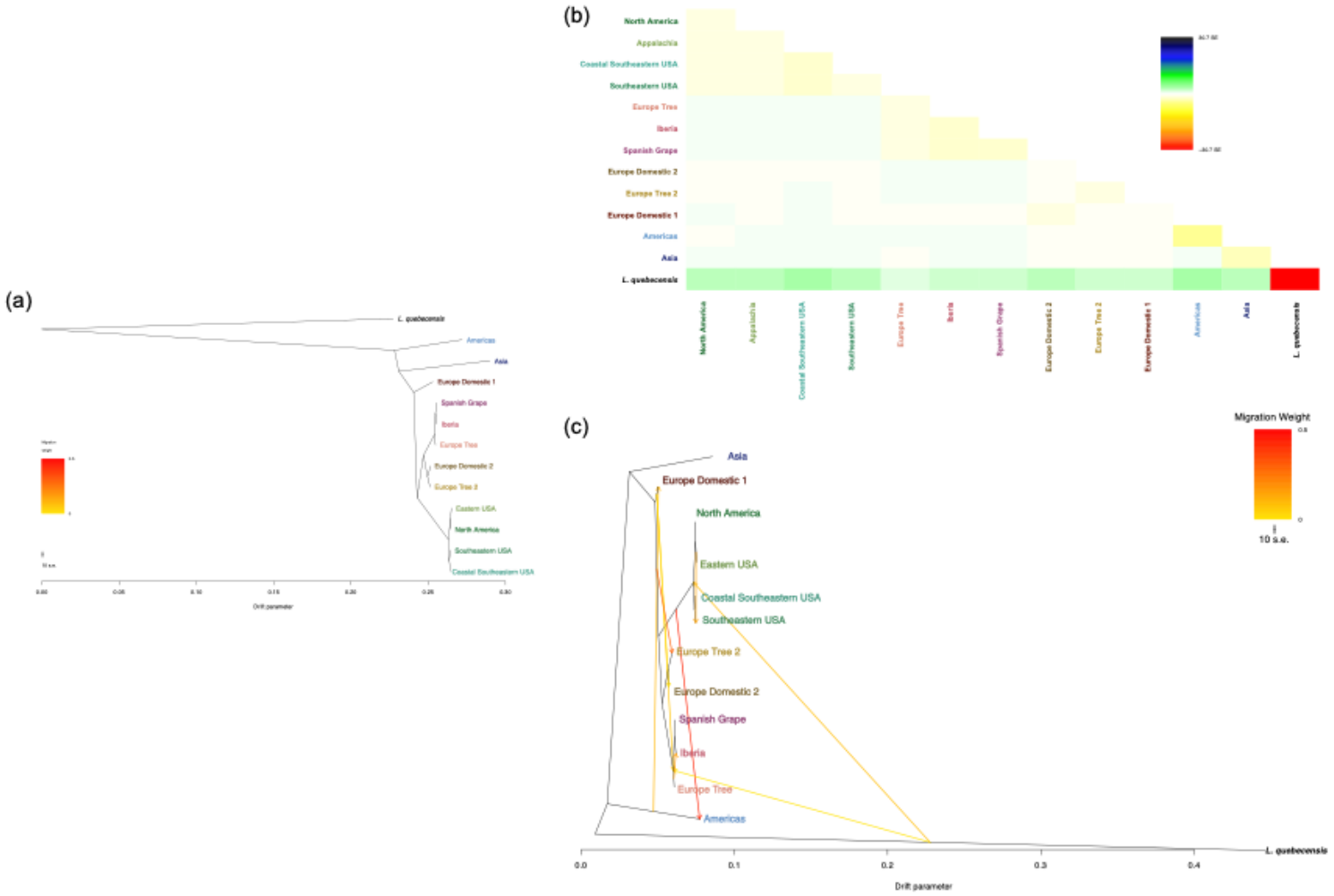


Figure 2.S10: Admixture graphs support multiple migration events across *L. thermotolerans* lineage. (a) Maximum likelihood tree generated by TreeMix with no migration events. (b) Plot showing residual fit for TreeMix graph with eight migration events. Residual covariance was calculated by dividing between each pair of populations by average standard error across all pair combinations (Pickrell & Pritchard, 2012). (c) Tree relationship between lineages and all eight migration events inferred using TreeMix. Clade names and colors correspond with previously identified ancestral lineages.

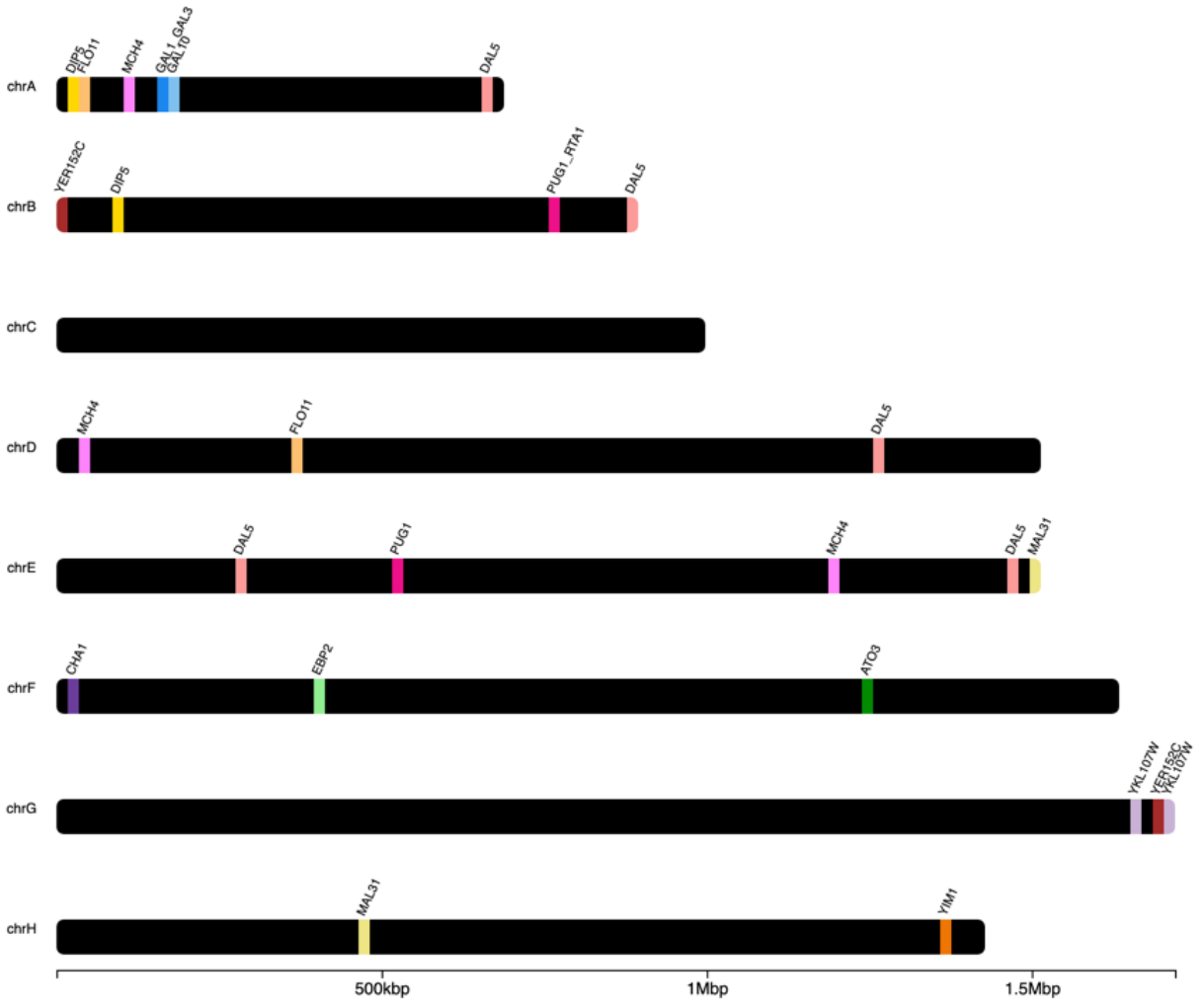


Figure 2.S11: Chromosomal location of genes of interest. Graphical depiction of the location of *Sacchromyces cerevisiae* homologs on *L. thermotolerans* chromosomes according to reference genome annotations. Regions are colored and labeled according to the gene of interest.

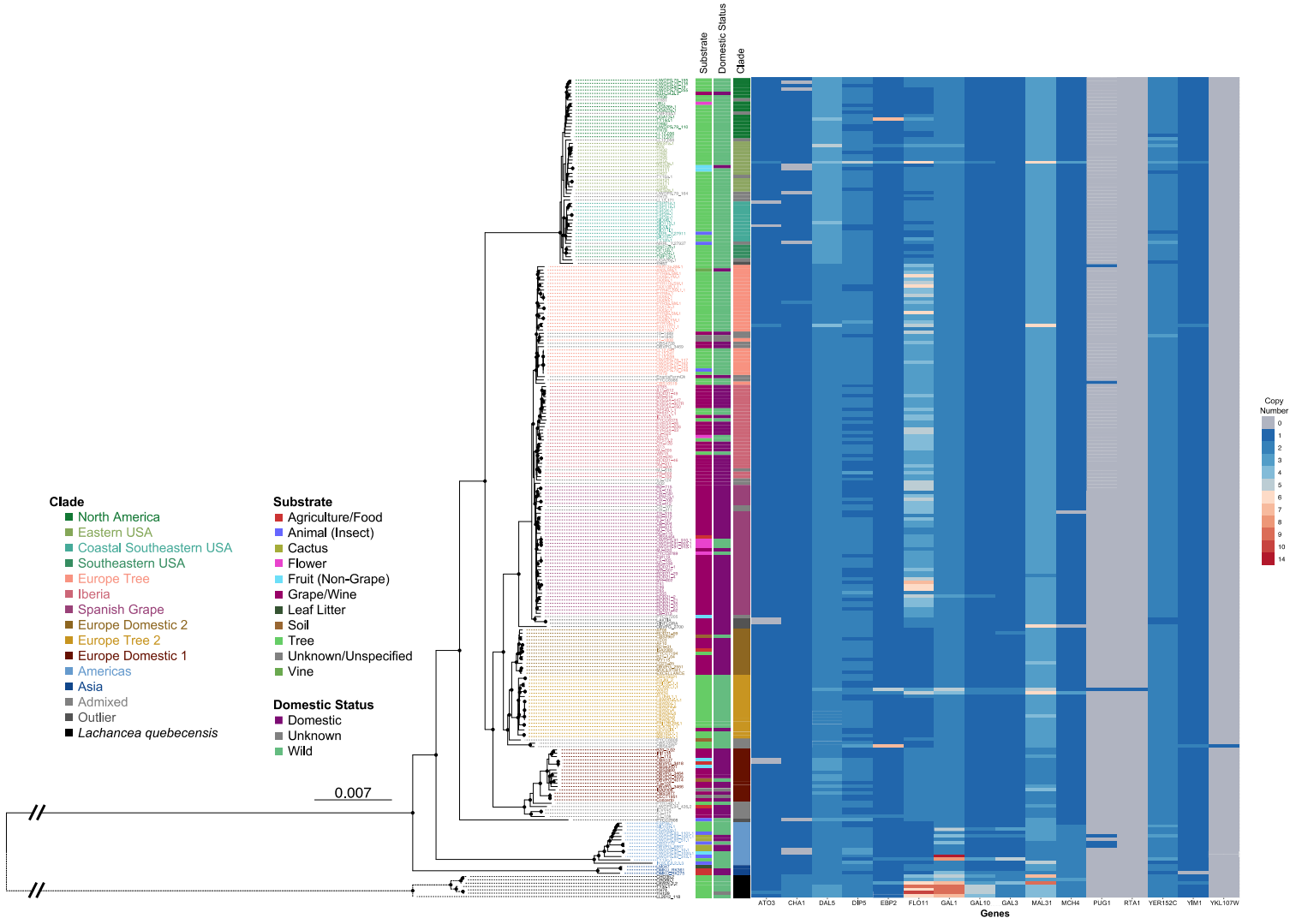


Figure 2.S12: Copy number variation across species tree. Whole-genome NJ tree as in Figure 1a, Strain names are colored by clade assignment when K = 15 (Fig. 1b); colored bars next to strain names indicate (i) substrate, (ii) domestic status of habitat, and (iii) clade. Heatmap is ordered based on NJ tree and shows copy numbers per strain for each of the 16 genes of interest.

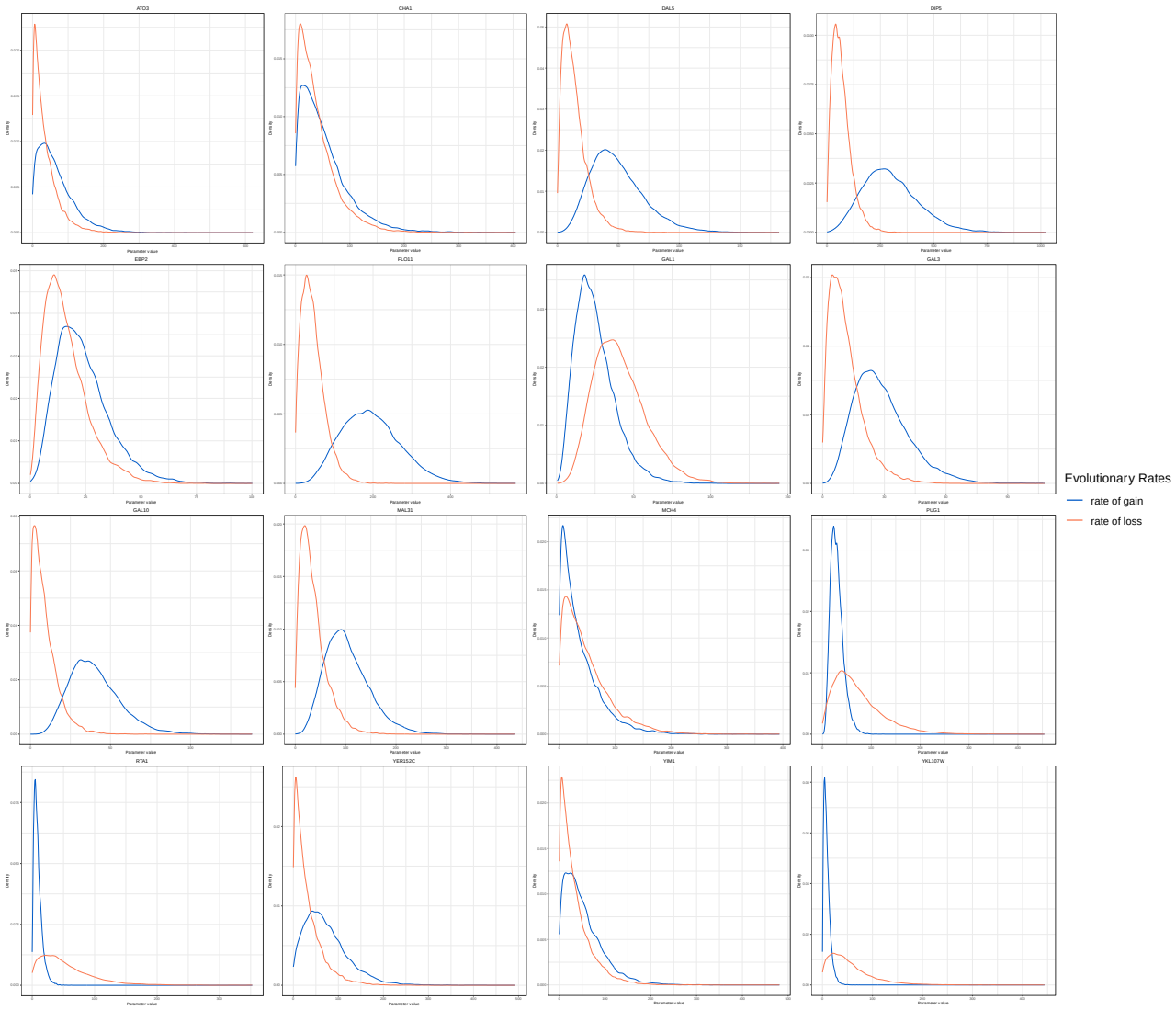


Figure 2.S13: Posterior distributions of evolutionary rate models. Graphs of the posterior distributions of rates of trait (i.e., copy number) gain and loss using the free rates model of discrete character estimation for all genes of interest.

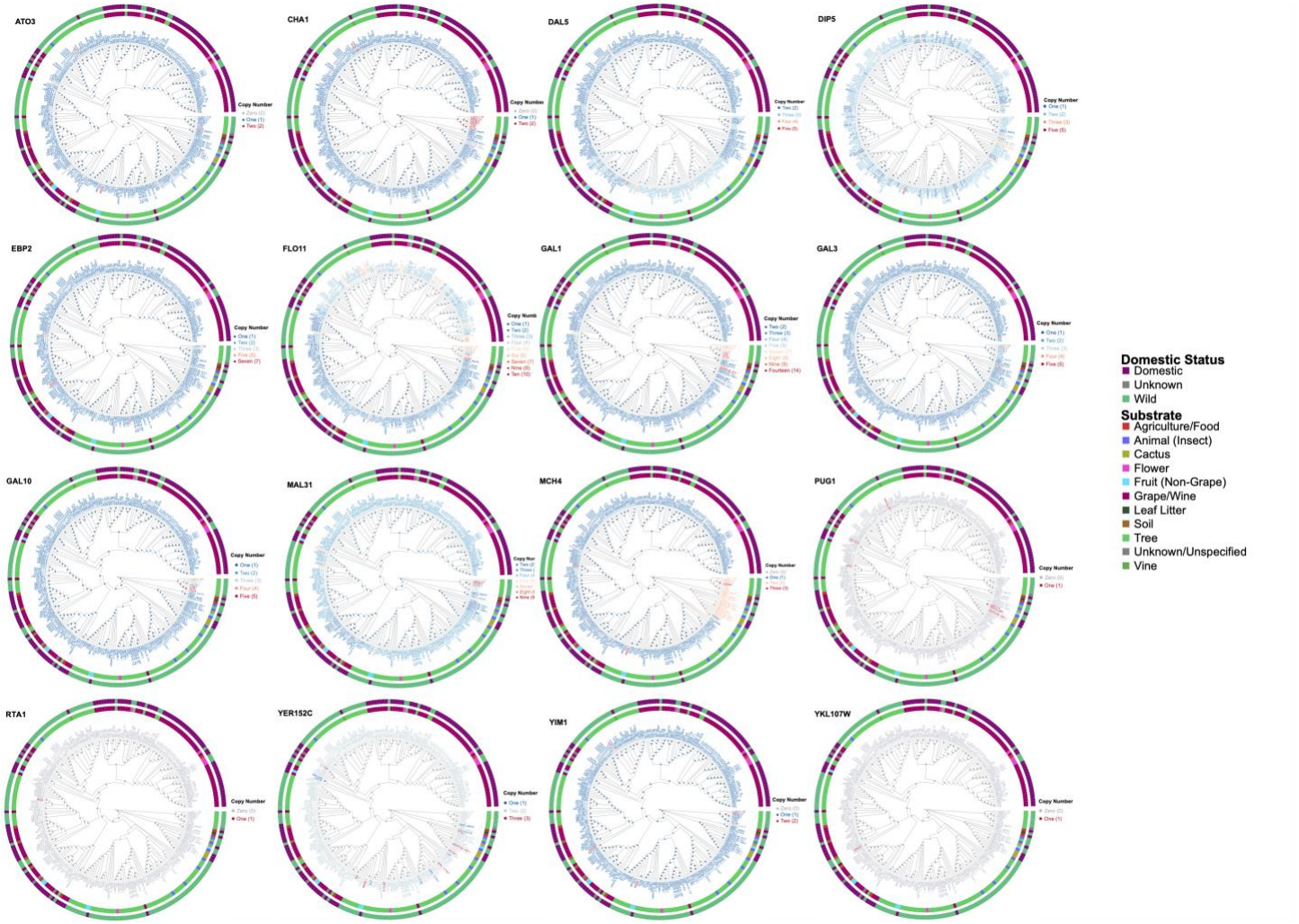


Figure 2.S14: Ancestral copy rate estimation trees. Cladograms of *L. thermotolerans* species tree rooted with *L. quebecensis* showing estimated ancestral copy number over the species evolutionary history for genes of interest. Color legend for copy number changes across genes; node and tip label are colored according to estimated or observed copy number, respectively. Nodes with a bootstrap value of 100 have been circled in black. Colored rings around each cladogram indicate substrate of origin and domestic status of habitat as seen in Figure 2.S12.

