

Substrate, temperature, and geographical patterns among nearly 2,000 natural yeast isolates

Keywords: Fungi, Yeasts, Ascomycota, Basidiomycota, *Saccharomyces*, Ecology

William J. Spurley*

wspurley@wisc.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

ORCID: 0000-0003-2280-3624

Kaitlin J. Fisher*

kjfisher5@wisc.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

ORCID: 0000-0002-7536-0508

Quinn K. Langdon*

qlangdon@stanford.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

Department of Biology, Stanford University, Stanford, CA 94305, USA

ORCID: 0000-0002-2141-3571

Kelly V. Buh

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

Martin Jarzyna

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

ORCID: 0000-0002-2485-6589

Max A. B. Haase

Max.Haase@nyulangone.org

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

Vilcek Institute of Graduate Biomedical Sciences and Institute for Systems Genetics, NYU Langone Health, New York, NY 10016, USA

ORCID: 0000-0002-8962-0446

Kayla Sylvester

kayla.sylvester@duke.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

Dept. of Molecular Genetics and Microbiology, Duke University, Durham, NC 27708, USA

ORCID: 0000-0003-4570-7432

Ryan V. Moriarty

rmoriarty@wisc.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, Microbiology Doctoral Training Program, University of Wisconsin-Madison, Madison, WI 53726, USA

ORCID: 0000-0001-7100-1906

Daniel Rodriguez

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

Angela Sheddan

asheddan@wc314.org

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

West Carroll High School, Savannah, IL 61074, USA

Sarah Wright

sdwright79@gmail.com

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

EAGLE School of Madison, Fitchburg, WI 53711, USA

Lisa Sorlie

sorlilis@bonduel.k12.wi.us

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

School District of Bonduel, Bonduel, WI 54107, USA

Amanda Beth Hulfachor

hulfachor@wisc.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

Dana A. Opulente*

dana.opulente@villanova.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

Department of Biology, Villanova University, Villanova, PA 19085, USA

ORCID: 0000-0003-3224-7510

Chris Todd Hittinger ^

cthittinger@wisc.edu

Laboratory of Genetics, J. F. Crow Institute for the Study of Evolution, Wisconsin Energy Institute, DOE Great Lakes Bioenergy Research Center, Center for Genomic Science Innovation, University of Wisconsin-Madison, Madison, WI 53726, USA

ORCID: 0000-0001-5088-7461

*WJS, KJF, QKL, and DAO should be considered joint first author

^Corresponding author

Data availability

The data that supports the findings of this study are available in the supplementary material of this article. R code for analyses for which custom scripts were written is archived at https://github.com/KatieFish/Spurley_Fisher_2021. ITS sequences for 1,905 isolates are available on GenBank under accession numbers OK050592-OK052496. Fifty-seven sequences failed to meet the submission criteria of GenBank due to sequence length and number of ambiguous bases and are provided in Suppl. Table 15. All supplemental materials are provided in a Figshare repository <https://doi.org/10.6084/m9.figshare.c.5599392>.

Author contributions

QKL, KVB, MJ, MABH, KS, RVM, DR, AS, SW, LS, & ABH isolated and genotyped strains. KVB, MJ, MABH, QKL, & DAO curated the dataset and compiled metadata using a database built by ABH. WJS & KJF conceived of and performed statistical analyses with input from DAO. WJS & KJF wrote the manuscript. CTH, DAO, & QKL edited the manuscript, and all co-authors approved it. KJF, QKL, DAO, & CTH provided mentorship at various stages of the study.

Acknowledgments

We thank all past and present participants in the Wild YEAST Program (<http://go.wisc.edu/wildyeast>), including all individuals that contributed samples to this study: Alison Coffey, Alysha Heimberg, Anita R. & S. Todd Hittinger, Anna S. Kropornicka, Kurt Ehlert, Ashanti Rogers, Beth Sargent, Bill Saucier, Bill Vagt, Brian P. H. Metzger, Meihua Christina Kuang, D. Leith Nye, Drew T. Doering, EmilyClare P. Baker, Erik D. Jessen, Erica L. Macke, Harry Iwatsuki, Jacob Barnes, Jin Kang, John E. Pool, Katie Burjek, Laura Beck, Leslie Shown, Mary B. O'Neill, Russell L. Wrobel, William G. Alexander, S. Smead and Christina Ruhaak, Mike Siegel, the Zasadil family, and Mel Langdon. This material is based upon work supported by the National Science Foundation under Grant Nos. DEB-1253634 (to C.T.H.), DEB-1442148 (to C.T.H.), DEB-2110403, and DGE-1256259 (to Q.K.L.); in part by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-SC0018409); and the USDA National Institute of Food and Agriculture (Hatch Project 1020204). W.J.S. was supported by a University of Wisconsin-Madison Genetics and Genomics Undergraduate Distinguished Research Fellowship while completing this work. K.J.F. is a Morgridge Metabolism Interdisciplinary Fellow of the Morgridge Institute for Research. Q.K.L. was also supported by the Predoctoral Training Program in Genetics, funded by the National Institutes of Health (Grant No. 5T32GM007133). C.T.H. is a Pew Scholar in the Biomedical Sciences and an H. I. Romnes Faculty Fellow, supported by the Pew Charitable Trusts and Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation, respectively.

Competing interests
None declared.

Take Away

- Analysis of environmental metadata of nearly 2,000 yeast isolates.
- Individual yeast taxa associate with specific substrates and plant genera.
- Optimal yeast isolation temperature differs depending on taxonomic rank.
- Substrate type and isolation temperatures affect isolated yeast diversity.

Abstract

Yeasts have broad importance as industrially and clinically relevant microbes and as powerful models for fundamental research, but we are only beginning to understand the roles yeasts play in natural ecosystems. Yeast ecology is often more difficult to study compared to other, more abundant microbes, but growing collections of natural yeast isolates are beginning to shed light on fundamental ecological questions. Here we used environmental sampling and isolation to assemble a dataset of 1,962 isolates collected from throughout the contiguous United States of America (USA) and Alaska, which were then used to uncover geographic patterns, along with substrate and temperature associations among yeast taxa. We found some taxa, including the common yeasts *Torulasporea delbrueckii* and *Saccharomyces paradoxus*, to be repeatedly isolated from multiple sampled regions of the US, and we classify these as broadly distributed cosmopolitan yeasts. A number of yeast taxon - substrate associations were identified, some of which were novel and some of which support previously reported associations. Further, we found a strong effect of isolation temperature on the phyla of yeasts recovered, as well as for many species. We speculate that substrate and isolation temperature associations reflect the ecological diversity of and niche partitioning by yeast taxa.

Introduction

Yeasts are fungi that spend at least part of their life cycle in a unicellular state and do not form fruiting bodies (Kurtzman, Fell and Boekout 2011). They are globally distributed, speciose, and diverse, and many have important relationships with humans. Yeasts are polyphyletic and are represented in the phyla Ascomycota (subphyla Saccharomycotina and Taphrinomycotina) and Basidiomycota (Nagy et al. 2014). The ascomycete subphylum Pezizomycotina, which is sister to the Saccharomycotina, also includes some dimorphic fungi that spend significant portions of their life cycle in a unicellular state, but they are not traditionally regarded as yeasts (Kurtzman, Fell and Boekout 2011). The phylogenetic, phenotypic, and genomic diversity within yeasts suggests that divergent yeasts have distinct ecological roles. Despite their importance and ubiquity, the ecology and distribution of most yeast species remain poorly understood. The advances in culture-independent metagenomic characterization of microbial communities that have enabled increasingly detailed characterization of bacterial communities (Franzosa et al. 2015) do not lend themselves well to the study of yeasts; typical methods for environmental detection of bacterial microbes are less useful for eukaryotic microbes that are low-abundance community members (Pereira-Marques et al. 2019). As a result, most research in yeast ecology is reliant on environmental sampling and isolation in synthetic media, either by direct plating (Boynton, Kowallik, Landermann and Stukenbrock 2019; Rose, Winston and Hieter 1990; Stefanini et al. 2012) or enrichment (Knight and Goddard 2015; Liti, Warringer and Blomberg 2017; Sniegowski, Dombrowski and Fingerman 2002).