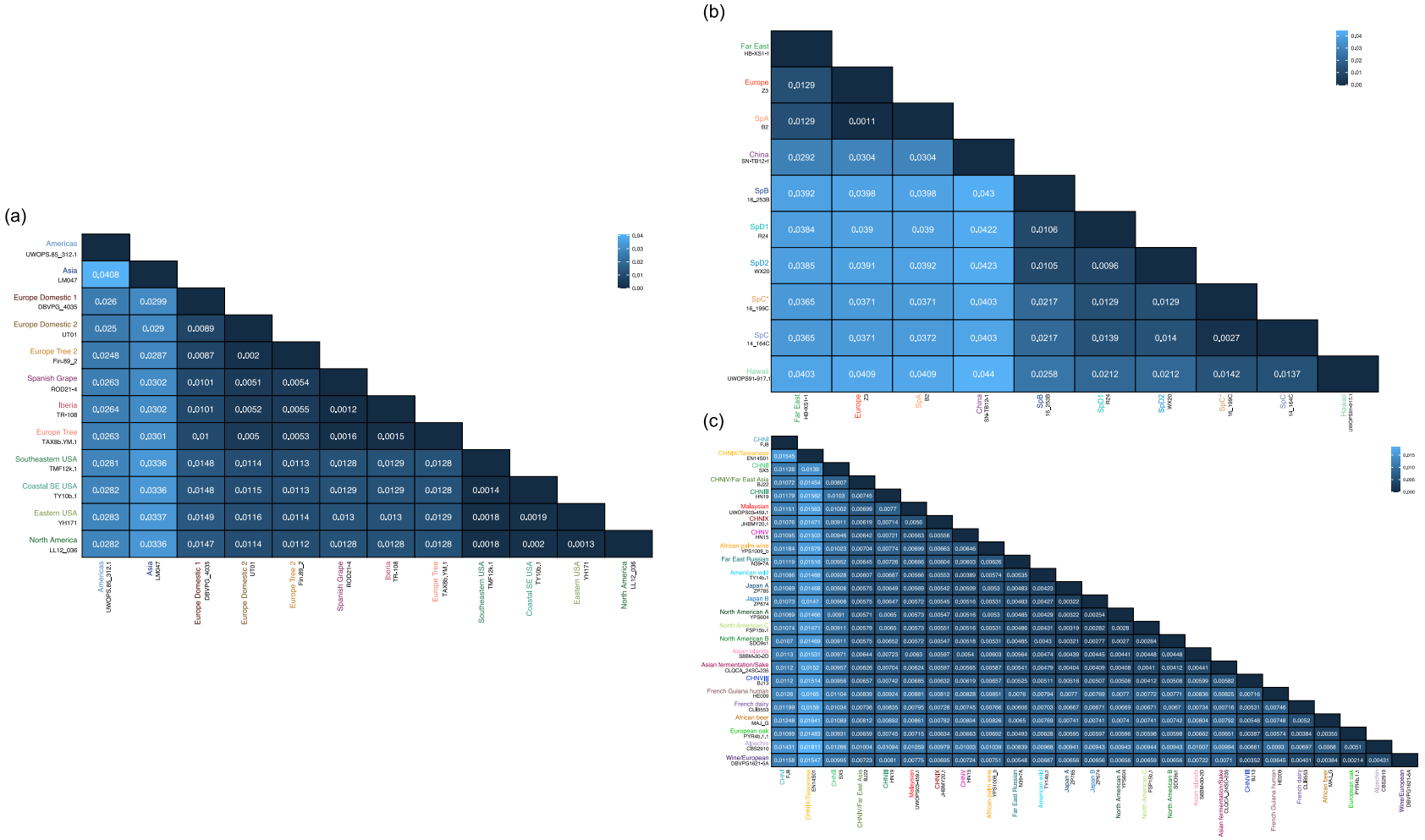


Figure 2.S15: Pairwise distances between *Lachancea thermotolerans* clades are comparable to those between *Saccharomyces paradoxus*. Plots showing the estimated pairwise distances between (a) *L. thermotolerans*, (b) *S. paradoxus*, and (c) *S. cerevisiae* clades. Intersections between each clade are labeled and colored according to pairwise distance. Clade names are colored according to colors previously used to represent them in (a) Figure 1b, (b) He *et al* (2022), and (c) Peña *et al* (2025). The strain used as a representative for each clade is listed beneath each clade name.

CHAPTER 3

GROWTH RATE DIFFERENCES AT HIGH TEMPERATURES ARE DRIVEN BY GENETIC BACKGROUND IN WILD *LACHANCEA THERMOTOLERANS* [1]

[1] Ward, A.K., Hose, J., McKibben, M.M.M, Gasch, A.P., & Bensasson, D. To be submitted to a peer-reviewed journal.

ABSTRACT

As rapid climate change continues to threaten global biodiversity, understanding how eukaryotic species adapt to their environments is essential. Laboratory experiments in the baker's yeast *Saccharomyces cerevisiae* suggest that adaptation is rapid and occurs frequently in microbes, yet little work has been done examining local adaptation in wild populations of fungal microorganisms. Here, we investigate the role of climate in adaptation of European and North American populations of the widely distributed yeast *Lachancea thermotolerans*. We found significant among-strain variation for growth rate across temperatures but were unable to explain this using climatic factors at geographic origin. Variation in growth across temperatures is mostly explained by a single divergence event, leading to wild populations native to North America growing more slowly than their conspecifics. Our results suggest that local temperature adaptation in natural populations of microbial species may be less common than expected.

KEYWORDS: adaptation, wild yeast, temperature-dependent fitness

INTRODUCTION

Rapid global climate change presents a major threat to global biodiversity, as species are unable to adapt at the rate niches are changing (Jezkova & Wiens, 2016; Quintero & Wiens, 2013). Understanding adaptive processes and identifying which organisms may be most vulnerable to temperature changes is critical for species preservation; however, there are few cost-effective eukaryotic models for thermal adaptation (Bay et al., 2017; Pacifici et al., 2015). Yeast species provide cost effective, tractable model systems that allow us to focus on universal eukaryotic cellular processes (Botstein & Fink, 2011). Past experimental evolution studies resulted in rescue mutations and adaptation in laboratory *Saccharomyces* yeast, suggesting that adaptation is a rapid and commonly occurring process for microbes (Fay et al., 2023; Lang et al.,

2011; Payen et al., 2013; Voordeckers et al., 2015; Wang et al., 2024). Furthermore, previous work identified local thermal adaptation in natural forest populations of *Saccharomyces paradoxus* (Leducq et al., 2014).

Lachancea thermotolerans, which diverged from the lineage leading to the *Saccharomyces* yeasts 125-150 million years ago (Shen et al., 2018), is known for its wide distribution across geographic regions and host substrates (Hranilovic et al., 2017). The species has primarily been studied for its impact on wine (Benito, 2018; Castrillo & Blanco, 2023; Jolly et al., 2014; Vilela, 2018) and beer (Canonico et al., 2019; Domizio et al., 2016; Osburn et al., 2018). In addition to its use in the fermentation industry, *L. thermotolerans* is an insect symbiont that has shown temperature dependent colonization in the microbiomes of both pests and agriculturally significant pollinators (Babcock et al., 2018; Kogan et al., 2023). *Lachancea thermotolerans* frequently co-occurs in natural environments with *S. cerevisiae* and *S. paradoxus* (Robinson et al., 2016; Xia et al., 2017) and with *S. cerevisiae* in winemaking environments (Binati et al., 2020; Gobbi et al., 2013). In fact, *L. thermotolerans* and *S. cerevisiae* have comparable domestication histories, with domesticated lineages in Europe for both species appearing to arise from the Mediterranean basin and showing gene gains and losses associated with adaptation to the winemaking environment (Almeida et al., 2015; Banilas et al., 2016; Hranilovic et al., 2018; Liti et al., 2009; Vicente et al., 2025). These ecological and genetic similarities, along with the large genetic distance between the *Lachancea* and *Saccharomyces* genera, make these species a compelling model system for increasing our understanding of temperature adaptation in natural environments, thus testing for parallel evolution.

We used North American and European *L. thermotolerans* to examine whether this non-*Saccharomyces* yeast could be adapted to local climatic conditions. Fifty-two strains from eight

distinct genetic lineages, with natural distributions across a range of temperatures and precipitation levels, were assayed for growth rate for a range of experimental temperatures (25 - 42 °C). Although one lineage grew slower than its conspecifics, we observed no evidence for local adaptation to temperature. These results suggest that adaptation of natural populations to local temperatures in microbial species may be less common than expected from experimental evolution studies.

MATERIALS AND METHODS

Yeast strains and phenotyping assays

Lachancea thermotolerans strains isolated from across Europe and North America and representing eight genetically distinct lineages (see Chapter 2) were selected for phenotyping (N = 57, Figure 3.1, Table 3.S1; Chapter 2; Osburn et al., 2018; Robinson et al., 2016). Strains were prepared for phenotyping in liquid YPD media on 96-well plates; 60 samples per plate including positive and negative controls. The first plate used *Saccharomyces cerevisiae* and *S. paradoxus* isolated from the same site with known thermal growth differences (Sweeney et al., 2004) as positive controls; the second plate included 9 *L. thermotolerans* strains repeated from the first as positive controls (Table 3.S1). There were three negative controls included in the first plate and two in the second (Table 3.S2). Strains were randomly assigned to wells, avoiding outer rows (Row 1, 12) and columns (Column A, H).

High throughput phenotyping was performed at the University of Wisconsin - Madison. Phenotypic characterization of yeasts was conducted by measuring optical density (OD) as strains grew to saturation across five different temperatures (25 °C, 30 °C, 37 °C, 39 °C, 42 °C) using an Infinite® 200 PRO microplate reader with i-control™ software. Screens of each of the

two plates were repeated in triplicate, with each replicate conducted on a different day. Twelve reads of OD were taken from each well at thirty-minute intervals and OD measurements were averaged for each time point. To determine optimal conditions and data collection intervals for phenotypic characterization, 14 *L. thermotolerans* strains from a broad geographic range were included on the first plate alongside strains from other species. Of these *L. thermotolerans* strains, 9 were repeated in the second plate with the remaining 41 strains (Table 3.S1).

Estimating growth rate

Growth curves using raw data (Figures S1, S2) were visualized using the package 'ggplot2' (v3.5.1; Wickham, 2016) in R (v4.2.1; https://www.R-project.org/). To estimate growth rate for each strain in each replicate, we log transformed growth data and fit a linear regression model to the data and obtained the maximum slope (Hall et al., 2014; Leducq et al., 2014) using the 'growthrates' package (v0.8.5; https://github.com/tpetzoldt/growthrates) in R. In addition to the growth rate, we estimate the lag time to growth, the maximum OD measurement, and model fit (r^2) for each strain (Table 3.S2). We used the mean maximum OD measurement ('max') for each strain to identify a 0.36 OD cutoff (Figure 3.S3) and assign a binomial character of 'No Growth' or 'Growth' to each strain across temperatures (Figures 3.5 and 3.S4, Table 3.S3).

Statistical analysis

All statistical and graphical analyses were completed in R (v4.2.1). Data for the daily maximum temperature, averaged over the hottest month of the year (Tmax) and the precipitation for the warmest quarter of the year (precipitation) was estimated using the WorldClim 2 dataset

(Fick & Hijmans, 2017) and the 'raster' (v3.6-31) R package for every host plant from a single pixel at 30 arc-second resolution. Using a linear regression model (LM), we modeled *L. thermotolerans* growth rate in different temperature conditions replicates (30 °C, 37 °C, 39 °C; Table S2) using an average maximum slope ('mslope') for each strain from three or six as the response variable (Table S3). Strains that were admixed (percent ancestry less than 90% from a single population; N = 6; Chapter 2; Ward et al., 2025) were removed from analyses.

The initial model included four explanatory variables and all their interactions: (i) a three-level factor describing experimental temperature condition, (ii) an eight-level factor of all strains grouped by clade, (iii) Tmax (in °C x 10) as a continuous variable estimated from a single pixel at 30-arc second resolution given the longitude and latitude of each isolate host, and (iv) precipitation (in millimeters) as a continuous variable estimated from a single pixel at 30-arc second resolution given the longitude and latitude of each isolate host. Upon model-checking (Crawley, 2015), we observed that errors were not normally distributed. Mean growth rate was square root transformed or log transformed to normalize errors; however, this did not improve model fit even when single-point outliers were removed or after removal of strains that did not grow (see above) from the analysis.

Given the non-normal distribution of errors within the data set, we used non-parametric tests broken down by each growth temperature: Kruskal-Wallis rank sum tests (Kruskal & Wallis, 1952) to compare mean growth rate within and between (i) the original eight-level factored clade variable, (ii) simplified four-level factored clade variable, (iii) further simplified two-leveled factored clade variable (Figures 3.2 and 3.3, Tables 3.1 and 3.2), and Spearman's rank correlation rho (Glasser & Winter, 1961; Spearman, 1904) to identify possible correlation between mean growth rate and (iv) Tmax or (v) precipitation for all factored levels (Figure 3.4,

Table 3.S4). In deciding critical values for *P*-values, we discuss the Bonferroni method. All statistics were performed in R (v.4.2.1). Maps depicting the temperature growth category for each strain (Figure 3.5) were drawn using the 'maps' and 'ggplot2' packages in R (v.4.2.1).

RESULTS AND DISCUSSION

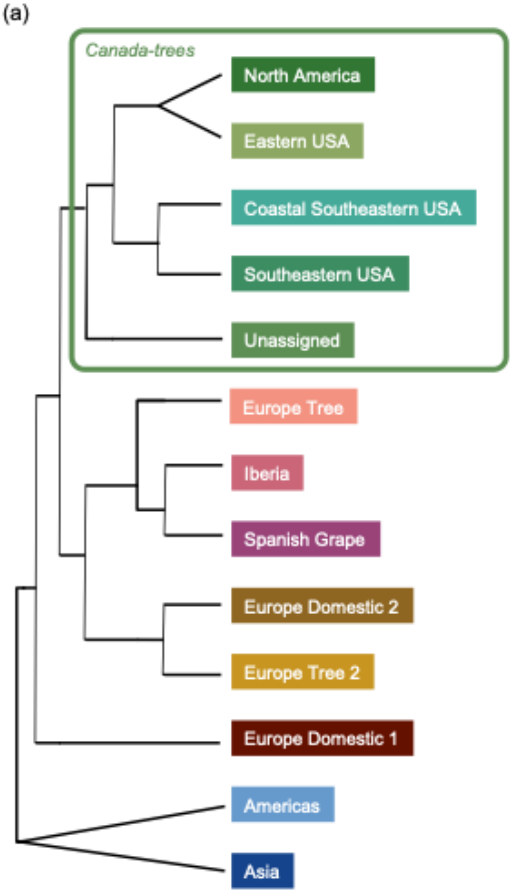
Strains from wild North American genetic lineages grow more slowly than other strains

High throughput phenotyping and genomics make it possible to identify the underlying genetic variation responsible for phenotypic differences driving adaptation (Bomblies & Peichel, 2022; Lai et al., 2024). Using 52 *Lachancea thermotolerans* isolates with known genetic lineages (Table 3.S1), we sought to identify differences in growth rate within and between lineages across multiple temperatures (Figure 3.1, Table 3.1). To understand how genetic background impacts growth rate across temperatures in *L. thermotolerans*, we compared the mean growth rate (maximum slope during exponential growth) for strains across each of the eight assigned genetic lineages represented, excluding admixed strains (Figure 3.1, Table 3.S1).

We found significant differences between growth rates of these eight lineages at 30 and 37 °C (Kruskal Wallis tests: $P < 0.01$; Figure 3.2 and Table 3.2). When we simplified to four lineages by combining (i) all Canada-trees lineages and (ii) Europe Tree and Iberia into Southern Europe, we found that growth rate differences were even more significant (Kruskal Wallis tests: $P < 0.005$). To see what was driving this difference, we further simplified to two groups, Canada-trees and non-Canada-trees, and found a significant difference in growth rate at 30 °C (Wilcoxon test: $P <$ 5e-10). When we removed the Canada-trees lineage from analyses, all significance across temperatures disappeared (Kruskal Wallis tests: $P > 0.05$). Repeating analyses using the two simplified groups, we found that strains in the Canada-trees lineage show

a significantly lower growth rate than strains in non-Canada-trees lineages at 30 and 37 °C (Kruskal Wallis tests: $P < 0.001$; Figure 3.3). As the Canada-trees lineage represents one divergence event, this suggests that the slower growth rate in these strains is a derived state that arose prior to the expansion and diversification of this lineage in North America.

(a)



(b)

	Strain Number	Geographic Origin
Canada-trees	33	
Coastal Southeastern USA	11	Florida, USA Georgia, USA Pennsylvania, USA
Eastern USA	13	Indiana, USA Pennsylvania, USA West Virginia, USA
North America	6	Georgia, USA Indiana, USA New Jersey, USA Pennsylvania, USA
Southeastern USA	2	Georgia, USA North Carolina, USA
Unassigned	1	Indiana, USA
Non-Canada-trees	19	
Americas	3	Florida , USA Pennsylvania, USA
Europe Tree	5	Greece Pennsylvania, USA
Europe Tree 2	8	France United Kingdom
Iberia	3	Portugal
Admixed Strains	6	Greece Georgia, USA New Jersey, USA Pennsylvania, USA

Figure 3.1: ***Lachancea thermotolerans*** **phylogeny and experimental design.** (a) Cladogram showing a simplification of the relationships among *L. thermotolerans* lineages; tip labels are clades described by population structure analyses when K = 15 as described in Chapter 2. (b) Table showing breakdown of strains used in this study by genetic lineage and geographical origin; all but Eastern USA strain YH109 were isolated from wild substrates.

Table 3.1: Summary of growth rate estimates for this study. Table shows the median mean growth rate as estimated for all temperature conditions, replicates, and lineages used in this study. Where no replicates exist, the value is listed as NA.