Isolation-dependent approaches are admittedly not ideal for studying ecology because most investigations are performed away from the sample collection site in laboratory conditions that differ drastically from natural conditions, because of the presence of bottlenecks in the isolation process, and because most fungal diversity is likely unculturable (James, Stajich, Hittinger and Rokas 2020). Despite these limitations, a broad and dense enough sampling strategy can begin to illuminate the cryptic ecology of yeasts. For example, exhaustive environmental sampling revealed that *S. cerevisiae* is not the purely domestic species it was once

believed to be (Török, Mortimer, Romano, Suzzi and Polsinelli 1996; Vaughan-Martini and Martini 1995), which has led to a much better understanding of the global distribution of *S. cerevisiae* (Naumov, Naumova and Sniegowski 1998; Peter et al. 2018; Wang, Liu, Liti, Wang and Bai 2012). Similar insights have been gained into the distributions of other members of the genus *Saccharomyces* through isolation-based surveys (Charron, Leducq, Bertin, Dubé and Landry 2014; Langdon et al. 2020; Nespolo et al 2020). While studies, such as these, are beginning to elucidate the distributions (and in some cases ancestral origins of) well-studied species, there is still much to be understood about the natural habitats and ecology of the vast majority of species.

Most surveys of the natural habitat of yeasts are limited in sampling depth and the breadth of substrate types sampled and therefore lack statistical power to assign preferred substrates to yeast taxa. Recent analyses that have overcome these limitations by using large datasets of isolates have been able to uncover associations between yeast species and habitats (Charron et al. 2014; Opulente et al. 2019; Sylvester et al. 2015). Connecting yeast species to biotic or abiotic substrates is the first step in inferring specific niches. Hypotheses regarding the ecological roles of yeasts in nutrient cycling, symbioses, and microbial interactions can be generated and tested once their preferred substrates are known. These connections also provide practical information about the natural reservoirs of yeasts that are significant to humans, both those that pose threats as emerging pathogens (Bensasson et al. 2019; Friedman and Schwartz 2019; Geddes-McAlister and Shapiro 2019; Opulente et al. 2019) and those that are potentially valuable for the development of bioindustrial capabilities (Cordente, Schmidt, Beltran, Torija and Curtin 2019; Hittinger, Steele and Ryder 2018; Spagnuolo, Yaguchi, and Blenner 2019; Wohlbach et al. 2011).

The most direct use of the characterizations of the natural reservoirs of different yeast taxa would be of applied use in the pragmatic design of further isolation protocols. Individual species or groups of species could be targeted by deep sampling of associated substrates. Likewise, surveys of yeast communities would benefit from sampling substrates that are known to yield high levels of diversity. Yeasts have historically been isolated mostly from fruit, soil, and bark, but there have been successful isolations from many other substrates, including aquatic environments, flowers, leaves, and insects (Han et a., 2015; Roth Jr, Orpurt and Ahearn 1964;

Sláviková, Vadkertiová and Vránová 2007; Stefanini et al. 2016). Coupling diversity of substrates with regional distribution of yeasts could lead to more precision in isolation strategies.

In an effort to better understand the distribution and natural ecology of yeasts, we have launched an ongoing initiative to sample yeast diversity from natural substrates. Previously, our lab investigated preferences and associations of yeasts among 589 isolates (Sylvester et al., 2015). This growing collection has also described associations of pathogenic yeasts and their geographic distribution in the United States of America (USA) (Opulente et al. 2019). Here, we have expanded this collection to curate a dataset of 1,962 isolates. To our knowledge, the collection analyzed here is the largest collection of natural yeast isolates gathered using common isolation protocols that has been published to date. We used this dataset to examine the spatial distribution patterns of diverse yeast species, look for associations between specific taxa and habitats, and examine overall yeast diversity among different types of sampled substrates. Using spatial and temporal isolation patterns, we also identified yeasts that exhibited a cosmopolitan distribution within our dataset. Due to the statistical power afforded by such a large dataset, we confirmed previously suggested substrate associations and identified new ones. Finally, we examined the diversity of yeasts associated with different substrate types to better understand where yeast communities are expected to be especially complex.

Methods

Sample collection and yeast isolation

Samples of natural substrates were collected across the continental USA and Alaska by participants in the Wild YEAST Program (<http://go.wisc.edu/wildyeast>), which includes academic and citizen scientists. Data recording the GPS location; substrate types; and, if applicable, plant species accompanied all collections. Sampling sites were generally natural areas, and sites that were strongly associated with humans or industry were avoided. Samples were collected using a sterile bag and without any human contact to prevent contamination. These samples were then either processed immediately or stored at 4°C for up to three weeks. Isolation of yeasts from samples was performed as described in Sylvester et al., 2015, with the addition of 4°C

as an isolation temperature in some cases. Briefly, a small amount of sample was inoculated in a 15-ml conical tube into 9ml of Wild Yeast Medium (5 g/L $(\text{NH}_4)_2\text{SO}_4$, 1 g/L Synthetic Complete Dropout mix (US Biological), 1.72 g/L Yeast Nitrogen Base w/o Amino Acids, & Carbohydrate w/o Ammonium Sulfate (US Biological), 0.1 g/L Ampicillin, 0.03 g/L Chloramphenicol), supplemented with either 8% or 0.8% glucose. The Sylvester et al. 2015 recipe included $(\text{NH}_4)_2\text{SO}_4$, but it was mistakenly excluded from the Materials and Methods. Small amounts of the same sample were used to inoculate three tubes which were then incubated at three different temperatures, 10°C, 22°C, or 30°C, until there was evidence of growth or fermentation (generally 1-2 days at warm temperatures, and up to two months at cold temperatures). Additional enrichments at 4°C were performed for many, but not all samples. 10 μL of culture was then passaged once more into 4 ml Wild Yeast Medium (1:400) in a 5-ml conical tube and incubated at their respective temperatures until growth was evident. Enriched cultures were plated to YPD agar plates. Plates were inspected for discrete colonies with distinct morphotypes. Morphotype distinction was based on variation in colony texture, color, surface, elevation, and margin. Colonies with spreading hyphal morphologies were typically avoided. This selection filtered isolates based on colony morphology, such that some Pezizomycotina species that form characteristic “yeast-like” colonies were occasionally isolated, while most Pezizomycotina species were not. Single colonies were selected for identification, restreaked to purify, and frozen down in 15% glycerol stocks.

Taxonomy was assigned through amplification and sequencing of the ITS region of the *rDNA* locus. Specifically, the entire ITS region was amplified using the oligonucleotides ITS1 and NL4 (McCullough, Clemons, McCusker and Stevens 1998; O’Donnell 1993), and the ITS1/2 region was sequenced using the oligonucleotide ITS4 (McCullough et al. 1998). ITS sequences were queried against the nucleotide database of GenBank using BLASTN, and query sequences that matched their best database hit by 97% or more were positively identified based on that search. Queries that did not return matches of at least 97% were sequenced at the D1/D2 domain of the *rDNA* locus using NL4 as a sequencing oligonucleotide to achieve a positive identification. Isolates that were more than 3% divergent in the D1/D2 domain were considered candidates for novel species, which will be formally described elsewhere. 51 such isolates were identified to the level of 22

different genera, and all potentially novel species are indicated in **Suppl. Table 1**. Some isolates can only be identified to the level of species complex due to poorly delineated species phylogenies. Because there are a number of candidate novel species and isolates belonging to species complexes, we use the term operational taxonomic unit (OTU) in lieu of species throughout.

A detailed protocol is provided in **Suppl. Note 1**.

Dataset curation

In some instances, we were unable to amplify the ITS or D1/D2 sequences of an isolate. Any strain that fell into this category was subject to three independent attempts to achieve a sequence. If the isolate failed three times, they were excluded from analyses.

Oftentimes, multiple isolates of the same OTU isolated from the same exact substrate sample appear in the dataset. To eliminate these possibly duplicate strains, prior to all analyses, except one category of analysis as described below, identical OTUs isolated from the same processed sample were eliminated. Exclusion of identical OTUs isolated from the same processed sample resulted in 1,516 unique isolations (**Fig. S1**). The same environmental sample was sometimes subjected to multiple different temperatures for isolation, and these were considered separate isolations for temperature analyses only.

Detection of cosmopolitan OTUs

We sought to use our extensive dataset of isolation locations to determine which OTUs in our data could be considered cosmopolitan across the expansive sampled region. Instead of simply determining which OTUs were frequently isolated temporally or spatially, we considered differences in sampling depth across the USA to identify those OTUs that were always isolated in regions that were deemed sufficiently sampled to detect them. We began by dividing the country into ten regions, the Arctic (Alaska) plus the nine climatic regions of the contiguous USA defined by the National Oceanic and Atmospheric Association (NOAA) (**Fig. S2**). For each OTU, we then calculated an expected discovery rate based on the most densely sampled climatic region in which that OTU was found. For example, *Torulaspora delbrueckii* was isolated 60 independent times out of 382 total samples collected in the Upper Midwest, giving it an isolation rate of once every 6.4 samples (382 total

samples/60 OTU isolations). Thus, we would then expect *T. delbrueckii* to be isolated in all regions that had been sampled at least 7 times. In this manner, we determined the regions from which each OTU was expected to have been isolated and compared these to the regions in which each OTU was actually found. We eliminated from consideration all yeasts that had only been isolated only once (103 OTUs) and all those yeasts that were only expected in a single region (75 OTUs, all expected only in the Upper Midwest due to high sampling density). We defined cosmopolitan yeasts as OTUs that were isolated either in all regions where they were expected or in all but one region where they were expected. For example, *T. delbrueckii* was expected in all regions except the Northern Rockies and Southwest, and it was found in all expected regions with the single exception of the Arctic. Since we allowed for one region of not being detected where expected, *T. delbrueckii* was therefore categorized as a cosmopolitan yeast OTU.