	25 °C	30 °C	37 °C	39 °C	42 °C
Canada-trees	0.101	0.122	0.0259	0.0135	0.0106
Coastal Southeastern USA	0.0988	0.126	0.0265	0.0135	0.0110
Eastern USA	0.103	0.122	0.0244	0.0129	0.00368
North America	0.101	0.119	0.0265	0.0121	0.0115
Southeastern USA	NA	0.121	0.0291	0.0124	0.00905
Unassigned	NA	0.110	0.0317	0.0175	0.0271
Non-Canada-trees	0.102	0.131	0.0527	0.0206	0.0153
Americas	0.102	0.129	0.104	0.0370	0.0145
Europe Tree	0.0948	0.131	0.0520	0.0254	0.0112
Europe Tree 2	0.104	0.132	0.0544	0.0195	0.0145
Iberia	NA	0.136	0.0497	0.0153	0.0189

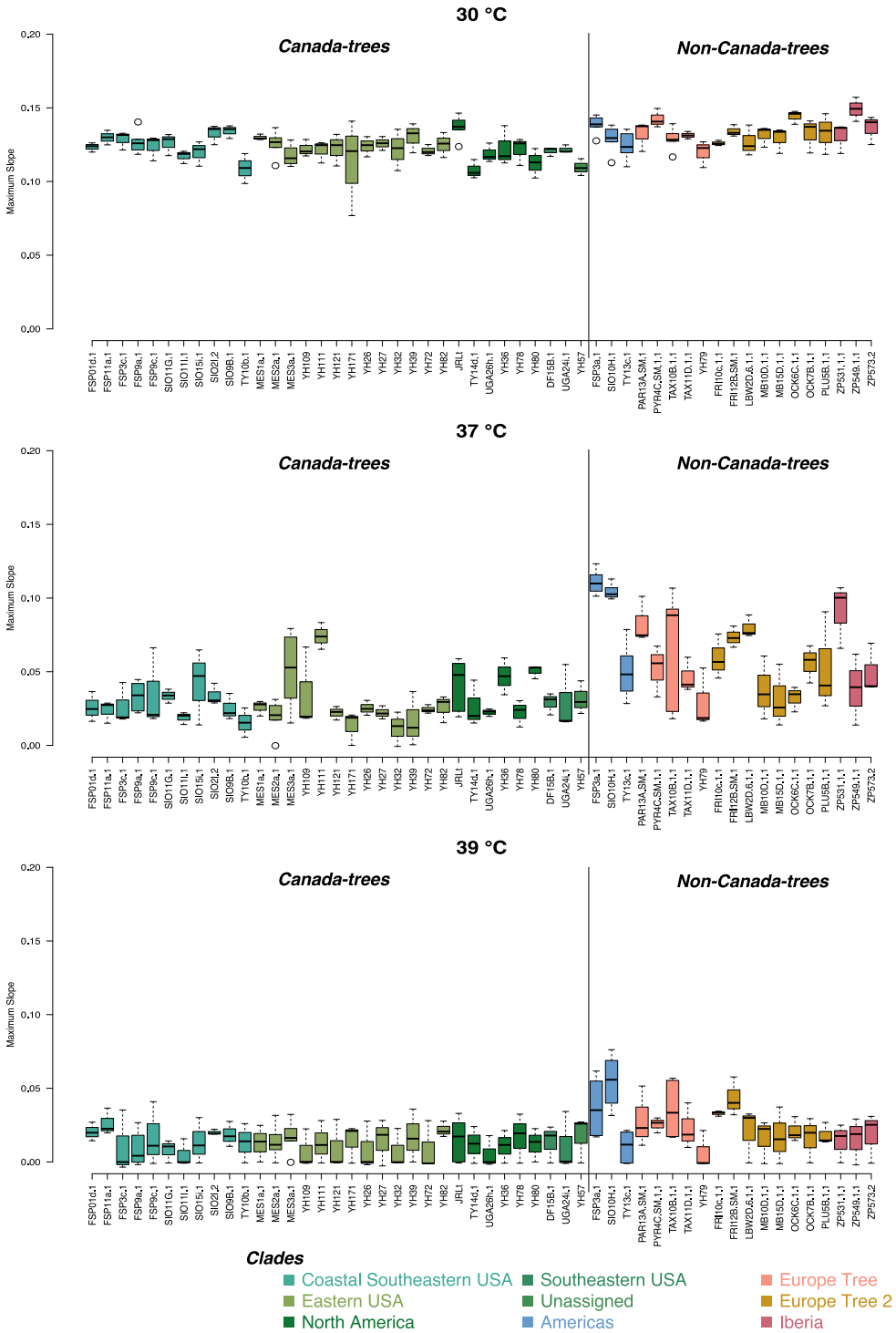


Figure 3.2: Variation in growth rate between strains within the same genetic lineages. Boxplots showing strain maximum slope (growth rate) across experimental temperatures (30 - 39 °C). The black line indicates the delineation of 'Canada-trees' strains, which showed a lower growth rate than other lineages. Boxplot fill corresponds with clade as defined in Figure 3.1a and Chapter 2. Strains are ordered alphabetically by and within clade. Admixed strains were not included.

Table 3.2: Statistically significant variation of thermal growth rate within genetic lineages. *P*-values for Kruskal-Wallis rank sum tests identifying the significant differences in mean within lineages at each temperature after simplifying to two groups. Across temperatures, only lineages within the non-Canada-trees group show significant variation in growth rate at 37 °C.

	Canada-trees	Non-Canada-trees
30 °C	0.0132	0.0182
37 °C	0.0419	2.5×10^{-4}***
39 °C	0.991	0.0236

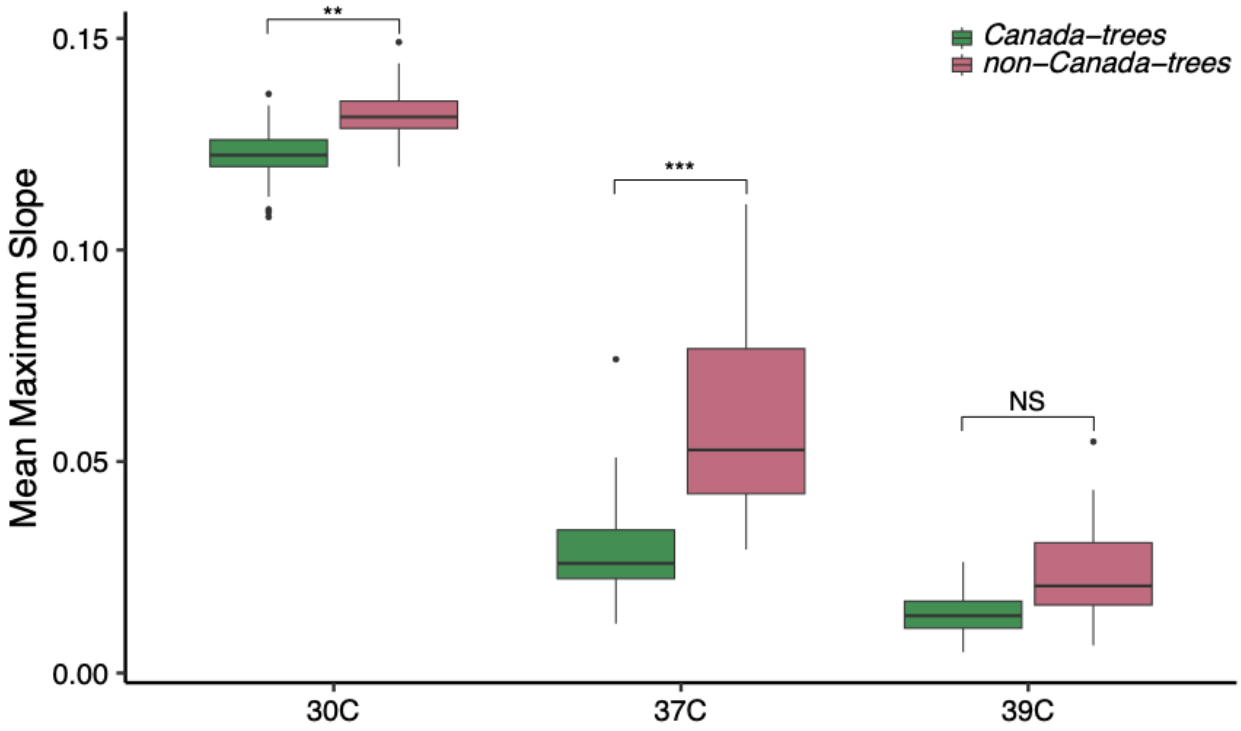


Figure 3.3: Significant differences in growth rate by genetic lineage. Boxplots showing mean maximum slope (growth rate) across all three experimental temperatures for the two genetic groups used in this study. We found significant differences (Kruskal Wallis tests: $P < 0.001$) between Canada-trees and other lineages at 30 °C and 37 °C. Boxplot fill corresponds with genetic lineage.

Significant variation in growth rate within L. thermotolerans lineages

Comparing the fastest growth rate (maximum slope during exponential growth) for each strain across 3-6 replicates, we found significant differences among strains outside of the Canada-trees lineage at 37 °C (Kruskal Wallis tests: $P < 0.01$; Figure 3.2 and Table 3.2). Additionally, we observed variation in maximum growth across temperatures and results that suggest strains within the Canada-trees lineage are less likely to show any growth at 39 °C (Figures 3.5, 3.S4, and 3.S5). These strains were not exposed to artificial selection in the laboratory, so this phenotypic variation likely represents standing genetic variation (Barrett & Schluter, 2008; Prezeworski et al., 2005). Looking within lineages, strains in the Canada-tree lineage do not differ in their maximum slope at 30, 37, or 39 °C (Kruskal Wallis tests: $P > 0.01$; Figure 3.2 and Table 3.2). For future thermal adaptation, *de novo* mutations will likely be needed.

No evidence for strains local thermal adaptation in Lachancea thermotolerans

While growth rate differences between conspecifics at different temperatures have been seen in lab environments through experimental evolution studies (Abrams & Brem, 2022; Fay et al., 2023), they have also been observed in wild populations. For example, marine bacteria in warmer temperatures grow more slowly than their conspecifics (Abreu et al., 2023), and previous work in the budding yeast *Saccharomyces paradoxus* has shown lineage-specific local adaptation in wild strains (Leducq et al., 2014).

Could differences among strains in growth rates be driven by local climatic differences in *L. thermotolerans*? To address this question, we looked for a correlation between mean maximum slope during exponential growth and (i) maximum temperature at site of isolation

(Tmax) or (ii) precipitation during the warmest quarter (precipitation). While Tmax did not show a significant relationship with growth rate (Spearman test, $P > 0.02$), we identified a correlation between growth rate and precipitation at 30 and 37°C (Figure 3.4; Spearman tests: $P = 0.01$), and at 39°C; though the latter would not be significant after Bonferroni correction (Spearman test: $P = 0.03$). This effect appears to be caused by the differences between genetic lineages rather than precipitation itself; there are no correlations between growth rates and precipitation or summer temperature when looking within the Canada lineage or after excluding this lineage (Spearman tests: $P > 0.1$; Table 3.S3). Strains in non-Canada-trees lineages co-occur with Canada-trees strains in North America (Figures 3.1 and 3.5, Table 3.S1), emphasizing that genetic background, not geographic origin, drives growth rates in *L. thermotolerans*.

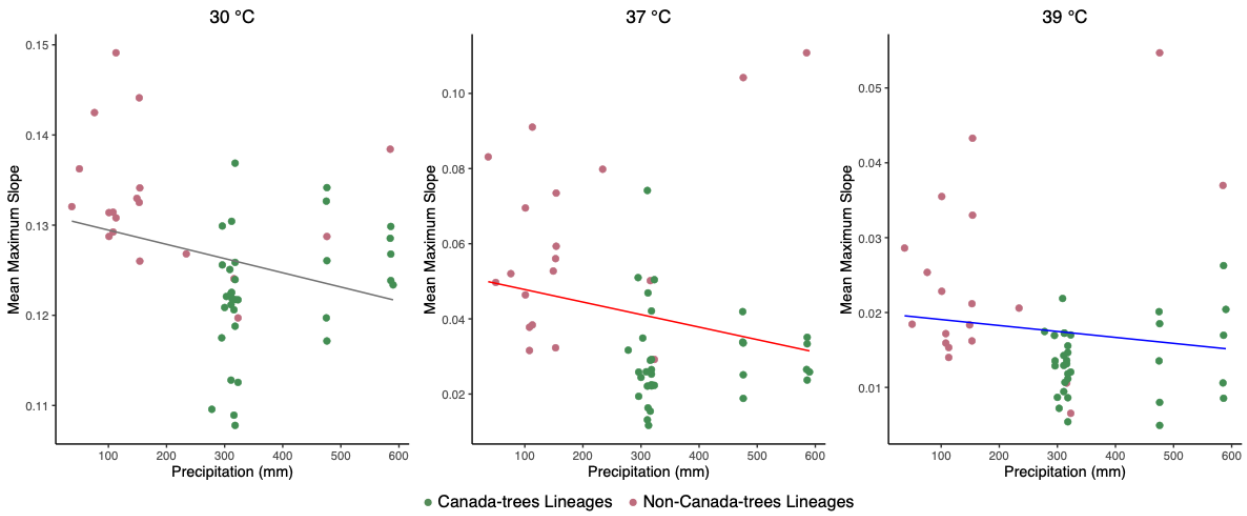


Figure 3.4: Possible correlation between summer precipitation and growth rate could be explained by lineage effects. Scatterplot with lines of best fit showing the relationship between precipitation during the warmest quarter of the year (precipitation) and mean maximum slope (growth rate) at 30 °C (Spearman test: P = 0.01), 37 °C (Spearman test: P = 0.01), and 39 °C (Spearman test: P = 0.03) is likely caused by the lower precipitation levels seen at the geographic source of Canada-tree strains. Points are colored according to whether they belong to Canada-trees or non-Canada-trees lineages.

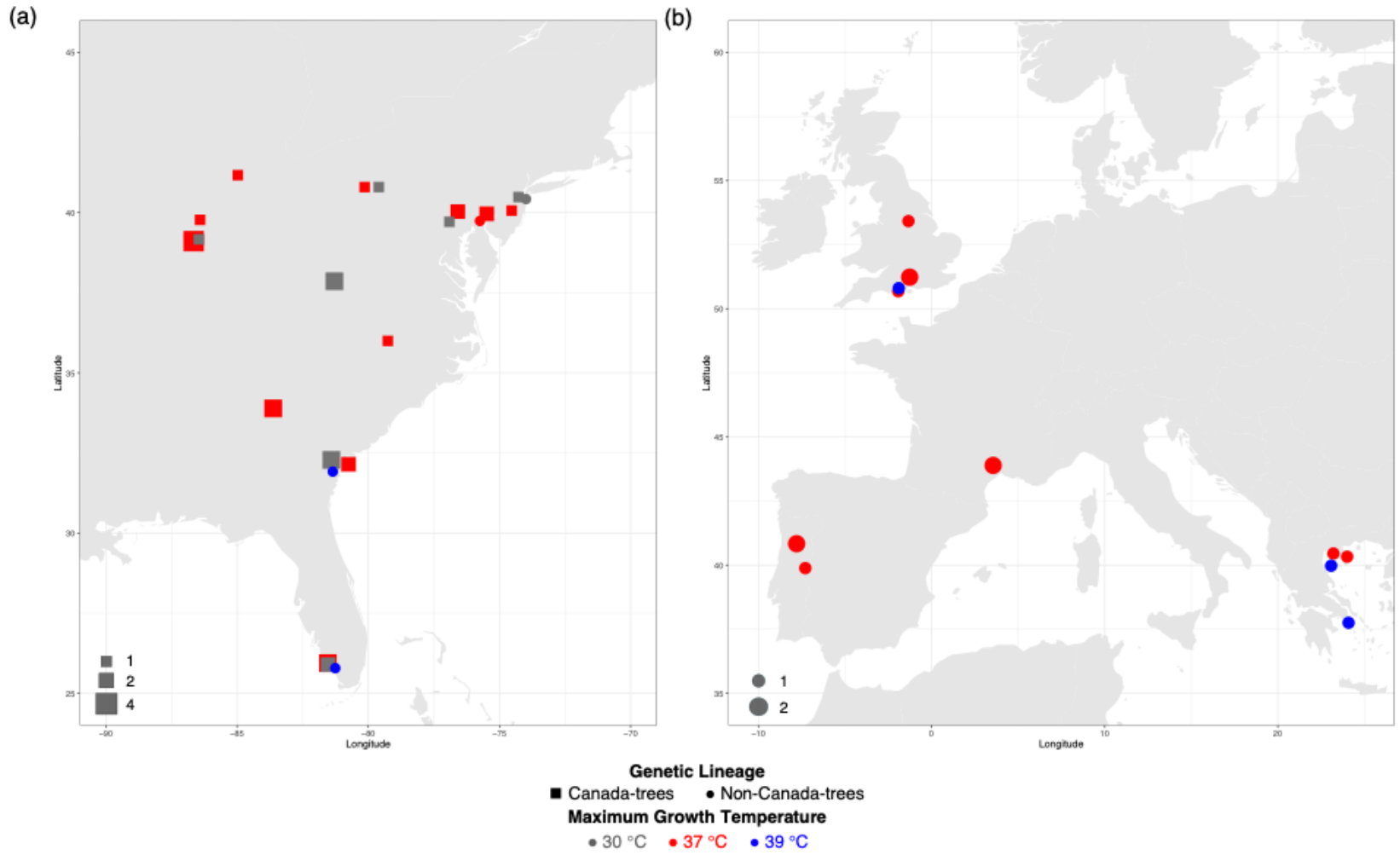


Figure 3.5: Maximum growth temperature and geographic sources of strains. Map showing the (a) North American and (b) European distributions of *L. thermotolerans* strains used in this study. Point shape was assigned based on whether they are from Canada-trees or non-Canada-trees lineages. Point sizes are scaled according to square-root-transformed sample sizes; in panel (a), all strains from non-Canada-trees lineages are single individuals. Isolate locations are colored according to highest temperature at which they grew (mean maximum growth exceeded 0.36 OD). Admixed strains were not included.

While we did not identify local thermal adaptation in *L. thermotolerans*, other environmental factors may be contributing to reduced growth rate in Canada-trees strains. In *S. cerevisiae*, laboratory experiments show that increased stress tolerance, such as survival of freeze-thaw cycles or in acidic or oxidative environments, is associated with lower growth rate (Zakrzewska et al., 2011). Previous work in *L. thermotolerans* has shown lactate production during fermentation is accompanied by decreased growth rate and an increase in the expression of stress-response genes (Battjes et al., 2023). It is possible that the reduced growth we see in the Canada-trees lineage results from the pleiotropic effect of loci responsible for a beneficial phenotype not measured in this study. Further exploration of the Canada-trees lineage examining potential trade-offs through additional phenotypic screening for growth in other environmental conditions, such as freeze-thaw cycles or alternative carbon or nitrogen sources, is needed to identify what advantage – if any – may be linked with reduced growth rate in *L. thermotolerans*.

The lack of thermal adaptation seems surprising because microbes can live in potentially large population, have rapid generation times and a past study of yeast in natural environments showed local thermal adaptation in a northern North American lineage of *S. paradoxus* (Leducq et al. 2014). Past experiments show rapid adaptation is possible for budding yeast in highly controlled laboratory experiments (Fay et al., 2023; Payen et al., 2013; Voordeckers et al., 2015; Wang et al., 2024), yet we do not see evidence that *L. thermotolerans* adapted as it expanded across geographic regions with different climates. This work highlights the need to examine evolutionary outcomes in natural environments, which are more complex than in the laboratory.

SUPPLEMENTARY TABLE INFORMATION

Due to their size, the following supplemental tables for this chapter are available on GitHub here: https://github.com/bensassonlab/data/tree/master/Ward2025_Dissertation

Table 3.S1. *Lachancea thermotolerans* strains used in this study. Metadata was compiled from publicly available data. Includes geographic and ecological origin, as well as genetic background (clade, admixture status, divergence event) and the phenotyping plate for each strain.

Table 3.S2. Calculated growth rate data from all replicates. Calculated growth rate data from the linear regression analysis of growth curves across all replicates and temperatures, including: slope at exponential growth ('mslope'), lag to growth ('lag'), maximum growth value, ('max') and R-squared ('rsq') value as a measure of model fit. Max temperature ('Tmax') is in degrees Celsius and precipitation during the warmest quarter is measured in millimeters. Whether or not there was growth based on maximum growth value of OD > 0.36 is included as a binary character under 'anyGrowth.'

Table 3.S3. Mean growth information. Mean value of maximum growth ('meanMax') and growth rate ('meanMSlope') across all replicates of temperatures 30C-39C for each strain, excluding negatives. Whether or not there was growth based on mean maximum growth value of OD > 0.36 is included as a binary character under 'anyGrowth.'

AUTHORS' CONTRIBUTIONS

Research was conceptualized and designed by AKW, DB, and APG. AKW and MMM prepared strains for phenotyping. JH performed high-throughput phenotyping. AKW and MMM

developed a script to identify growth rate parameters. AKW performed analyses and data visualization. AKW wrote this chapter with feedback from DB.