Association analyses

We first determined if yeasts belonging to any specific taxonomic groups (phyla and subphyla categories) were isolated more frequently than expected by chance from specific substrates or isolation temperatures. One-tailed Fisher's exact tests followed by Benjamini Hochberg post-hoc corrections were used to test for enrichment of taxonomic groups in substrate-types and isolation temperatures. We then used permutations to identify individual OTUs that were isolated more frequently than random expectation from different substrates. All substrate samples associated with a unique isolation were permuted by randomly reassigning isolation OTUs across substrate samples without replacement for 10,000 permutations. P-values were assigned based on the number of permutations in which an OTU was assigned to a substrate at an equal or higher rate than observed in our data. Benjamini-Hochberg adjustments were applied to correct for multiple testing. Expected frequencies of OTU-substrate association were calculated from permutation means and contrasted to observed frequencies in figures. An identical method was applied to all substrate genera and isolation temperatures associated with a unique isolation. We did not examine our data for depleted combinations because we cannot statistically distinguish between true depletion and undersampling. Permutation analyses were performed with custom R scripts (see Data Availability).

Diversity analyses

Sample-based alpha diversity analyses were carried out using the methods outlined in Chao et al. 2014 using the associated iNEXT R package (Hsieh, Ma, and Chao 2020). Asymptotic diversity estimates were calculated for substrate-types and isolation temperatures. To minimize effects due to low sampling when comparing substrate types, asymptotic estimates of species richness and Shannon entropy were calculated only for substrates that had an estimated sample coverage greater than 0.75. Estimated isolation temperature diversity by taxonomic group was only calculated for the two most heavily sampled groups, Saccharomycotina and Basidiomycota. To focus analyses on total species diversity, all rarefactions were performed with an order number (q) set to 0. Beta diversity among isolation temperatures was estimated by calculating the Jaccard distance (d_j) using the R vegan package (Oksanen et al. 2019).

Results & Discussion

Curated dataset of isolates

The final dataset curated for this study consisted of 1,962 isolations of 244 unique operational taxonomic units (OTUs) isolated from 688 unique samples (**Fig. S1A, Suppl. Table 1**). Rarefaction indicated that the true richness is not represented in the sampled range and that this sample is an underestimate of richness (**Fig. S3A**). 483 strains in this study were also included in a previously published survey of wild yeast isolates (Sylvester et al. 2015), as were an additional 34 strains from subsequent taxon-specific studies (Krause et al. 2018; Langdon et al. 2020; Leducq et al. 2014; Opulente et al. 2019; Peris et al. 2016; **Suppl. Table 1**). 106 strains from Sylvester et al. 2015 were excluded due to differences between isolation protocols, missing metadata, or other discrepancies that prevented inclusion in the present analyses. Accompanying metadata for each isolate included GPS coordinates of sample collection, the general substrate type, the specific substrate type, the genus and species of biotic substrates, the incubation temperature used in laboratory isolation, and the ITS sequences used to identify the species. Yeasts from both major divisions or phyla of Dikarya, Basidiomycota (76 OTUs) and Ascomycota (168 OTUs), were represented in the data. Within Ascomycota,

isolates belonged to the subphyla Saccharomycotina (150 OTUs), Pezizomycotina (17 OTUs), and Taphrinomycotina (1 OTU) (**Fig. S3B, Fig. S4**). Pezizomycotina and Taphrinomycotina were represented by relatively few isolates due to systematic avoidance of colonies belonging to the former (see Methods) and the presumed low abundance and need for specialized enrichment for the latter (Benito, Calderón and Benito et al. 2018).