REFERENCES

Abrams, M. B., & Brem, R. B. (2022). Temperature-dependent genetics of thermotolerance between yeast species. *Frontiers in Ecology and Evolution*, *10*, 859904. https://doi.org/10.3389/fevo.2022.859904

Abreu, C. I., Bello, M. D., Bunse, C., Pinhassi, J., & Gore, J. (2023). Warmer temperatures favor slower-growing bacteria in natural marine communities. *SCIENCE ADVANCES*, *9*, eade8352. https://doi.org/10.1126/sciadv.ade8352

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

Babcock, T., Borden, J., Gries, R., Carroll, C., Moore, M., & Gries, G. (2018). *Lachancea thermotolerans*, a Yeast Symbiont of Yellowjackets, Enhances Attraction of Three Yellowjacket Species (*Hymenoptera: Vespidae*) to Fruit Powder. *Environmental Entomology*, *47*(6), 1553–1559. https://doi.org/10.1093/ee/nvy139

Banilas, G., Sgouros, G., & Nisiotou, A. (2016). Development of microsatellite markers for *Lachancea thermotolerans* typing and population structure of wine-associated isolates. *Microbiological Research*, *193*, 1–10. https://doi.org/10.1016/j.micres.2016.08.010

Barrett, R., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Battjes, J., Melkonian, C., Mendoza, S. N., Haver, A., Al-Nakeeb, K., Koza, A., Schrubbers, L., Wagner, M., Zeidan, A. A., Molenaar, D., & Teusink, B. (2023). Ethanol-lactate transition of *Lachancea thermotolerans* is linked to nitrogen metabolism. *Food Microbiology*, *110*, 104167. https://doi.org/10.1016/j.fm.2022.104167

Bay, R. A., Rose, N., Barrett, R., Bernatchez, L., Ghalambor, C. K., Lasky, J. R., Brem, R. B., Palumbi, S. R., & Ralph, P. (2017). Predicting Responses to Contemporary Environmental Change Using Evolutionary Response Architectures. *The American Naturalist*, *189*(5), 463–473. https://doi.org/10.1086/691233

Benito, S. (2018). The impacts of *Lachancea thermotolerans* yeast strains on winemaking. *Applied Microbiology and Biotechnology*, *102*(16), 6775–6790. https://doi.org/10.1007/s00253-018-9117-z

Binati, R. L., Lemos Junior, W. J. F., Luzzini, G., Slaghenaufi, D., Ugliano, M., & Torriani, S. (2020). Contribution of non-*Saccharomyces* yeasts to wine volatile and sensory diversity: A study on *Lachancea thermotolerans*, *Metschnikowia spp*. And *Starmerella bacillaris* strains isolated in Italy. *International Journal of Food Microbiology*, *318*, 108470. https://doi.org/10.1016/j.ijfoodmicro.2019.108470

Bomblies, K., & Peichel, C. L. (2022). Genetics of adaptation. *Proceedings of the National Academy of Sciences*, *119*(30), e2122152119. https://doi.org/10.1073/pnas.2122152119

Botstein, D., & Fink, G. R. (2011). Yeast: An Experimental Organism for 21st Century Biology. *Genetics*, *189*(3), 695–704. https://doi.org/10.1534/genetics.111.130765

Brown, M. V., Ostrowski, M., Messer, L. F., Bramucci, A., Van De Kamp, J., Smith, M. C., Bissett, A., Seymour, J., Hobday, A. J., & Bodrossy, L. (2024). A marine heatwave drives significant shifts in pelagic microbiology. *Communications Biology*, *7*(1), 125. https://doi.org/10.1038/s42003-023-05702-4

Canonico, L., Galli, E., Ciani, E., Comitini, F., & Ciani, M. (2019). Exploitation of Three Non-Conventional Yeast Species in the Brewing Process. *Microorganisms*, *7*(1). https://doi.org/10.3390/microorganisms7010011

Castrillo, D., & Blanco, P. (2023). Characterization of Indigenous Non-*Saccharomyces* Yeast Strains with Potential Use in Winemaking. *Frontiers in Bioscience-Elite*, *15*(1), 1. https://doi.org/10.31083/j.fbe1501001

Crawley, M. J. (2015). *Statistics: An introduction using R* (Second edition). Wiley.

Domizio, P., House, J. F., Joseph, C. M. L., Bisson, L. F., & Bamforth, C. W. (2016). *Lachancea thermotolerans* as an alternative yeast for the production of beer. *Journal of the Institute of Brewing*, *122*(4), 599–604. https://doi.org/10.1002/jib.362

Fay, J. C., Alonso-del-Real, J., Miller, J. H., & Querol, A. (2023). Divergence in the *Saccharomyces* Species' Heat Shock Response Is Indicative of Their Thermal Tolerance. *Genome Biology and Evolution*, *15*(11), evad207. https://doi.org/10.1093/gbe/evad207

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Glasser, G. J., & Winter, R. F. (1961). Critical Values of the Coefficient of Rank Correlation for Testing the Hypothesis of Independence. *Biometrika*, *48*(3/4), 444–448.

Gobbi, M., Comitini, F., Domizio, P., Romani, C., Lencioni, L., Mannazzu, I., & Ciani, M. (2013). *Lachancea thermotolerans* and *Saccharomyces cerevisiae* in simultaneous and sequential co-fermentation: A strategy to enhance acidity and improve the overall quality of wine. *Food Microbiology*, *33*(2), 271–281. https://doi.org/10.1016/j.fm.2012.10.004

Gonçalves, P., Valério, E., Correia, C., De Almeida, J. M. G. C. F., & 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

Hall, B. G., Acar, H., Nandipati, A., & Barlow, M. (2014). Growth Rates Made Easy. *Molecular Biology and Evolution*, *31*(1), 232–238. https://doi.org/10.1093/molbev/mst187

Hranilovic, A., Bely, M., Masneuf-Pomarede, I., Jiranek, V., & Albertin, W. (2017). The evolution of *Lachancea thermotolerans* is driven by geographical determination, 102nthropization and flux between different ecosystems. *PloS ONE*, *12*(9). https://doi.org/10.1371/journal.pone.0184652

Hranilovic, A., Gambetta, J. M., Schmidtke, L., Boss, P. K., Grbin, P. R., Masneuf-Pomarede, I., Bely, M., Albertin, W., & Jiranek, V. (2018). Oenological traits of *Lachancea thermotolerans* show signs of domestication and allopatric differentiation. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-33105-7

Jezkova, T., & Wiens, J. J. (2016). Rates of change in climatic niches in plant and animal populations are much slower than projected climate change. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1843), 20162104. https://doi.org/10.1098/rspb.2016.2104

Jolly, N. P., Varela, C., & Pretorius, I. S. (2014). Not your ordinary yeast: Non- *Saccharomyces* yeasts in wine production uncovered. *FEMS Yeast Research*, *14*(2), 215–237. https://doi.org/10.1111/1567-1364.12111

Kogan, H. V., Elikan, A. B., Glaser, K. F., Bergmann, J. M., Raymond, L. M., Prado-Irwin, S. R., & Snow, J. W. (2023). Colonization of Honey Bee Digestive Tracts by Environmental Yeast *Lachancea thermotolerans* Is Naturally Occurring, Temperature Dependent, and Impacts the Microbiome of Newly Emerged Bees. *Microbiology Spectrum*, *11*(2), e05194-22. https://doi.org/10.1128/spectrum.05194-22

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, *47*(260), 583–621.

Lai, W.-Y., Nolte, V., Jakšić, A. M., & Schlötterer, C. (2024). Evolution of Phenotypic Variance Provides Insights into the Genetic Basis of Adaptation. *Genome Biology and Evolution*, *16*(4), evae077. https://doi.org/10.1093/gbe/evae077

Lang, G. I., Botstein, D., & Desai, M. M. (2011). Genetic Variation and the Fate of Beneficial Mutations in Asexual Populations. *Genetics*, *188*(3), 647–661. https://doi.org/10.1534/genetics.111.128942

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

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, 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

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

Pacifici, 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), 215–224. https://doi.org/10.1038/nclimate2448

Payen, C., Di Rienzi, S. C., Ong, G. T., Pogachar, J. L., Sanchez, J. C., Sunshine, A. B., Raghuraman, M. K., Brewer, B. J., & Dunham, M. J. (2013). The Dynamics of Diverse Segmental Amplifications in Populations of *Saccharomyces cerevisiae* Adapting to Strong Selection. *G3: Genes|Genomes|Genetics*, *4*(3), 399–409. https://doi.org/10.1534/g3.113.009365

Prezeworski, M., Coop, G., & Wall, J. D. (2005). THE SIGNATURE OF POSITIVE SELECTION ON STANDING GENETIC VARIATION. *Evolution*, *59*(11), 2312–2323. https://doi.org/10.1111/j.0014-3820.2005.tb00941.x

Quintero, I., & Wiens, J. J. (2013). Rates of projected climate change dramatically exceed past rates of climatic niche evolution among vertebrate species. *Ecology Letters*, *16*(8), 1095–1103. https://doi.org/10.1111/ele.12144

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

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

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, *15*(1), 72–101.

Sweeney, J., Kuehne, H., & Sniegowski, P. (2004). Sympatric natural and populations have different thermal growth profiles. *FEMS Yeast Research*, *4*(4–5), 521–525. 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

Vicente, J., Friedrich, A., Schacherer, J., Freel, K., Marquina, D., & Santos, A. (2025). Whole-Genome Sequencing and Phenotyping Reveal Specific Adaptations of *Lachancea thermotolerans* to the Winemaking Environment. *Molecular Ecology*, e17667. https://doi.org/10.1111/mec.17667

Vilela, A. (2018). *Lachancea thermotolerans*, the Non-*Saccharomyces* Yeast that Reduces the Volatile Acidity of Wines. *Fermentation*, *4*(3), 56. https://doi.org/10.3390/fermentation4030056

Voordeckers, K., Kominek, J., Das, A., Espinosa-Cantú, A., De Maeyer, D., Arslan, A., Van Pee, M., van der Zande, E., Meert, W., Yang, Y., Zhu, B., Marchal, K., DeLuna, A., Van Noort, V., Jelier, R., & Verstrepen, K. J. (2015). Adaptation to High Ethanol Reveals Complex Evolutionary Pathways. *PLOS Genetics*, *11*(11), e1005635. https://doi.org/10.1371/journal.pgen.1005635

Wang, P., Driscoll, W. W., & Travisano, M. (2024). Genomic sequencing reveals convergent adaptation during experimental evolution in two budding yeast species. *Communications Biology*, *7*(1), 1–9. https://doi.org/10.1038/s42003-024-06485-y

Ward, A. K., Scopel, E. F. C., Shuman, B., Momany, M., & Bensasson, D. (2025). Low level contamination confounds population genomic analysis. *bioRxiv*. https://doi.org/10.1101/2025.01.17.633387

Wickham, H. (2016). *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Xia, W., Nielly-Thibault, L., Charron, G., Landry, C. R., Kasimer, D., Anderson, J. B., & Kohn, L. M. (2017). Population genomics reveals structure at the individual, host-tree scale and persistence of genotypic variants of the undomesticated yeast *Saccharomyces paradoxus* in a natural woodland. *Molecular Ecology*, *26*(4), 995–1007. https://doi.org/10.1111/mec.13954

Zakrzewska, A., van Eikenhorst, G., Burggraaff, J. E. C., Vis, D. J., Hoefsloot, H., Delneri, D., Oliver, S. G., Brul, S., & Smits, G. J. (2011). Genome-wide analysis of yeast stress

survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. *Molecular Biology of the Cell*, *22*(22), 4435–4446. https://doi.org/10.1091/mbc.e10-08-0721

Table 3.S4: Climatic variables do not significantly impact growth rate in *Lachancea thermotolerans*. P-values for Spearman correlation tests for relationships between growth rate and two climatic variables: max summer temperature (Tmax) and precipitation during the warmest quarter (Precipitation) when grouped by genetic lineage.

	Tmax	Precipitation
30 °C		
Canada-trees	0.317	0.116
Non-Canada-trees	0.974	0.155
37 °C		
Canada-trees	0.245	0.675
Non-Canada-trees	0.949	0.326
39 °C		
Canada-trees	0.593	0.912
Non-Canada-trees	0.827	0.889

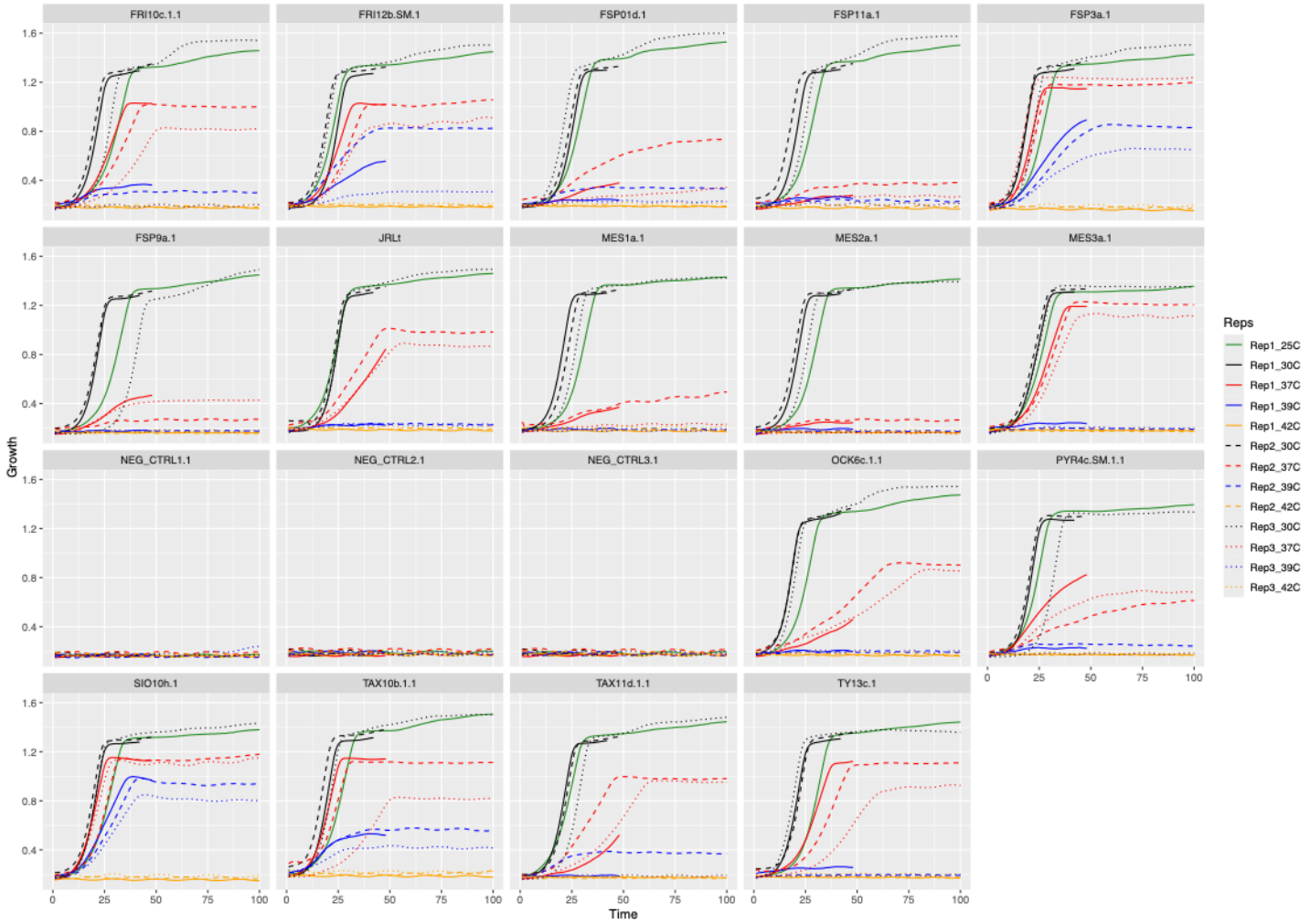


Figure 3.S1: Growth curves for first phenotyping plate. Growth curves for each of the 16 *Lachancea thermotolerans* strains and 3 negative controls from the pilot phenotyping plate over 50 hours. Plots show yeast density in YPD media measured in OD; OD was recorded every 30 minutes. Curves are colored according to growth temperature; line type was assigned based on replicate number. Only one replicate for the 25 °C condition was conducted. Strains from the first replicate (Rep1) were grown for 21 (30 °C) or 24 (30 – 37 °C) hours; all other replicates were grown for 72 hours.

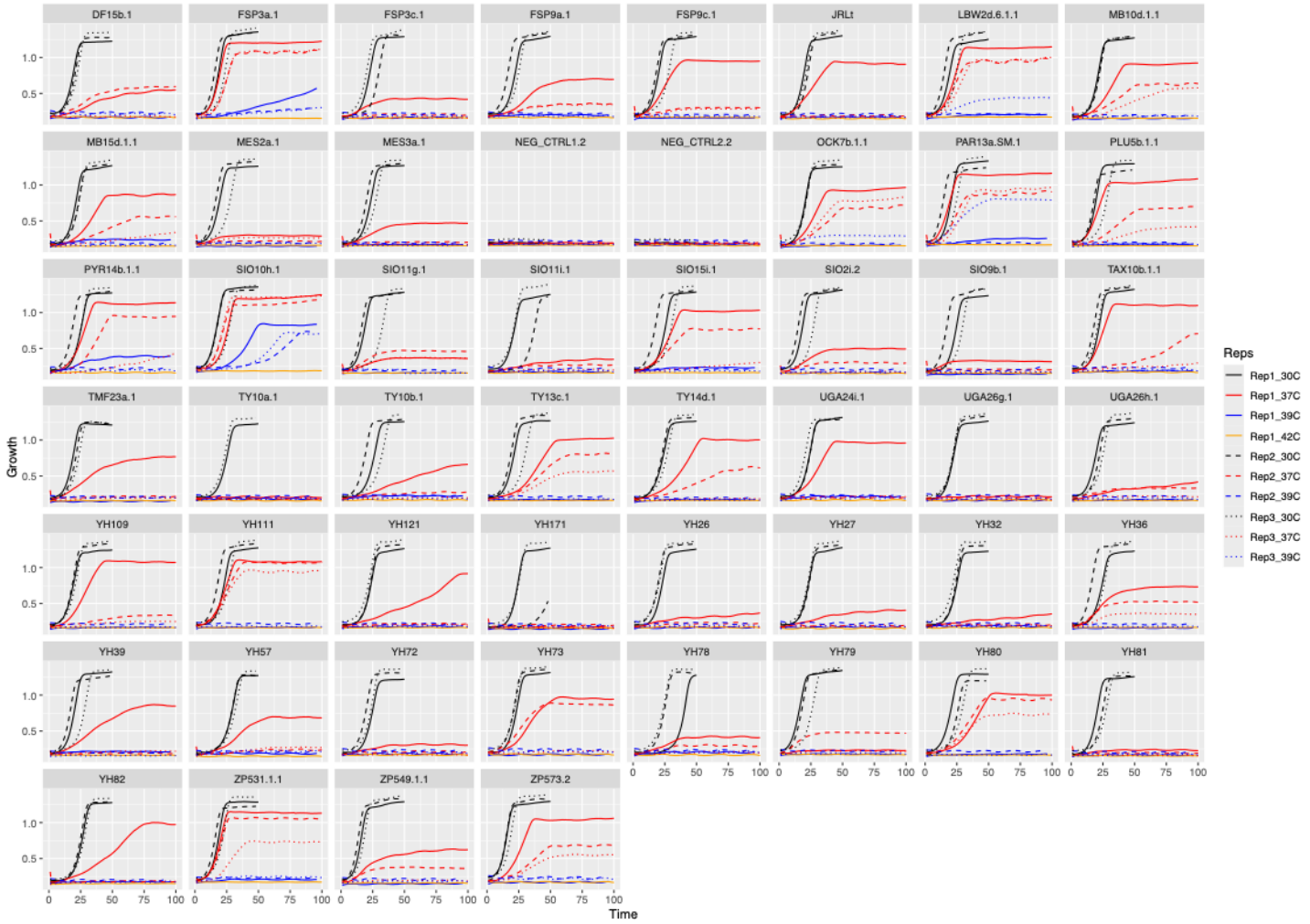


Figure 3.S2: Growth curves for second phenotyping plate. Growth curves for each of the 50 *Lachancea thermotolerans* strains and 2 negative controls from the second phenotyping plate over 50 hours. Plots show yeast density in YPD media measured in OD; OD was recorded every 30 minutes. Curves are colored according to growth temperature; line type was assigned based on replicate number. Only one replicate for the 42 °C condition was conducted. Replicates for 30 °C were grown for 25 hours; all other temperatures were grown for a minimum of 48 hours.