Geospatial variation among isolations

The Upper Midwest was overrepresented in our dataset

Isolates were obtained from all major climate regions of the contiguous United States and Alaska, except for the Southwest (**Fig. S2**). The Upper Midwest was the most densely sampled region with 869 unique isolations of 166 OTUs, followed by 144 isolations of 59 OTUs from the Ohio Valley, 140 isolations of 57 OTUs from the Northeast, 135 isolations of 59 OTUs from the Northwest, 104 isolations of 32 OTUs from the Southeast, 40 isolations of 33 OTUs from the West, 32 isolations of 24 OTUs from the South, and 8 isolations of 8 OTUs from the Northern Rockies. Zero isolates were collected from the Southwest. An additional 44 isolations of 24 OTUs came from Alaskan samples, which we classified as Arctic.

Rarely and frequently isolated OTUs

Of the 244 unique OTUs, 103 were singletons that were found in only one isolation (**Suppl. Table 2**), while the remaining 141 generally were isolated fewer than 30 times. The singleton OTUs isolated in this study were primarily isolated from bark and soil, but singletons were not enriched for any substrate category (**Fig. S5A**). Singletons were evenly distributed across sampled regions (**Fig. S5C**).

Seventeen yeast OTUs were isolated over 20 independent times in this dataset (**Suppl. Table 3**). Among these were two *Saccharomyces* species, *Saccharomyces paradoxus* and *Saccharomyces cerevisiae*, for which considerable population genetic work has already been done (Liti et al. 2009; Xia et al. 2017; Peter et al. 2018, Hénault et al. 2017). Other frequently isolated OTUs are of broad interest but have been subjected to relatively little population genetic analysis. These include *Kluyveromyces lactis*, *T. delbrueckii*, *Hanseniaspora uvarum*,

and the *Metschnikowia pulcherrima* spp. complex. The collection described herein is teeming with potential for population-level studies for these and more OTUs.

12 cosmopolitan OTUs

Although many yeast species are frequently referred to as cosmopolitan, there are no clear criteria by which yeasts are or are not classified as such. We examined our dataset for yeasts that were widely distributed across the USA by considering both the number and spatial distribution of isolations for each OTUs. Briefly, we defined an OTU-specific isolation rate and then looked for OTUs that were generally able to be isolated in all regions where sampling was dense enough to detect them (**Suppl. Table 4**, see Methods). While this approach cannot exclude any yeasts from being broadly distributed, it can identify those OTUs in our dataset that are most likely cosmopolitan in distribution. We determined that 12 OTUs in our dataset were cosmopolitan and generally found when expected across climatic regions (*Candida railenensis*, *Candida sake*, *Papiliotrema flavescens*, *Pichia manshurica*, *Cyberlindnera saturnus*, *Debaryomyces hansenii*, *Leucosporidium scottii*, *Curvibasidium cygneicollum*, *S. paradoxus*, *Scheffersomyces ergatensis*, *T. delbrueckii*, and *Wickerhamomyces anomalus*, **Suppl. Table 5**). Only 6 of these OTUs were also among the 17 OTUs isolated over 20 times, which suggests that frequent isolation in individual regions may not indicate otherwise broad geographic distributions. Two of the OTUs we identified as cosmopolitan were isolated 10 or fewer times (*D. hansenii*, *Le. scottii*), but were nonetheless always detected in regions with sufficient sampling (**Fig. S6A**, **Fig. S7C**). Cosmopolitan OTUs ranged from being isolated in just two regions (due to requiring sufficiently dense sampling to detect them) to being isolated in a maximum of 7 regions (**Fig. S6B**, **Fig. S7C**). No OTU was isolated from all nine sampled USA regions. Cosmopolitan yeasts were not isolated at equal rates among substrates, but rather were significantly enriched for soil samples ($P_{\text{adj}}= 0.003$, **Fig. S7A**).

Our approach to defining cosmopolitan isolates is imperfect and affected by sampling biases in our data, but the clear statistical criteria represent an improvement over the ad hoc subjective assessments that are sometimes made. The approach only considered total climatic region sampling density and not the sampling densities of any specific substrates. Most substrates were not uniformly sampled across regions, and yeasts with

high substrate-specificities may have been missed. The climatic regions we use to categorize sampling locations were also unevenly sampled, which would likely lead to common species being missed by chance and rare species not being considered because they would not be expected in low sampled regions. Unfortunately, powerful approaches for determining distributions for yeast OTUs remain elusive. They are often not abundant enough to be detected by metagenome sequencing or to be reliably isolated in all samples. The clandestine biogeography of yeasts is mainly studied using the same approaches we use here – laborious environmental sampling and laboratory isolation. While it is possible that many more yeast OTUs in our collection may also have wide distributions, we can at least conclude that the eleven OTUs we identify as cosmopolitan show patterns of ubiquitous distributions, at least across the USA. Soil was one of the most broadly sampled substrate types across climatic regions, which may be one reason the OTUs we identify as cosmopolitan seem to be soil-associated taxa. Conversely, soil is a ubiquitous substrate, and the strong statistical enrichment for cosmopolitan yeasts among soil samples may suggest soil as a reservoir of cosmopolitan yeasts.