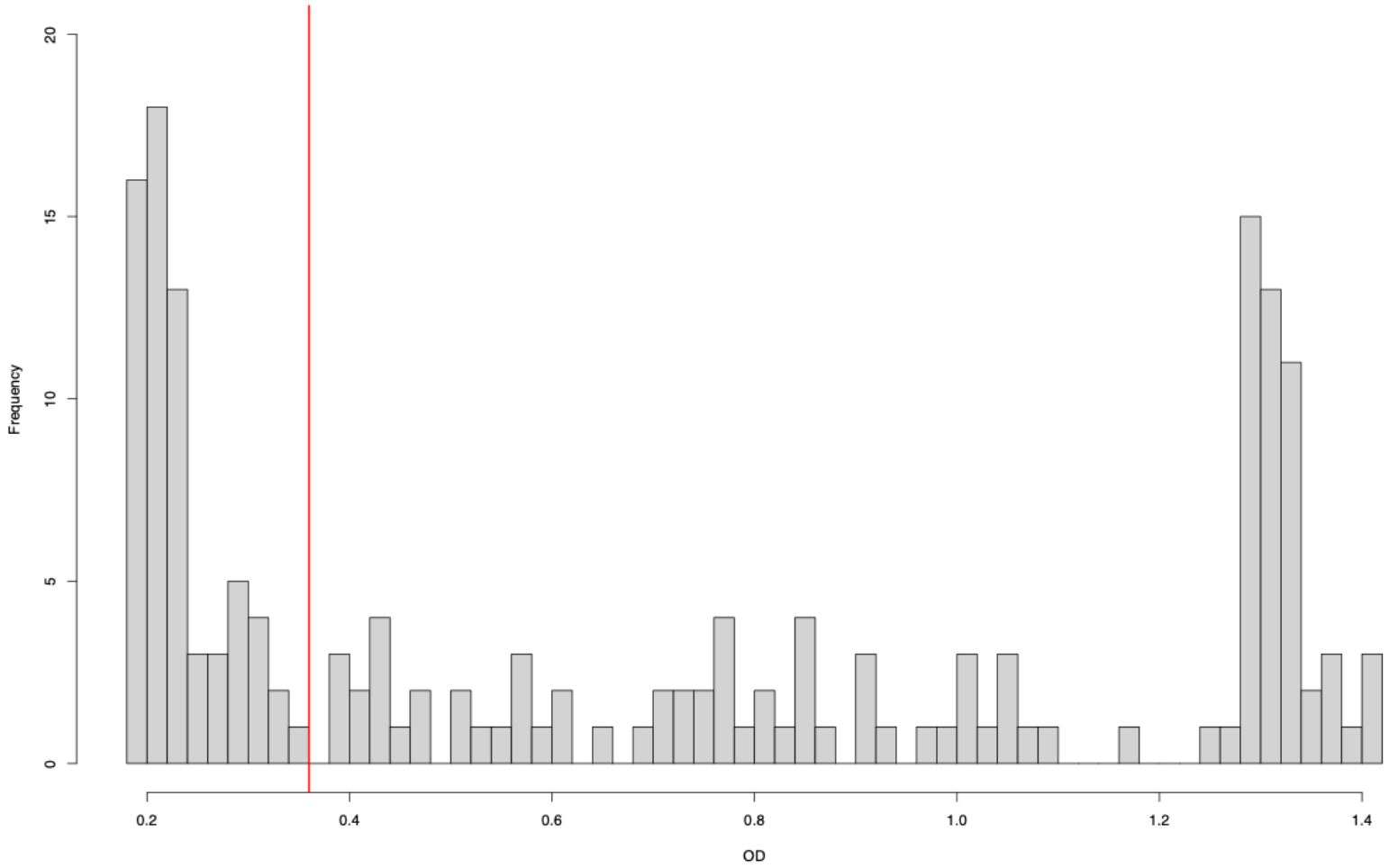


Figure 3.S3: Identifying a cut-off for strain growth. Histogram showing the frequency of maximum growth (optical density, OD) averaged across replicates for each strain. The histogram includes average maximum OD for three temperatures (30 °C, 37 °C, and 39 °C). Negative controls were excluded. The red line at 0.36 OD indicates the threshold for categorizing a strain as having grown or not grown in a temperature condition.

Figure 3.S4: Percent of strain replicates showing growth across experimental temperatures. Stacked bar plot showing the percent of replicates for each strain that met the growth threshold of 0.36 OD across each experimental growth temperature. Growth is indicated by a light blue and no growth is indicated by a salmon pink. Strain names are colored according to clade as defined in Figure 3.1a. Clades are ordered alphabetically within genetic groupings; strains are ordered alphabetically within clade.

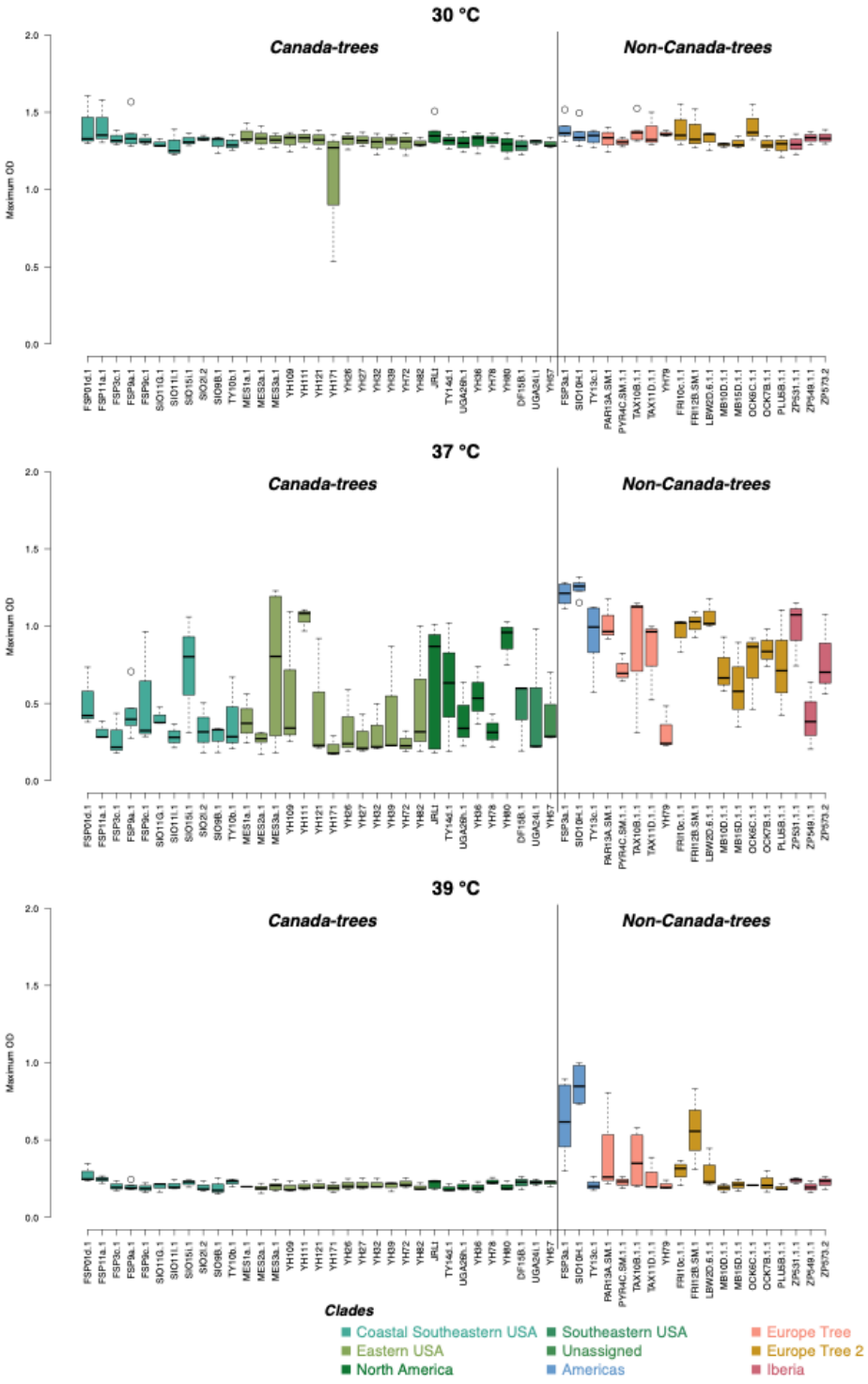


Figure 3.S5: Variation in maximum strain growth (carrying capacity) among strains and genetic lineages. Boxplots showing maximum optical density (OD) at saturation in YPD media at all experimental temperatures for all strains included in statistical analyses. The black line separates 'Canada-trees' strains from strains of other lineages. Boxplot fill corresponds with clade as defined in Figure 3.1a. Strains are ordered alphabetically by and within clade. Admixed strains were not included.

CHAPTER 4

LOW LEVEL CONTAMINATION CONFOUNDS POPULATION GENOMIC ANALYSIS[1]

[1] Ward, A.K. [†], Scopel, E.F.C. [†], Shuman, B., Momany, M., & Bensasson, D. *bioRxiv.* doi: 10.1101/2025.01.17.633387. Submitted to *BMC Genomics*, February 2025.

[†]Authors contributed equally. All authors agreed with the inclusion of this manuscript.

ABSTRACT

Genome sequence contamination has a variety of causes and can originate from within or between species. Previous research focused primarily on cross-species contamination or on prokaryotes. This paper visualizes B-allele frequency to test for intra-species contamination and measures its effects on phylogenetic and admixture analysis in two fungal species. Using a standard base calling pipeline, we found that contaminated genomes superficially appeared to produce good quality genome data. Yet as little as 5-10% genome contamination was enough to change phylogenetic tree topologies and make contaminated strains appear as hybrids between lineages (genetically admixed). We recommend the use of B-allele frequency plots to screen genome resequencing data for intra-species contamination.

KEYWORDS*:* phylogenomic, population genomics, heterozygosity, BAF plots, population structure, single nucleotide polymorphism (SNP) calls, cross-contamination

BACKGROUND

The contamination of high-throughput sequence data is a known challenge in genome biology that can lead to incorrect inferences (Goig et al., 2020; Merchant et al., 2014; Prous et al., 2020; Wilson et al., 2018). Low level sample contamination can occur in laboratories during DNA extraction or in culture, at sequencing centers during amplification steps, or even in silico if barcodes are not easily distinguished after multiplexing (Ballenghien et al., 2017; Clark et al., 2019; Cornet & Baurain, 2022; Dickins et al., 2014). Most existing tools detect contamination that occurs between species (Cornet & Baurain, 2022). Yet analysis of bacterial genomes suggests within-species contamination is more likely to lead to mistakes in base calling, species identification or phylogenetic analysis (Pightling et al., 2019). Furthermore, analysis of RNAseq

data for animal mtDNA shows that intra-species contamination can result in the overestimation of heterozygosity and incorrect inference of balancing selection (Ballenghien et al., 2017).

Most tools for detecting intra-species contamination compare read data to sequence databases for prokaryotes or particular genes or species (Cornet & Baurain, 2022). A more broadly applicable approach for the detection of within-species contamination is to identify short read data with unusual frequencies of variant alleles after mapping to a reference (Dickins et al., 2014). A similar approach, the visualization of variant (or B) allele frequencies in plots, is commonly used to determine ploidy or aneuploidy (Bensasson et al., 2019; Zhu et al., 2016). Using B-allele frequency plots, we encountered low level intra-species contamination in public data for two model fungal species: *Saccharomyces cerevisiae* and *Aspergillus fumigatus*. To determine whether low levels of intra-species contamination are cause for concern, we tested the sensitivity of a standard base calling pipeline, phylogenomic and admixture analyses to within-species contamination using read data that we contaminated in silico to known degrees (0 - 50%).

METHODS

To understand the effects of intra-species contamination on base calls and phylogenomic analysis, we created contaminated mixtures with various levels of contamination for *A. fumigatus* and *S. cerevisiae*; 0, 1, 5, 10, 20, 30, 40, and 50%. Using published short-read data (Kang et al., 2022; Scopel et al., 2021) (Table 4.S3), *S. cerevisiae* haploid, heterozygous diploid, triploid, and tetraploid genomes (CBS1479, DBVPG1074, NPA05a1, UCD 06-645) were contaminated with reads from a haploid donor (CLIB219.2b). For *A. fumigatus*, the recipient and donor were haploid strains eAF749 and eAF163 respectively. The reads for each mixture were randomly

sampled without replacement using seqtk sample (v1.2 for *S. cerevisiae*; 1.3 for *A. fumigatus*; https://github.com/lh3/seqtk). For each *S. cerevisiae* strain (12 Mbp genome), we used 8 million paired reads; 4 million for each simulated fastq file. For each *A.* fumigatus (29 Mbp genome) we used 6 million reads; 3 million for each simulated fastq file.

For base calling, we mapped reads to reference genomes using Burrows-Wheeler Aligner (bwa mem, v0.7.17; Li & Durbin, 2009). The reference genomes were Sac-Cer Apr2011/sacCer3 from strain S288c at UCSC for *S. cerevisiae* and ASM265v1 from strain Af293 for *A. fumigatus* (Nierman et al., 2005). Consensus sequences were generated using SAMtools mpileup and BCFtools call -c (v1.6; Li et al., 2009) with indels removed and read depth limited to a maximum of 100,000 reads. Mapped alignments were converted to fasta format using vcfutils.pl vcf2fq (from BCFtools) and seqtk seq with a phred- scaled quality threshold of 40 to define low quality base calls. Mitochondrial DNA was removed for downstream analyses. We generated BAF plots using vcf2alleleplot.pl with default options and counted high quality heterozygous and homozygous sites in fasta files using basecomp.pl (Bensasson, 2018).

To test the effects of contamination on phylogenetic analyses, we compared each recipient or contaminated genome to a panel of reference strains with known phylogenetic positions (Table 4.S3). For *S. cerevisiae*, we randomly selected up to 2 strains (where available) from each of the known 26 lineages described in Scopel et al. (Scopel et al., 2021), which resulted in a total of 52 reference panel strains including the donor strain. Recent genetic admixture is common in *S. cerevisiae* (Liti et al., 2009) and can complicate phylogenetic analysis, but prior analyses show that none of the strains used here were admixed (Peter et al., 2018; Scopel et al., 2021). For *A. fumigatus*, we used one-dimensional k-means clustering to

categorize 168 strains (Kang et al., 2022) into 52 clusters based on their pairwise genetic distances from a single strain (CF098), then randomly chose a single strain from each cluster. Genetic distances were estimated using the dnadist function of PHYLIP (v3.697) with default parameters and a 0.5:1 transition:transversion ratio (Felsenstein, 1993) and we used python to perform the k-means clustering (getGenDist.py; Scopel, 2024)). Neighbor-joining trees were constructed using MEGA (v10.0.5; Kumar et al., 2016) with the Tamura-Nei model (Tamura & Nei, 1993) and 100 boot-strap replicates. Maximum likelihood trees were estimated using RAxML (v8.2.11 for *S. cerevisiae* and 8.2.12 for *A. fumigatus*; Stamatakis, 2014) with a GTRGAMMA model and 100 bootstrap replicates. For visualization, trees were rooted with EN14S01, GE14S017B and HN6 for *S. cerevisiae* and JN10 for *A. fumigatus* then right ladderized using the ape package (v5.8; Paradis & Schliep, 2019) in R (v4.3.3).

For analysis of population structure and genetic admixture in *S. cerevisiae*, we used ADMIXTURE (v1.3.0; Alexander et al., 2009). Each recipient or contaminated genome was merged into an alignment with the sequence of 52 reference panel strains using BCFtools view with the –min-ac 1 option (v1.15.1). Low-quality reads (phred score under 40) were filtered in VCFtools (v0.1.16; Danecek et al., 2011). Alignment files were converted to text and binary files using PLINK (v1.9b 6.21; Purcell et al., 2007). Genomes were assigned to populations (genetic clusters) in repeated ADMIXTURE runs with default parameters and varying num- bers of populations (K); from 2 to 26 with five replicates per K. Resultant ancestry proportions were aligned across K values using CLUMPAK distruct (v1.1; Kopelman et al., 2015) and results were visualized using the R package pophelper (v2.3.1; Francis, 2017). The Fisher's exact test and other analyses and visualizations were performed in R (v4.3.3).

RESULTS AND DISCUSSION

Within-species contamination in public short read genome data

B-allele frequency plots are routinely used to distinguish homozygous or haploid genome data from heterozygous diploids or polyploids (Bensasson et al., 2019; Zhu et al., 2016). Single nucleotide polymorphisms (SNPs) in heterozygous diploids differ from the reference genome at read allele frequencies of 1.0 or 0.5, triploids at 1.0, 0.67 or 0.33 and so on (Figure 4.1). In screening short-read genome data for aneuploidy, we observed public read samples with appreciable levels of intra-species contamination (over 5%) in *S. cerevisiae* (Figure 4.S1, Table 4.S1) and *A. fumigatus* (data not shown). *Saccharomyces cerevisiae* strains are mostly homozygous diploids and *A. fumigatus* strains are usually haploid. We screened *S. cerevisiae* genome data for 1,357 strains sequenced to high read depth (over 30×) and found 8 genomes with at least 5% intra-species contamination (Peña et al., 2025). Most of these (N = 6) showed 5 - 10% contamination, and two showed 10 - 20% contamination. Higher levels of contamination would be difficult to distinguish from polyploidy using our methods, but are probably less likely.

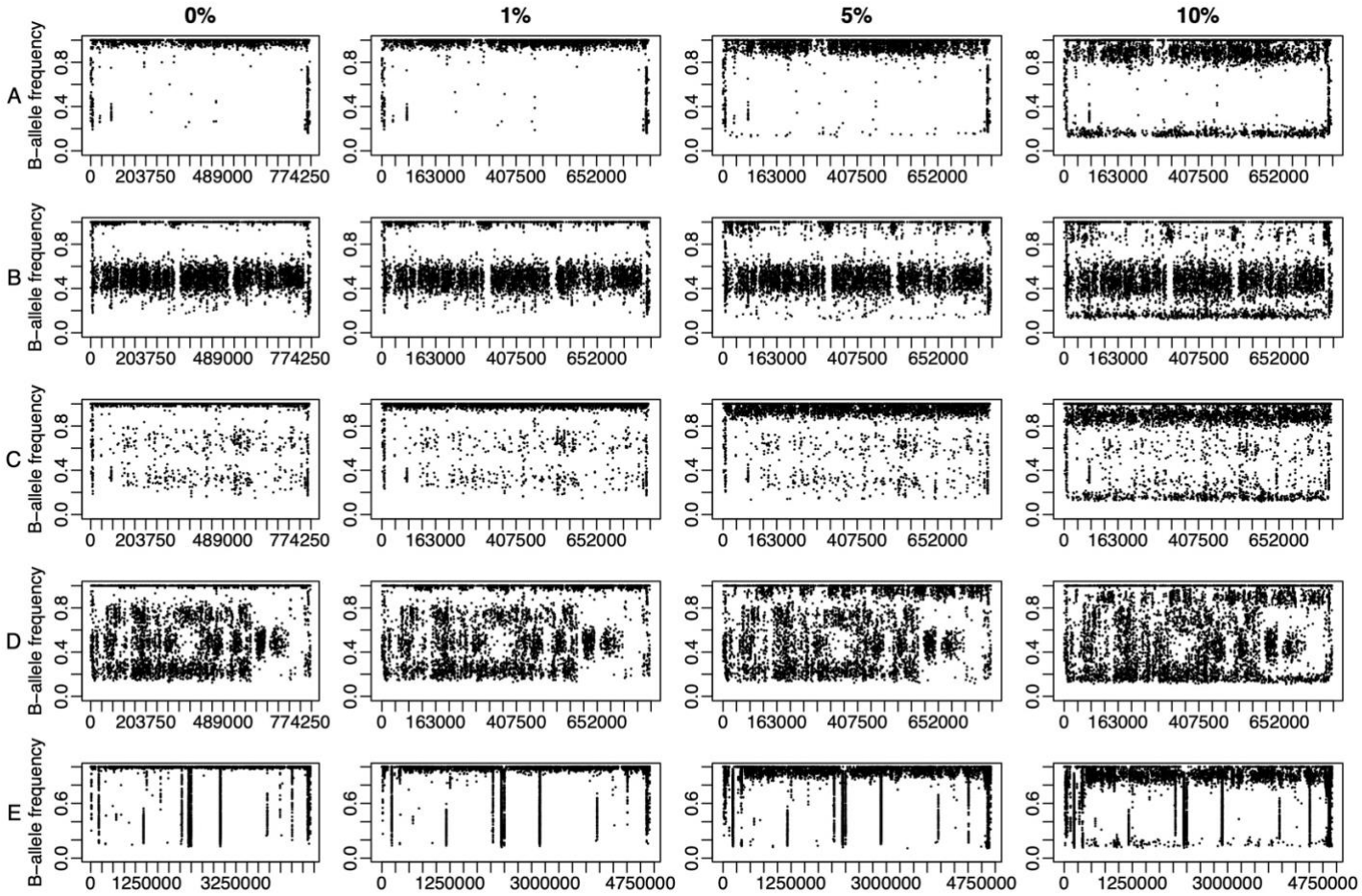


Figure 4.1: Intra-species contamination is recognizable in B-allele frequency plots at 5% contamination. Plots show base calls for resampled genome data contaminated *in silico* to: 0%, 1%, 5% and 10%. Points show the frequency of non-reference "B" alleles along chromosome II for *S. cerevisiae* (A-D) and along chromosome 1 for *A. fumigatus* (E) for A) a haploid, B) diploid, C) triploid, D) tetraploid and E) haploid. In contaminated mixtures, a substantial fraction of SNP differences from the reference genome appear below their expected frequency of 1.0, at the level expected if the contaminating strain has the same allele as the reference e.g. 0.95 for 5% contamination with a strain matching the reference. In repetitive regions variants appear at many allele frequencies appearing as vertical lines on the plots.

The extent of intra-species contamination in public genome data that we observe for *S. cerevisiae* (0.59%) is lower overall than that reported for bacteria at levels expected to affect base calling (*Escherichia coli* for 0.87%, *Salmonella enterica* 1.48%, *Listeria monocytogenes* 2.22%; Pightling et al., 2019). The percentage of *S. cerevisiae* read samples with contamination do however vary greatly by study: from under 0.2% to 15% (Table 4.S1; Fisher's exact test, $P = 2\times10-6$). This is consistent with past observations that the extent of contamination can differ substantially among studies and sequencing centers (Ballenghien et al., 2017; Goig et al., 2020).

The effects of in-silico contamination on base calling

Most contaminated data show only low levels of contamination (5 - 10%; 6 out of 8 contaminated genomes), so correct base calls outnumbered incorrect calls by ten to twenty-fold. To determine whether such low-level contamination impacts base calling, we examined in silico simulations of read data with known levels of added contamination using a standard base calling pipeline. We applied a phred-scaled quality filter (Q40) that labels sites as "low quality" data if they have estimated error rates above 1 in 10,000; a consensus base call would be represented with an "N" and therefore treated as missing data in downstream analyses. The proportion of low-quality base calls does not increase with increasing levels of contamination (Table 4.S2). The number of high quality heterozygous base calls does increase with increasing contamination, but in haploids and triploids heterozygosity only reaches the levels seen in diploids and tetraploids with 20% contamination or above. Surprisingly, even the number of high quality homozygous base calls increases slightly at 20% contamination for haploids, and with any amount of contamination at higher ploidy levels (Table 4.S2). These simple quality checks that are easily performed without population genomic analyses suggest that contamination at the

levels usually observed in public databases do not greatly affect base calling. However, these checks do not address the effects of contamination on variant or SNP sites in particular, which are likely affected differently than invariant sites and are critical for downstream applications.

Low level contamination affects population genomic analyses

Intra-species contamination likely results in erroneous heterozygous calls at SNP sites. It is therefore not surprising that contamination in past work led to mistakes in estimating the inbreeding statistic, F_{IT}, which relies on correct heterozygous base calls (Ballenghien et al., 2017). Other important population genomic statistics, such as Tajima's D and ratios of non-synonymous to synonymous diversity were less sensitive to intra-species contamination (Ballenghien et al., 2017). The inference of individual ancestry from allele frequency data is useful for estimating population structure and identifying genetic admixture (Alexander et al., 2009). It also uses heterozygous base calls and is therefore likely sensitive to contamination. Here we tested the effect of contamination at 5% and 10% contamination on the inference of individual ancestry from allele frequency data using the software ADMIXTURE (Alexander et al., 2009). In all ADMIXTURE runs, 5% contamination did not affect results (Figure 4.S2). The estimation of ancestry was affected however by 10% contamination. In most runs the contaminated strain appeared admixed between donor and recipient lineages and mostly to a greater extent (25%) than the expected 10% contamination level (Figure 4.S3).

In contrast to allele frequency analyses, we expect phylogenetic analyses to be more robust to low levels of contamination because most phylogenetic software treat heterozygous sites as missing data, and we do not expect low level contamination to result in homozygous calls for the minority allele. To test the impact of contamination on phylogenetic analysis, we

included a contaminated strain in within-species phylogenomic trees for *A. fumigatus* and *S. cerevisiae*.

Surprisingly, the phylogenetic placement of the recipient strain changed considerably even with only 10% contamination (Figure 4.2, Figures 4.S4-6). This was true for *S. cerevisiae* and *A. fumigatus* using neighbor joining distance or maximum likelihood approaches. At 10% contamination, we observed major shifts in phylogenetic position (Figures 4.2, 4.S4-6) and by 20% contamination the recipient *S. cerevisiae* strain clustered with the donor strain (Figures 4.2 and 4.S4). For A. *fumigatus* we did not include the donor strain in the phylogeny, yet we still saw major changes to tree topology (Figures 4.S5-6). Using neighbor-joining distance, we even saw a small effect on tree topology at 5% contamination in S. cerevisiae (Figure 4.2).

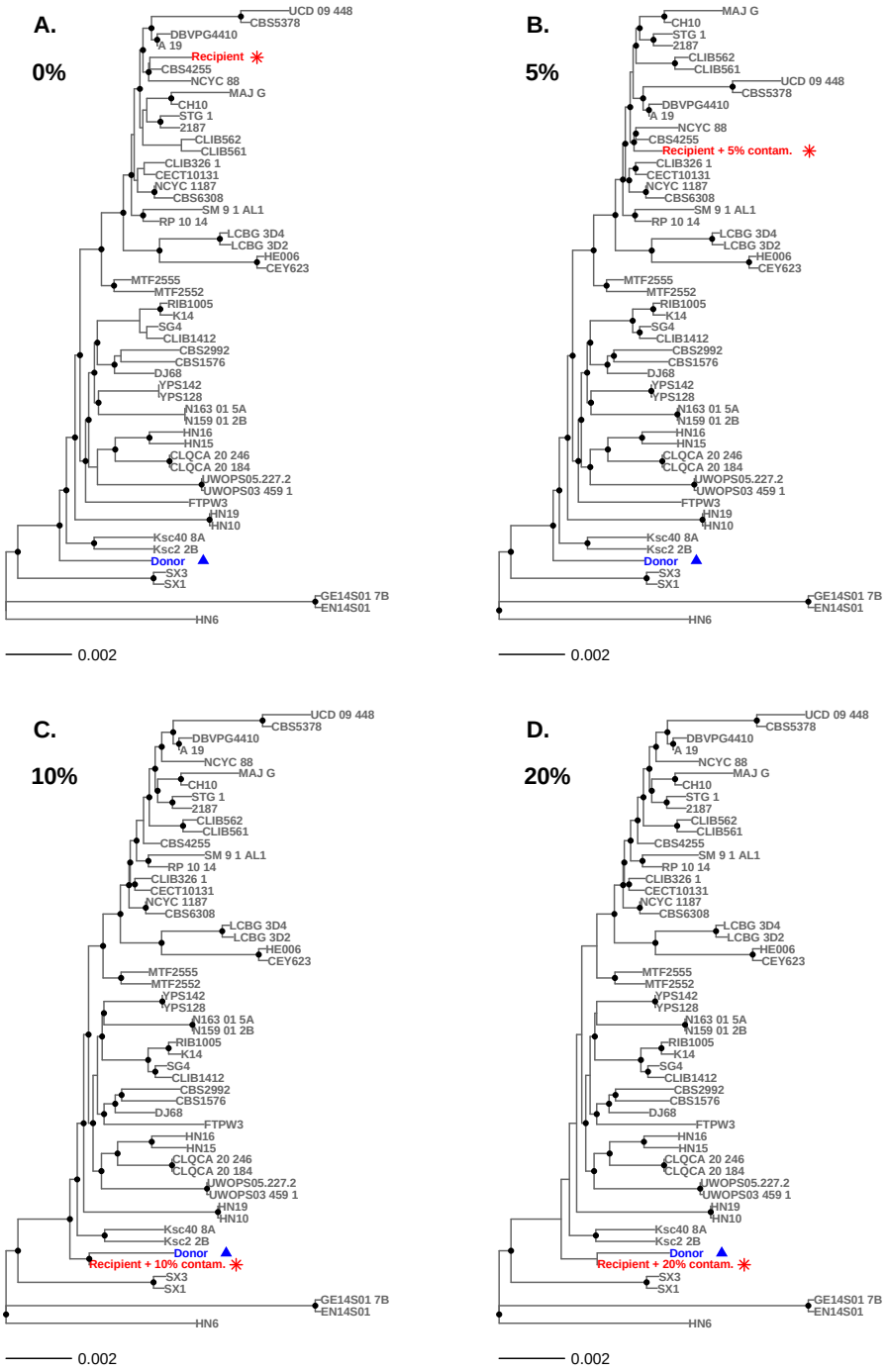


Figure 4.2: Change in topology for neighbor joining *S. cerevisiae* phylogenetic trees starting at 5% and 10% contamination. Panel A shows donor and recipient genomes in the absence of contamination; B shows the recipient with 5% contamination results in a tree with slightly altered topology; and at higher levels of contamination, 10% in C and 20% in D, the recipient clusters with the donor strain.

How could low level (5-10%) contamination alter tree topology? In contaminated data, differences between the donor and recipient sequence appear as heterozygous sites (Table 4.S2) (Ballenghien et al., 2017). These sites no longer contribute to estimates of genetic distance between the contaminated strain and donor lineage using most phylogenetic software (Lischer et al., 2014). In addition, the donor alleles will be called in regions where the recipient genome has low quality sequence or deletions relative to the reference, which could explain the increase in homozygous base calls at increasing levels of contamination (Table 4.S2). In cases where the donor genome has more high-quality regions mapping to the reference than the recipient genome (as in this study; Table 4.S2) enough homozygous base calls might result from donor reads to change tree topology. The chances of seeing an effect of contamination on cluster analyses also increase with increasing divergence between donor and recipient genomes (Pightling et al., 2019). The strains we used in this study are from genetically distinct lineages (Figures 4.2, 4.S2-6), so our analyses probably represent a worst-case scenario.

CONCLUSIONS

Here we show that within-species contamination of genome data can lead to incorrect phylogenies or inference of genetic admixture, even at the low levels seen in public databases. Contamination has led to incorrect conclusions in the past, but most reports are on between-species contamination or Sanger sequencing studies (Goig et al., 2020; Merchant et al., 2014; Prous et al., 2020; Wilson et al., 2018). Analysis of intra-species contamination in bacteria show that it can be especially damaging (Pightling et al., 2019). Using eukaryotic models, we show the importance of screening for intra-species contamination in short read genome data, especially because phylogenetic analyses can be more sensitive to contamination (5-10%, Figure 2) than

previously recognized (40-50%) (Pightling et al., 2019). The visualization of SNPs in mapped read data with B-allele frequency plots provides a means to synchronously assess ploidy, heterozygosity and potential contamination (Figure 4.1), all important information for downstream phylogenetic or population genomic analyses.

ACKNOWLEDGEMENTS

We would like to thank Jacqueline Peña for help with data management ahead of publication, and Momany and Bensasson lab members for helpful discussion. This work was funded by a National Science Foundation grant (IOS, no. 1946046) awarded to DB.

SUPPLEMENTARY TABLE INFORMATION.

Due to size, the following supplemental tables for this chapter are available on GitHub here: https://github.com/bensassonlab/data/tree/master/Ward2025_Dissertation

Table 4.S3. Summary of the strains used for *in silico* contamination and population genomic analyses.

AUTHORS' CONTRIBUTIONS

All authors conceived and designed the experiment. EFCS screened for contamination in public data, developed the *in-silico* sampling approach and performed base call analyses; AKW performed the phylogenetic and admixture analyses; BS performed all analyses for *A. fumigatus*. All authors wrote the manuscript.

REFERENCES

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

Ballenghien, M., Faivre, N., & Galtier, N. (2017). Patterns of cross-contamination in a multispecies population genomic project: Detection, quantification, impact, and solutions. *BMC Biology*, *15*(1), 25. https://doi.org/10.1186/s12915-017-0366-6

Bensasson, D. (2018). *bensassonlab/scripts: Code for yeast population genomic analyses* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.1488147

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

Clark, M. A., Stankiewicz, S. H., Barronette, V., & Ricke, D. O. (2019). Letter to the Editor – Detecting HTS Barcode Contamination. *Journal of Forensic Sciences*, *64*(3), 961–962. https://doi.org/10.1111/1556-4029.14027

Cornet, L., & Baurain, D. (2022). Contamination detection in genomic data: More is not enough. *Genome Biology*, *23*(1), 60. https://doi.org/10.1186/s13059-022-02619-9

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

Dickins, B., Rebolledo-Jaramillo, B., Su, M. S.-W., Paul, I. M., Blankenberg, D., Stoler, N., Makova, K. D., & Nekrutenko, A. (2014). Controlling for Contamination in Re-

Sequencing Studies With a Reproducible Web-Based Phylogenetic Approach. *BioTechniques*, *56*(3), 134–141. https://doi.org/10.2144/000114146

Felsenstein, J. (1993). *PHYLIP (Phylogeny Inference Package)* (Version 3.697) [Computer software].

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

Goig, G. A., Blanco, S., Garcia-Basteiro, A. L., & Comas, I. (2020). Contaminant DNA in bacterial sequencing experiments is a major source of false genetic variability. *BMC Biology*, *18*(1). https://doi.org/10.1186/s12915-020-0748-z

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*, *12*(2), jkab427. https://doi.org/10.1093/g3journal/jkab427

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

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

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., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lischer, H. E. L., Excoffier, L., & Heckel, G. (2014). Ignoring Heterozygous Sites Biases Phylogenomic Estimates of Divergence Times: Implications for the Evolutionary History of Microtus Voles. *Molecular Biology and Evolution*, *31*(4), 817–831. https://doi.org/10.1093/molbev/mst271

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, 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

Merchant, S., Wood, D. E., & Salzberg, S. L. (2014). Unexpected cross-species contamination in genome sequencing projects. *PeerJ*, *2*, e675. https://doi.org/10.7717/peerj.675

Nierman, W. C., Pain, A., Anderson, M. J., Wortman, J. R., Kim, H. S., Arroyo, J., Berriman, M., Abe, K., Archer, D. B., Bermejo, C., Bennett, J., Bowyer, P., Chen, D., Collins, M., Coulsen, R., Davies, R., Dyer, P. S., Farman, M., Fedorova, N., … Denning, D. W. (2005). Genomic sequence of the pathogenic and allergenic filamentous fungus

Aspergillus fumigatus. *Nature*, *438*(7071), 1151–1156. https://doi.org/10.1038/nature04332

Paradis, E., & Schliep, K. (2019). 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

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. https://doi.org/10.1111/mec.17669

Peter, J., De Chiara, M., 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), 339–344. https://doi.org/10.1038/s41586-018-0030-5

Pightling, A. W., Pettengill, J. B., Wang, Y., Rand, H., & Strain, E. (2019). Within-species contamination of bacterial whole-genome sequence data has a greater influence on clustering analyses than between-species contamination. *Genome Biology*, *20*(1), 286. https://doi.org/10.1186/s13059-019-1914-x

Prous, M., Lee, K. M., & Mutanen, M. (2020). Cross-contamination and strong mitonuclear discordance in Empria sawflies (Hymenoptera, Tenthredinidae) in the light of phylogenomic data. *Molecular Phylogenetics and Evolution*, *143*, 106670. https://doi.org/10.1016/j.ympev.2019.106670

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., 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

Scopel, E. F. C. (2024). *Eduardoscopel/getGenDist* (Version 1.2) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.13881957

Scopel, E. F. C., Hose, J., Bensasson, D., & Gasch, A. P. (2021). Genetic variation in aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages. *Genetics*, *217*(4), iyab015. https://doi.org/10.1093/genetics/iyab015

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

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*. https://doi.org/10.1093/oxfordjournals.molbev.a040023

Wilson, C. G., Nowell, R. W., & Barraclough, T. G. (2018). Cross-Contamination Explains "Inter and Intraspecific Horizontal Genetic Transfers" between Asexual Bdelloid Rotifers. *Current Biology*, *28*(15), 2436-2444.e14. https://doi.org/10.1016/j.cub.2018.05.070

Zhu, Y. O., Sherlock, G., & Petrov, D. A. (2016). Whole Genome Analysis of 132 Clinical *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation. *G3: Genes|Genomes|Genetics*, *6*(8), 2421–2434. https://doi.org/10.1534/g3.116.029397

Table 4.S1: Differences among studies in rates of intra-species *Saccharomyces cerevisiae* contamination. Data from Peña *et al* (2025) after excluding genomes with low read depth and studies with fewer than 10 high depth genomes. Rates appear different (Fisher's exact test, $P = 2 \times 10^{-6}$) even after excluding PRJEB11698 (Fisher's exact test, $P = 0.002$).

SRA identifier	Uncontaminated	Contaminated	%	Study
ERP014555	915	0	0%	Peter *et al*, 2018
PRJNA396809	260	5	2%	Duan *et al*, 2018
PRJEB7601	55	0	0%	Almeida *et al*, 2015
PRJNA1090965	43	0	0%	Peña *et al*, 2025
PRJEB11698	17	3	15%	Barbosa *et al*, 2016

Table 4.S2: Intra-species contamination does not lower the quality of base calls.

% from donor.[1]	Ploidy	Homozygous	Heterozygous	Low Quality	P(LQ)[2]
A. fumigatus					
100%	haploid	28,056,212	27,124	1,299,087	0.044
0%	haploid	27,191,111	26,332	2,166,779	0.074
1%		27,194,419	26,678	2,163,182	0.074
5%		27,303,362	30,663	2,050,218	0.070
10%		27,616,509	44,955	1,722,747	0.059
20%		28,018,909	75,478	1,289,836	0.044
30%		28,104,794	83,560	1,195,866	0.041
S. cerevisiae					
100%	haploid	11,461,857	3,697	604,481	0.050
0%	haploid	11,430,299	2,706	638,192	0.053
1%		11,432,806	2,784	635,631	0.053
5%		11,448,751	4,868	617,607	0.051
10%		11,456,000	24,042	591,174	0.049
20%		11,464,068	72,671	534,541	0.044
30%		11,471,388	80,901	519,008	0.043
40%		11,473,538	81,923	515,830	0.043
50%		11,474,291	82,281	514,720	0.043
0%	diploid	11,426,767	62,889	581,602	0.048
1%		11,427,772	62,895	580,591	0.048
5%		11,431,346	64,035	575,878	0.048
10%		11,430,347	76,131	564,783	0.047
20%		11,435,185	104,413	531,692	0.044
30%		11,439,882	108,814	522,575	0.043
40%		11,441,299	108,547	521,427	0.043
50%		11,441,050	105,473	524,656	0.043
0%	triploid	11,421,825	8,610	637,335	0.053
1%		11,423,877	8,665	635,663	0.053
5%		11,425,285	10,582	631,913	0.052
10%		11,415,094	28,950	624,236	0.052
20%		11,413,683	75,713	579,163	0.048
30%		11,419,975	83,775	564,165	0.047
40%		11,423,533	84,218	559,800	0.046
50%		11,423,917	83,808	560,476	0.046
0%	tetraploid	11,347,790	46,526	676,765	0.056
1%		11,349,837	46,590	674,801	0.056
5%		11,376,595	47,656	646,986	0.054
10%		11,414,485	59,758	596,994	0.049
20%		11,439,265	96,163	535,856	0.044
30%		11,446,409	101,228	523,665	0.043
40%		11,449,966	98,988	522,346	0.043
50%		11,452,171	95,042	524,088	0.043

[1] Percent of reads from the donor (contaminant) genome. The recipient genome has 0% of reads from donor. For *Aspergillus fumigatus*, the donor strain was eAF163 and the recipient was eAF749. For *Saccharomyces cerevisiae*, the donor strain was CLIB219.2b and haploid, heterozygous diploid, triploid, and tetraploid recipient strains were CBS1479, DBVPG1074, NPA05a1, UCD 06-645 respectively. Donor and recipient genomes were resampled to the same read depth as contaminated genomes.

[2] Proportion of base calls that are low quality (phred-scaled quality lower than Q40).