Substrate and temperature associations

Substrate – higher taxonomic group associations

We were able to assign one of 40 specific substrate categories to 1,520 isolations, which had yielded 243 unique yeast OTUs (**Fig. S8A**). Contingency tables were used to identify associations between higher taxonomic groups of yeast (phylum Basidiomycota and the three Ascomycota subphyla) and substrate. When taxonomic group representation across substrate types was examined (**Suppl. Table 6**), we found that leaves were enriched for Basidiomycota ($P_{\text{adj}}= 0.015$) but found no other significant enrichment of taxonomic groups across substrate types.

Individual OTU - substrate associations

Permutations were used to identify associations between specific taxonomic units and sampled substrates that occurred more often than expected by chance. Of the 702 observed combinations (**Fig. S9A**), we found 23 yeast OTU-by-substrate associations that occur more frequently than expected by chance ($P_{\text{adj}} < 0.05$,

Fig. 1A, Suppl. Table 7). Soil, which was the most heavily sampled substrate, was significantly associated with two Basidiomycota and three Saccharomycotina OTUs (*Mrakia* spp. complex, $P_{\text{adj}}= 0.012$; *Trichosporon porosum*, $P_{\text{adj}}= 0.039$; *T. delbrueckii*, $P_{\text{adj}}= 0.013$; *Cy. saturnus*, $P_{\text{adj}}= 0.025$; *S. paradoxus*, $P_{\text{adj}}= 0.039$). Bark was also heavily sampled but yielded only two significantly associated OTUs, both of which were Saccharomycotina (*Lachancea kluyveri*, $P_{\text{adj}}= 0.033$; *Sc. ergatensis*, $P_{\text{adj}}= 0.045$). Leaf samples were also associated with two OTUs, both of which were Basidiomycota (*Rhodotorula nothofagi*, $P_{\text{adj}}= 0.013$; *Mrakia gelida*, $P_{\text{adj}}= 0.013$). Fungal samples were significantly associated with three Saccharomycotina OTUs (*H. uvarum*, *Suhomyces bolitotheri*, *Teunomyces cretensis/kruisii* spp. complex, $P_{\text{adj}}< 0.0001$); fruit with one Basidiomycota OTU and two Saccharomycotina OTUs (*Cu. cygneicollum*, $P_{\text{adj}}= 0.021$; *Pichia kudriavzevii*, $P_{\text{adj}}= 0.039$; *H. uvarum*, $P_{\text{adj}}= 0.032$); and sand with two Saccharomycotina OTUs (*Kazachstania serrabonitensis*, $P_{\text{adj}}< 0.0001$; *Pichia scaptomyzae*, $P_{\text{adj}}= 0.025$). Feathers, flowers, insects, lichens, and needles were associated with a single OTU each (*Peterozyma toletana*, $P_{\text{adj}}= 0.046$; *Barnettozyma californica*, $P_{\text{adj}}= 0.039$; *Kwoniella newhampshirensis*, $P_{\text{adj}}= 0.039$; *Scleroconidioma sphagnicola*, $P_{\text{adj}}< 0.0001$; *Sydowia polyspora*, $P_{\text{adj}}= 0.013$; respectively). Plant matter samples (excluding matter that fell into other plant-related categories) were associated with a single OTU (*Candida mycetangii*, $P_{\text{adj}}= 0.025$).