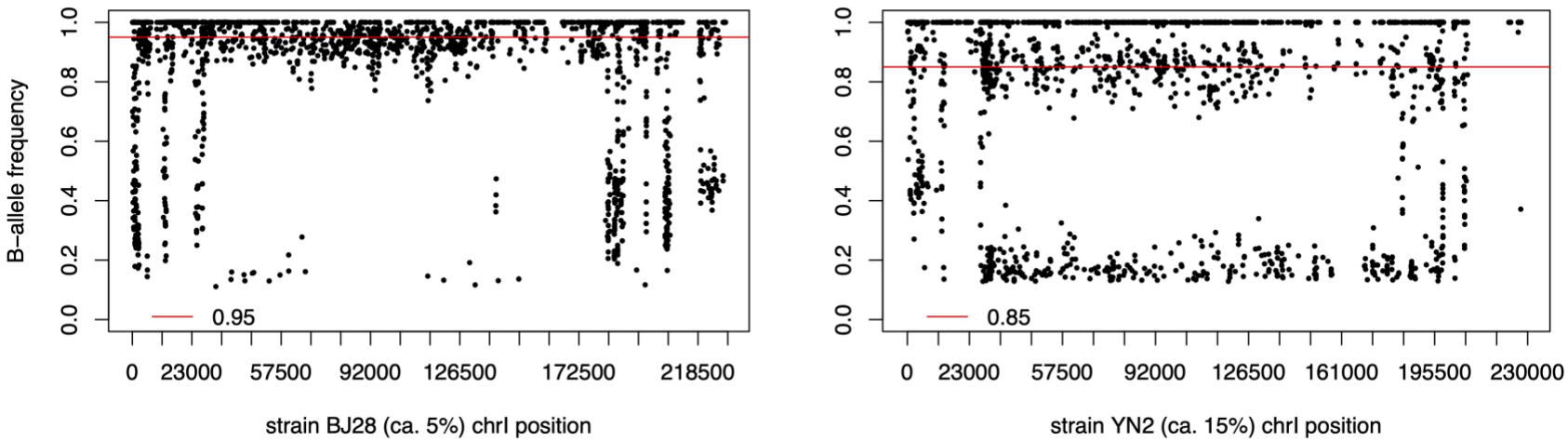


Figure 4.S1: B-allele frequency plots show intraspecies contamination in public *S. cerevisiae* genome data. In the left panel, the points clustering around the red line at 0.95 suggest contamination levels of 5% where the sequenced strain (BJ28) has SNPs that differ from the reference strain, but the contaminating strain allele matches the reference. The right panel shows B-allele frequencies for strain YN2 suggesting 15% contamination.

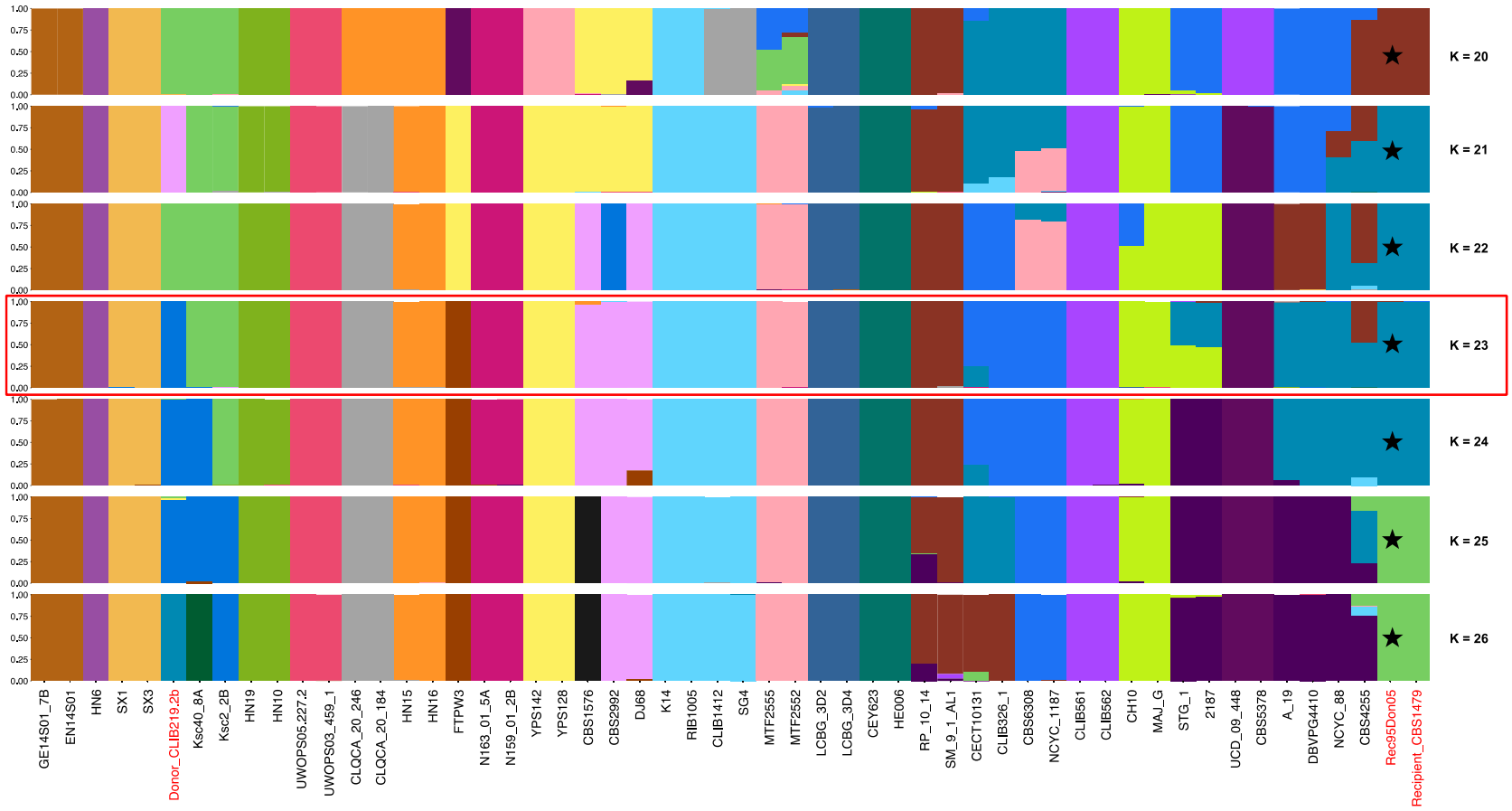


Figure 4.S2: No effect of 5% contamination on analysis of genetic admixture. The plot shows the ancestry proportions for each individual arranged in the order seen in Figure 4.S5. These plots show the runs with the highest log-likelihoods for each assumed number of populations (K = 20 - 26), and the run with the most clustering similarity to the phylogenetic analyses (K = 23) is highlighted with a red box. Individual genomes highlighted with red text are the donor genome (Donor CLIB10 2B), and the recipient genome showing the ancestry proportions expected with 0% contamination (Recipient CBS1479) and an *in-silico* mix of recipient with 5% contamination from the donor (Rec95Don05, black star) showing the same results as the uncontaminated recipient.

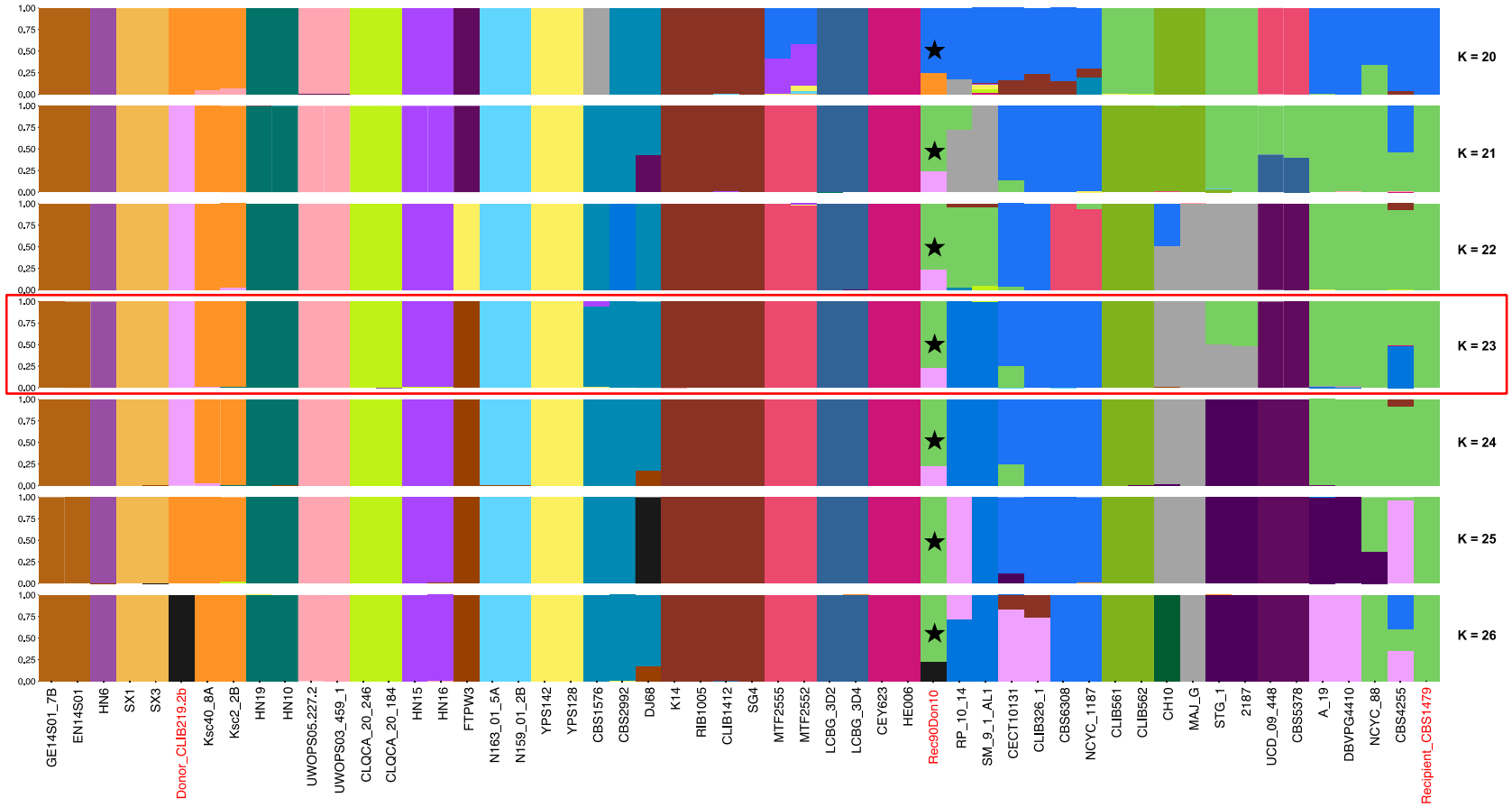


Figure 4.S3: Contamination levels of 10% result in incorrect calls of genetic admixture. The plot shows the ancestry proportions for each individual arranged in the order seen in maximum likelihood phylogenetic analyses Figure 4.S5. These plots show the runs with the highest log-likelihoods for each assumed number of populations (K = 20 - 26), and the run with the most clustering similarity to the phylogenetic analyses (K = 23) is highlighted with a red box. Individual genomes highlighted with red text are the donor genome (Donor CLIB10 2B), and the recipient genome showing the ancestry proportions expected with 0% contamination (Recipient CBS1479) and an *in-silico* mix of recipient with 10% contamination from the donor (Rec90Don10, black star). In most runs, the genome with 10% contamination shows 25% admixture between donor and recipient lineages.

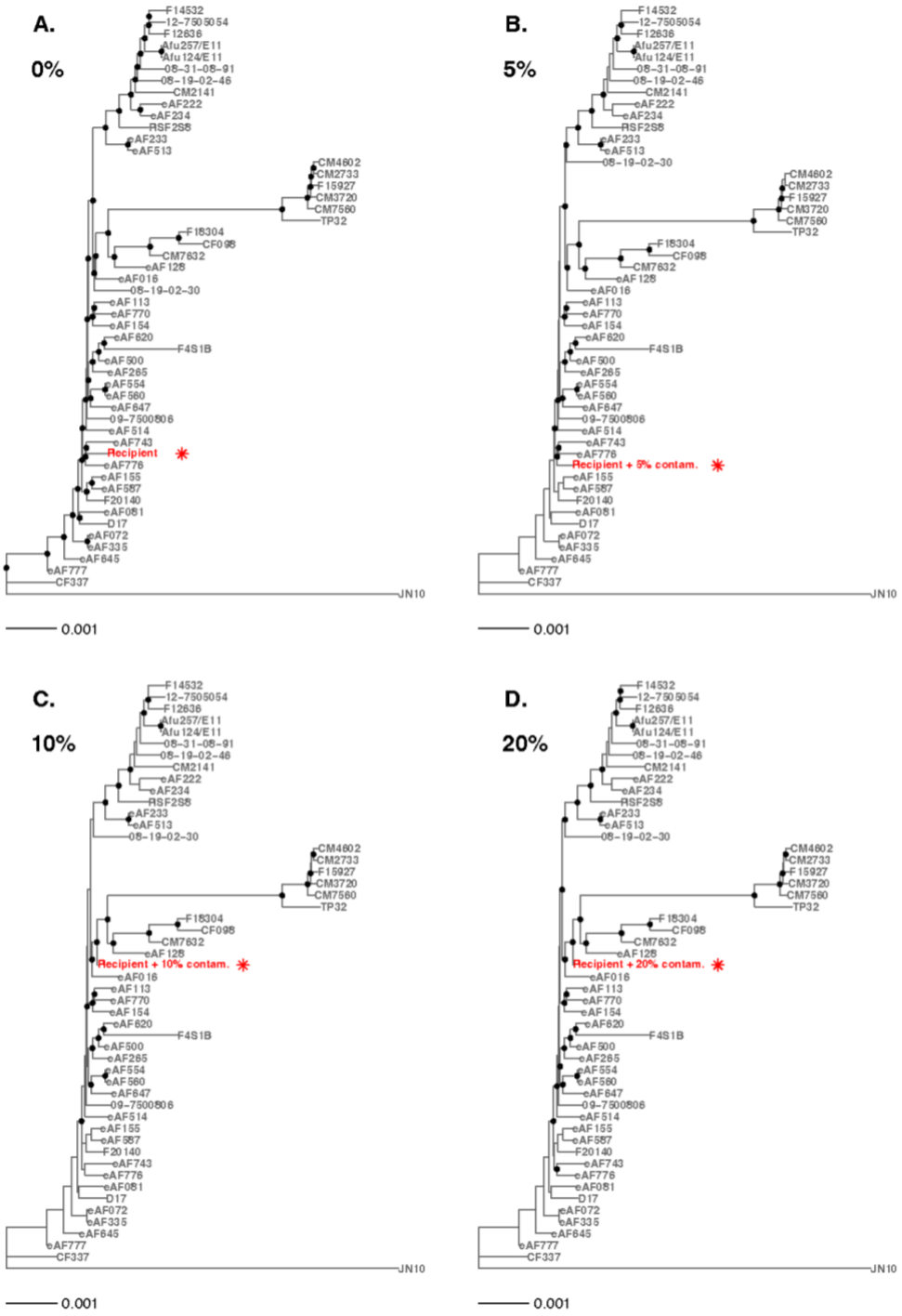


Figure 4.S4: Change in topology for neighbor joining *A. fumigatus* phylogenetic trees with 10% cross-contamination. Panel A shows the recipient genome in the absence of cross-contamination; B shows the recipient with 5% contamination; C shows a new topology at 10% contamination and D shows the same new topology at 20%.

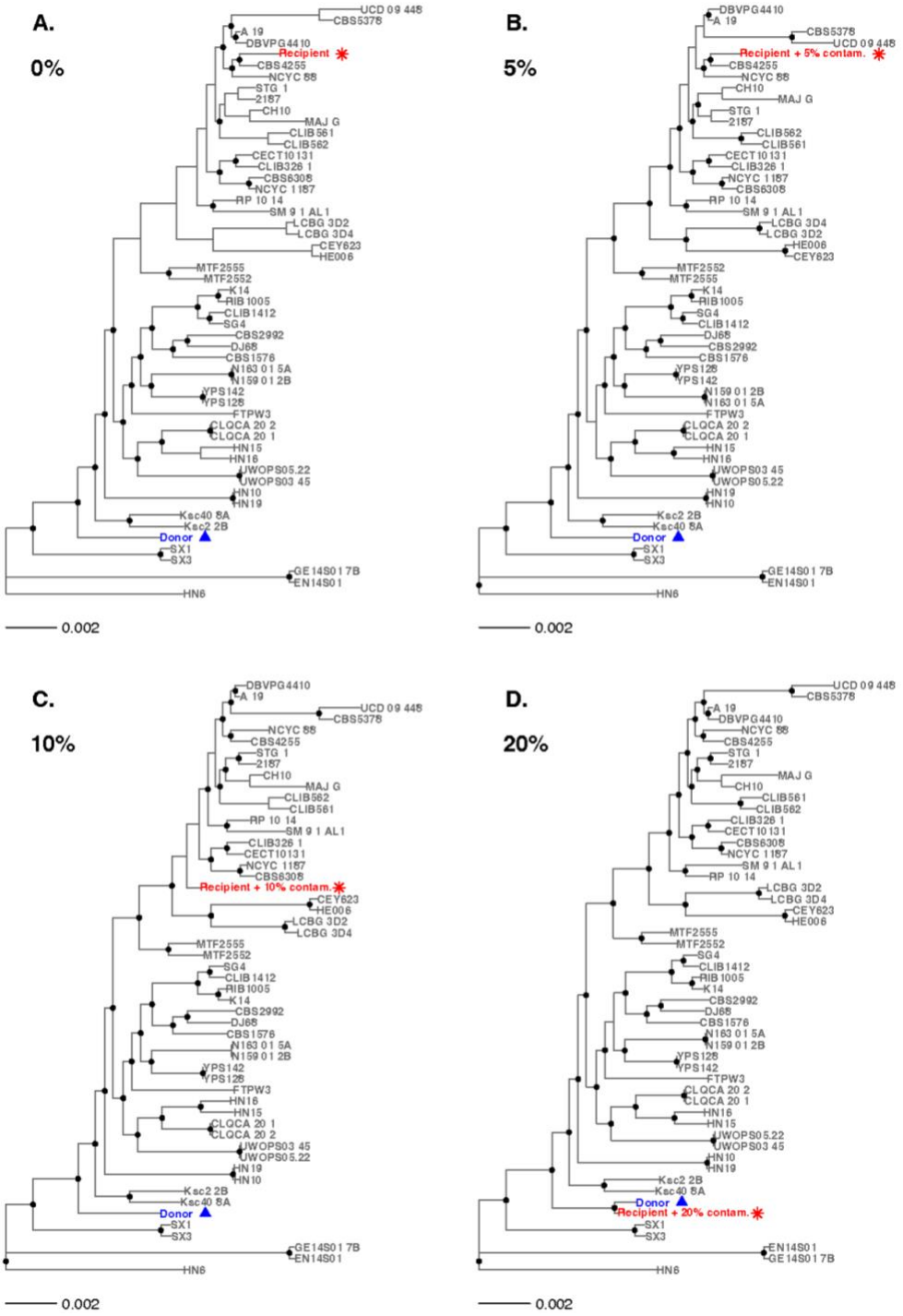


Figure 4.S5: Change in topology for maximum likelihood *S. cerevisiae* phylogenetic trees with 10% cross-contamination. Panel A shows donor and recipient genomes in the absence of cross-contamination; B shows the recipient with 5%; C shows a new topology at 10%; D shows a more greatly altered topology at 20% contamination.

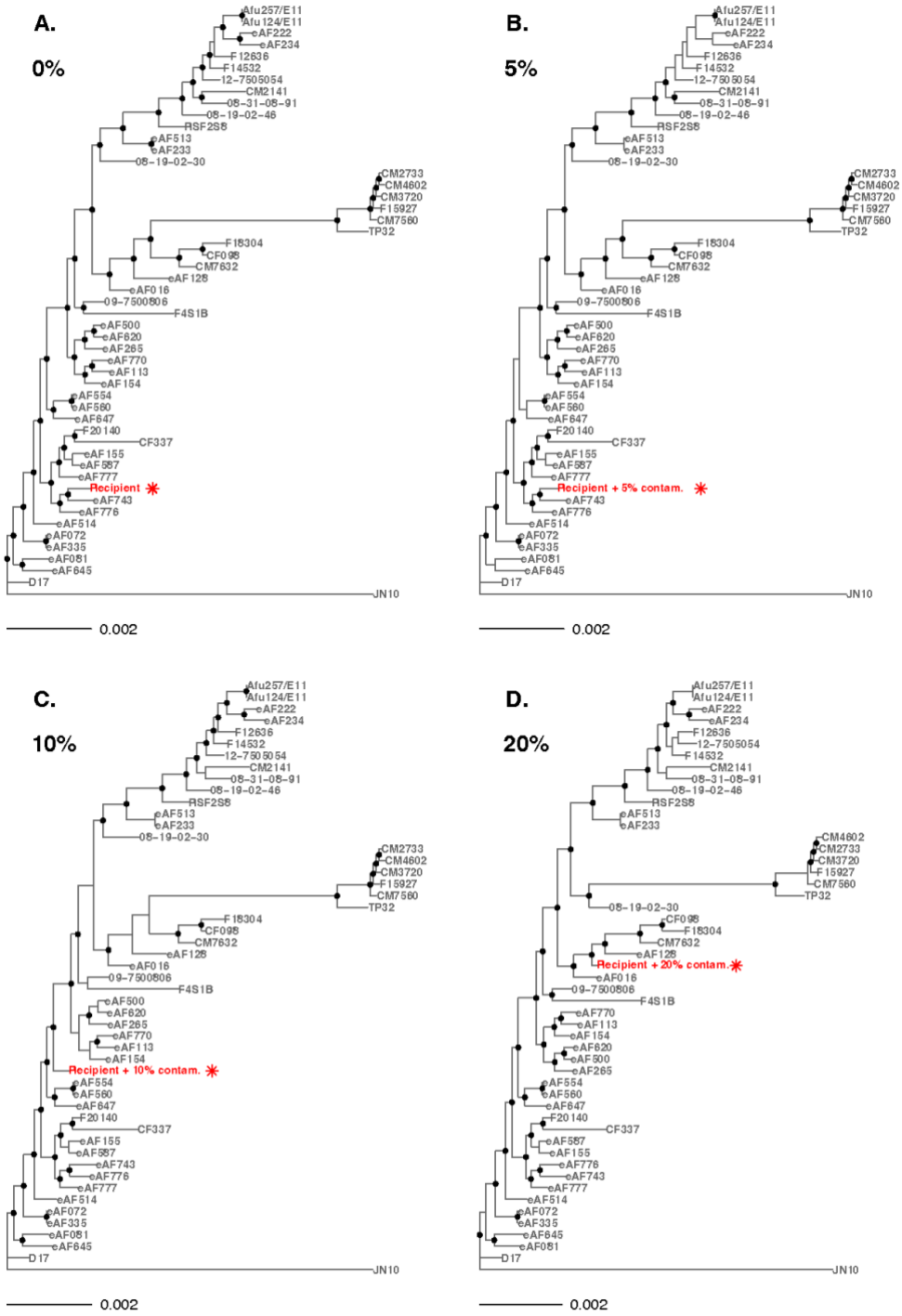


Figure 4.S6: Change in topology for maximum likelihood *A. fumigatus* phylogenetic trees with 10% cross-contamination. Panel A shows the recipient genome in the absence of cross-contamination; B shows the recipient with 5%; C shows a new topology at 10%; D shows a more greatly altered topology at 20% contamination.

SUPPLEMENTAL 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.: A population genomics insight into the Mediterranean origins of wine yeast domestication. Molecular Ecology 24(21), 5412–5427 (2015) 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.: Evidence of Natural Hybridization in Brazilian Wild Lineages of *Saccharomyces cerevisiae*. Genome Biology and Evolution 8(2), 317–329 (2016) https://doi.org/10.1093/gbe/evv263

Duan, S.-F., Han, P.-J., Wang, Q.-M., Liu, W.-Q., Shi, J.-Y., Li, K., Zhang, X.-L., Bai, F.-Y.: The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. Nature Communications 9(1), 2690 (2018) https://doi.org/10.1038/s41467-018-05106-7

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. https://doi.org/10.1111/mec.17669

Peter, J., De Chiara, M., 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), 339–344. https://doi.org/10.1038/s41586-018-0030-5

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

The population genomics and environmental niches of model species *Saccharomyces cerevisiae* and its sister species *S. paradoxus* have been well characterized (Bai et al., 2022; He et al., 2022; Leducq et al., 2014; Peña et al., 2025; Robinson et al., 2016; Xia et al., 2017); these species provide an excellent model system in which to study contrasting evolutionary dynamics within the same genus (Yue et al., 2017). Expansion of research into the clinical, ecological, and industrial roles of and whole genome resources for non-model yeasts provides new avenues to address questions about adaptive processes within a broader context (Peter & Schacherer, 2016).

Lachancea thermotolerans is a rising model for population genomics due to its wide distribution across geographic regions and substrates and distinct population structure (Banilas et al., 2016; Freel et al., 2014; Hranilovic et al., 2017; Porter et al., 2019), co-occurrence with both *S. cerevisiae* and *S. paradoxus* in woodland environments (Robinson et al., 2016; Spurley et al., 2022), and its frequent use alongside *S. cerevisiae* in fermentative processes (Benito, 2018; Binati et al., 2020; Gobbi et al., 2013; Postigo et al., 2023). In addition to this overlap in the fermentation industry, *L. thermotolerans* and *S. cerevisiae* both have domesticated lineages associated with winemaking in the Mediterranean Basin (Almeida et al., 2015; Banilas et al., 2016; Hranilovic et al., 2018). The large genetic distance between *Saccharomyces* and *Lachancea* (~125-150 mya; Shen et al., 2018) and the genetic and ecological overlaps between these species make them a compelling model system for investigating parallel and convergent evolution. In this dissertation, I utilize a combination of phylogenetic and genomic approaches to

understand phylogeography, population structure, and adaptation in *L. thermotolerans* to further develop the species as a model system.

In Chapter 2, I examine the phylogeographic distribution and population structure of *Lachancea thermotolerans*. Previous analyses have characterized lineage divergence according to geography and ecological origin; however, these studies primarily use strains isolated from domestic European populations (Freel et al., 2014; Hranilovic et al., 2017; Vicente et al., 2025). To better understand the breadth of population structure across habitats, I incorporated additional publicly available *L. thermotolerans* whole-genome sequences from woodland environments (Hranilovic et al., 2017; Xia et al., 2017) and expanded whole-genome resources for wild forest isolates across the United States and Europe (Osburn et al., 2018; Robinson et al., 2016).

Analyses of these sequences expanded the number of known genetic lineages within the *L. thermotolerans* species tree from 6 to 12: two distinct lineages from northern and western or southern European forests, three distinct lineages within the Iberian Peninsula, and four wild lineages separated by geography in North America. Highly structured woodland lineages have also been seen in *S. cerevisiae* and *S. paradoxus* (Leducq et al., 2014; Peña et al., 2025), which suggests that this may be a consistent pattern across *Saccharomycotina*. The genetic distance among *L. thermotolerans* lineages appears comparable to those found in *S. paradoxus*, which has shown subspecies formation and reproductive isolation between the most distantly related lineages (Delneri et al., 2003). This presents interesting directions for further analyses looking into the diversity within *L. thermotolerans*. Additionally, I found evidence for gene flow between lineages within continents and between the Americas and Europe. Future work quantifying potential migration events using time divergence analysis incorporating previous dating estimates (Shen et al., 2018) or lab estimates of mutation rate in *S. cerevisiae* or *S. paradoxus*

(Kaya et al., 2021; Zhu et al., 2014) may allow better understanding of the environmental factors that have shaped the distribution and genetic diversity within the species. Receding glaciers after the Last Glacial Maximum and human activity impacted *Lachancea cidri* and *S. cerevisiae* (Peña et al., 2025; Villarreal et al., 2022); time divergence analysis would allow us to see if these events may have also shaped the phylogeography of *L. thermotolerans*. The copy numbers of genes connected with the domestication suite were previously used to compare lineages within *L. thermotolerans* (Vicente et al., 2025). In contrast, I found increased copy numbers of these genes in wild lineages and was unable to recapitulate the finding of high copy numbers in domestic lineages. Additional research using the same approach of Vicente *et al*. (2025), combined measures of read depth and BLAST matches in *de novo* assembly, should clarify whether our contrasting findings result from the needed addition of wild strains or from methodological differences. This would improve understanding of the potential influence of the winemaking environment on genomic content across the *L. thermotolerans* species tree.

Phenotypic variation within a species can be attributed to genetic differences. In Chapter 3, I assess variation in growth rate across different thermal conditions to determine whether *L. thermotolerans* strains could be locally adapted to temperature at their geographic origin. Previous work in natural forest populations of *S. paradoxus* have shown local climatic adaptation (Leducq et al., 2014), and past experimental evolution studies in laboratory *S. cerevisiae* suggest that adaptation is rapid and occurs frequently in microbes (Fay et al., 2023; Lang et al., 2011; Wang et al., 2024). I sought to determine if this non-*Saccharomyces* yeast shows evidence for local climatic adaptation to elevated temperature by assaying fifty-two strains of *L. thermotolerans* isolated from a range of local temperatures and precipitation levels for growth rate across a range of experimental temperatures (25 – 42 °C). While strains from one genetic

lineage grew more slowly than their conspecifics in other lineages and grew less overall at high temperatures, I found no evidence that thermal growth rate differences were driven by factors beyond genetic background.

Although there was no evidence of local precipitation amount or temperature impacting the growth speed of *L. thermotolerans*, other environmental factors may be shaping lineage response to temperature changes. In *L. thermotolerans*, lactate production during fermentation is accompanied by decreased growth rate and increased expression of genes related to stress response (Battjes et al., 2023). Laboratory experiments have shown that increased stress tolerance, such as withstanding nutrient depletion and survival in freeze-thaw cycles, is associated with a lower growth rate in *S. cerevisiae* (Zakrzewska et al., 2011). This suggests the possibility of a trade-off between stress tolerance and growth rate in *L. thermotolerans*, which can be explored in future work. Another area for future work would be to explore competition between species, which incentivizes niche separation (Wadgymar et al., 2022) and may become more frequent as climate change alters natural species distributions (Cuartero et al., 2024; Leducq et al., 2014). As *L. thermotolerans* co-occurs with many different microbial species, including *Saccharomyces* yeasts, assays determining how well *L. thermotolerans* competes with other species for resources and if any other climatic variables may shape niche separation in *L. thermotolerans* would be useful experiments to increase our understanding of environmental adaptation.

Contamination of sequence data can lead to incorrect inferences about species history and population dynamics and can occur at any point from DNA extraction in the laboratory to *in silico* experiments and analyses (Ballenghien et al., 2017; Clark et al., 2019). Chapter 4 discusses a collaborative project demonstrating the potential impact of intra-species contamination on

genome data and downstream genomic analyses. The majority of tools focus on between-species contamination; those that do detect intra-species contamination compare data against sequence databases (Cornet & Baurain, 2022). Using B-allele frequency plots (Bensasson et al., 2019; Zhu et al., 2016), we found low-level intra-species contamination in public data for the fungal models *S. cerevisiae* and *Aspergillus fumigatus*. Base calling analyses for *in silico* contaminated strains across various levels of contamination (0 - 50%) for both species showed that contamination at levels normally seen in public data (1-10%) does not have an obvious impact on base calling. Incorporating sequences from known lineages of *A. fumigatus* and *S. cerevisiae* (Kang et al., 2022; Scopel et al., 2021) for population genomic analyses with these contaminated strains told a different story; strains that were contaminated at levels as low as 10% resulted in major changes to (i) allele frequency analysis results and (ii) tree topology in both maximum-likelihood and neighbor-joining methods. These results emphasize the importance of quality checks in mapped read data, especially via visualization of single nucleotide polymorphisms, to identify potential contamination prior to use in population genomic analyses.

Overall, this dissertation provides an investigation into the population history, genetic architecture, and phenotypic profiles of an industrially and ecologically relevant yeast using a variety of analyses and approaches. More work is needed to develop this species as a model system that could be used in conjunction with other yeast species to understand repeated evolution. My dissertation highlights how expanding genomic resources of non-*Saccharomyces* yeasts can provide additional insight into evolutionary and population dynamics across eukaryotes.

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

Bai, F.-Y., Han, D.-Y., Duan, S.-F., & Wang, Q.-M. (2022). The Ecology and Evolution of the Baker's Yeast *Saccharomyces cerevisiae*. *Genes*, *13*(2), 230. https://doi.org/10.3390/genes13020230

Ballenghien, M., Faivre, N., & Galtier, N. (2017). Patterns of cross-contamination in a multispecies population genomic project: Detection, quantification, impact, and solutions. *BMC Biology*, *15*(1), 25. https://doi.org/10.1186/s12915-017-0366-6

Banilas, G., Sgouros, G., & Nisiotou, A. (2016). Development of microsatellite markers for *Lachancea thermotolerans* typing and population structure of wine-associated isolates. *Microbiological Research*, *193*, 1–10. https://doi.org/10.1016/j.micres.2016.08.010

Battjes, J., Melkonian, C., Mendoza, S. N., Haver, A., Al-Nakeeb, K., Koza, A., Schrubbers, L., Wagner, M., Zeidan, A. A., Molenaar, D., & Teusink, B. (2023). Ethanol-lactate transition of *Lachancea thermotolerans* is linked to nitrogen metabolism. *Food Microbiology*, *110*, 104167. https://doi.org/10.1016/j.fm.2022.104167

Benito, S. (2018). The impacts of *Lachancea thermotolerans* yeast strains on winemaking. *Applied Microbiology and Biotechnology*, *102*(16), 6775–6790. https://doi.org/10.1007/s00253-018-9117-z

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

Binati, R. L., Lemos Junior, W. J. F., Luzzini, G., Slaghenaufi, D., Ugliano, M., & Torriani, S. (2020). Contribution of non-*Saccharomyces* yeasts to wine volatile and sensory diversity: A study on *Lachancea thermotolerans*, *Metschnikowia spp.* and *Starmerella bacillaris* strains isolated in Italy. *International Journal of Food Microbiology*, *318*, 108470. https://doi.org/10.1016/j.ijfoodmicro.2019.108470

Clark, M. A., Stankiewicz, S. H., Barronette, V., & Ricke, D. O. (2019). Letter to the Editor – Detecting HTS Barcode Contamination. *Journal of Forensic Sciences*, *64*(3), 961–962. https://doi.org/10.1111/1556-4029.14027

Cornet, L., & Baurain, D. (2022). Contamination detection in genomic data: More is not enough. *Genome Biology*, *23*(1), 60. https://doi.org/10.1186/s13059-022-02619-9

Cuartero, J., Querejeta, J. I., Prieto, I., Frey, B., & Alguacil, M. M. (2024). Warming and rainfall reduction alter soil microbial diversity and co-occurrence networks and enhance pathogenic fungi in dryland soils. *Science of The Total Environment*, *949*, 175006. https://doi.org/10.1016/j.scitotenv.2024.175006

Delneri, D., Colson, I., Grammenoudi, S., Roberts, I. N., Louis, E. J., & Oliver, S. G. (2003). Engineering evolution to study speciation in yeasts. *Nature*, *422*(6927), 68–72. https://doi.org/10.1038/nature01418

Fay, J. C., Alonso-del-Real, J., Miller, J. H., & Querol, A. (2023). Divergence in the *Saccharomyces* Species' Heat Shock Response Is Indicative of Their Thermal Tolerance. *Genome Biology and Evolution*, *15*(11), evad207. https://doi.org/10.1093/gbe/evad207

Freel, K. C., Friedrich, A., Hou, J., & Schacherer, J. (2014). Population Genomic Analysis Reveals Highly Conserved Mitochondrial Genomes in the Yeast Species *Lachancea thermotolerans*. *Genome Biology and Evolution*, *6*(10), 2586–2594. https://doi.org/10.1093/gbe/evu203

Gobbi, M., Comitini, F., Domizio, P., Romani, C., Lencioni, L., Mannazzu, I., & Ciani, M. (2013). *Lachancea thermotolerans* and *Saccharomyces cerevisiae* in simultaneous and sequential co-fermentation: A strategy to enhance acidity and improve the overall quality of wine. *Food Microbiology*, *33*(2), 271–281. https://doi.org/10.1016/j.fm.2012.10.004

He, P., Shao, X., Duan, S., Han, D., Li, K., Shi, J., Zhang, R., Han, P., Wang, Q., & Bai, F. (2022). Highly diverged lineages of *Saccharomyces paradoxus* in temperate to subtropical climate zones in China. *Yeast*, *39*(1–2), 69–82. https://doi.org/10.1002/yea.3688

Hranilovic, A., Bely, M., Masneuf-Pomarede, I., Jiranek, V., & Albertin, W. (2017). The evolution of *Lachancea thermotolerans* is driven by geographical determination, anthropisation and flux between different ecosystems. *PLoS ONE*, *12*(9). https://doi.org/10.1371/journal.pone.0184652

Hranilovic, A., Gambetta, J. M., Schmidtke, L., Boss, P. K., Grbin, P. R., Masneuf-Pomarede, I., Bely, M., Albertin, W., & Jiranek, V. (2018). Oenological traits of *Lachancea thermotolerans* show signs of domestication and allopatric differentiation. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-33105-7

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*, *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

Lang, G. I., Botstein, D., & Desai, M. M. (2011). Genetic Variation and the Fate of Beneficial Mutations in Asexual Populations. *Genetics*, *188*(3), 647–661. https://doi.org/10.1534/genetics.111.128942

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

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

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. https://doi.org/10.1111/mec.17669

Peter, J., & Schacherer, J. (2016). Population genomics of yeasts: Towards a comprehensive view across a broad evolutionary scale. *Yeast*, *33*(3), 73–81. https://doi.org/10.1002/yea.3142

Porter, T. J., Divol, B., & Setati, M. E. (2019). *Lachancea* yeast species: Origin, biochemical characteristics and oenological significance. *Food Research International*, *119*, 378–389. https://doi.org/10.1016/j.foodres.2019.02.003

Postigo, V., Esteban, S., & Arroyo, T. (2023). *Lachancea thermotolerans*, an Innovative Alternative for Sour Beer Production. *Beverages*, *9*(1), 20. https://doi.org/10.3390/beverages9010020

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

Scopel, E. F. C., Hose, J., Bensasson, D., & Gasch, A. P. (2021). Genetic variation in aneuploidy prevalence and tolerance across *Saccharomyces cerevisiae* lineages. *Genetics*, *217*(4), 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

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

Vicente, J., Friedrich, A., Schacherer, J., Freel, K., Marquina, D., & Santos, A. (2025). Whole-Genome Sequencing and Phenotyping Reveal Specific Adaptations of *Lachancea thermotolerans* to the Winemaking Environment. *Molecular Ecology*, e17667. https://doi.org/10.1111/mec.17667

Villarreal, P., Villarroel, C. A., O'Donnell, S., Agier, N., Quintero-Galvis, J. F., Peña, T. A., Nespolo, R. F., Fischer, G., Varela, C., & Cubillos, F. A. (2022). Late Pleistocene-dated divergence between South Hemisphere populations of the non-conventional yeast *L. cidri*. *Environmental Microbiology, 24*(12), 5615–5629. https://doi.org/10.1111/1462-2920.16103

Wadgymar, S. M., DeMarche, M. L., Josephs, E. B., Sheth, S. N., & Anderson, J. T. (2022). Local Adaptation: Causal Agents of Selection and Adaptive Trait Divergence. *Annual Review of Ecology, Evolution, and Systematics, 53*(1), 87–111. https://doi.org/10.1146/annurev-ecolsys-012722-035231

Wang, P., Driscoll, W. W., & Travisano, M. (2024). Genomic sequencing reveals convergent adaptation during experimental evolution in two budding yeast species. *Communications Biology, 7*(1), 1–9. https://doi.org/10.1038/s42003-024-06485-y

Xia, W., Nielly-Thibault, L., Charron, G., Landry, C. R., Kasimer, D., Anderson, J. B., & Kohn, L. M. (2017). Population genomics reveals structure at the individual, host-tree scale and persistence of genotypic variants of the undomesticated yeast *Saccharomyces paradoxus*

in a natural woodland. *Molecular Ecology*, *26*(4), 995–1007. https://doi.org/10.1111/mec.13954

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

Zakrzewska, A., van Eikenhorst, G., Burggraaff, J. E. C., Vis, D. J., Hoefsloot, H., Delneri, D., Oliver, S. G., Brul, S., & Smits, G. J. (2011). Genome-wide analysis of yeast stress survival and tolerance acquisition to analyze the central trade-off between growth rate and cellular robustness. *Molecular Biology of the Cell*, *22*(22), 4435–4446. https://doi.org/10.1091/mbc.e10-08-0721

Zhu, Y. O., Sherlock, G., & Petrov, D. A. (2016). Whole Genome Analysis of 132 Clinical *Saccharomyces cerevisiae* Strains Reveals Extensive Ploidy Variation. *G3: Genes|Genomes|Genetics*, *6*(8), 2421–2434. https://doi.org/10.1534/g3.116.029397

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). https://doi.org/10.1073/pnas.1323011111