When possible, substrate genera were identified for plant and fungal substrates directly sampled (**Fig. S8B**). Substrate genera were also assigned to samples indirectly associated with plant genera (e.g. soil sampled from base of tree). In total, 66 substrate genera were identified across 1,061 isolations comprising 198 OTUs (**Fig. S9B, Suppl. Table 8**). Seventeen of 652 observed substrate genera – yeast associations were found more often than expected by chance (**Fig. 1B**). *Quercus*, the genus to which oak trees belong, was heavily sampled due to reported associations with *Saccharomyces* species (Naumov 1996; Naumov, Naumova and Sniegowski 1998; Sniegowski et al. 2002; Kowallik and Greig 2016; Wang et al. 2012; Zhang, Skelton, Garner and Goddard 2010). Two Saccharomycotina OTUs were significantly associated with *Quercus* spp., (*Kluyveromyces lactis*, $P_{\text{adj}}< 0.0001$; *Lachancea fermentati*, $P_{\text{adj}}= 0.011$). Notably, none of the *Quercus*-associated OTUs belong to the genus *Saccharomyces*, although both do belong to the Saccharomycetaceae, the

family to which *Saccharomyces* belongs. Spruce trees of the genus *Picea* were associated with the dimorphic Pezizomycotina OTU *Sy. polyspora* ($P_{\text{adj}} = 0.007$) and the Saccharomycotina OTU *Kazachstania servazzii* ($P_{\text{adj}} = 0.037$). *Amanita* was the only fungal genus found to be significantly associated with yeast OTUs. *Amanita*, a speciose Basidiomycota genus containing many poisonous and edible mushroom species, was found to be associated with the Basidiomycota OTU *Vanrija humicola* ($P_{\text{adj}} = 0.013$) and the Saccharomycotina *Te. cretensis/kruisii* spp. complex ($P_{\text{adj}} = 0.022$). Eleven additional plant genera were associated with a single OTU each (*Cercis* – flowering shrubs, *M. pulcherrima* spp. complex, $P_{\text{adj}} < 0.0001$; *Betula* – birch trees, *Kwoniella betulae*, $P_{\text{adj}} = 0.022$; *Acer* – maple trees, *T. delbrueckii*, $P_{\text{adj}} = 0.039$; *Thuja* – cypress trees, *D. hansenii*, $P_{\text{adj}} = 0.025$; *Taxus* – yew trees, *Scheffersomyces coipomoensis*, $P_{\text{adj}} = 0.007$; *Taraxacum* – dandelions, *B. californica*, $P_{\text{adj}} = 0.007$; *Pinus* – pine trees, *Schwanniomyces polymorphus*, $P_{\text{adj}} < 0.0001$; *Morus* – mulberries, *Meyerozyma carpophila*, $P_{\text{adj}} = 0.046$; *Festuca* – perennial tufted grasses, *Wickerhamomyces onychis*, $P_{\text{adj}} < 0.0001$; *Alnus* – alder trees, *Nakazawaea anatomiae/populi* spp. complex, $P_{\text{adj}} = 0.007$; *Fagus* – beech trees, *Candida boleticola*, $P_{\text{adj}} = 0.041$).

Associations were examined at both the level of general substrate type and substrate genus because these categories do not overlap entirely. For example, the association of *K. lactis* and the *Quercus* genus of oaks would have been missed at the substrate level, as *K. lactis* had independent isolations from soil 15 times, bark 17 times, and leaves three times, as well as other various substrates like duff, fungi, and sawdust. Nonetheless, we did find corroboration between the two analyses. This consistency was most obvious among the three OTUs that were associated with a biotic substrate and also with a corresponding genus into which that substrate falls: *B. californica* (*Taraxacum* and flower), *Sy. polyspora* (*Picea* and needles), and *Te. cretensis/kruisii* spp. complex (*Amanita* and fungus). These analyses were performed independently on non-identical datasets (**Fig. S9A-B**), and so these corroborative findings lend high confidence in these associations.

In contrast to many previous studies (Naumov 1996; Naumov, Naumova and Sniegowski 1998; Sniegowski et al. 2002; Kowallik and Greig 2016; Wang et al. 2012; Zhang, Skelton, Garner and Goddard 2010), we did not find associations between *Saccharomyces* species and the oak genus *Quercus*. We isolated *S.*

cerevisiae 21 independent times with only five isolations from *Quercus*-associated substrates (Suppl. Table 9). *S. paradoxus* was isolated 43 independent times with 7 isolations from *Quercus* substrates (Suppl. Table 9). The absence of associations cannot be explained by lack of sampling as *Quercus* was the most deeply sampled plant genus in our dataset (Fig. S8B). The absence of statistical associations is noteworthy as oak trees have long been thought of as the natural habitat of *Saccharomyces* species, particularly *S. cerevisiae* and *S. paradoxus* (Kowallik and Greig 2016; Sniegowski et al. 2002; Wang et al. 2012; Zhang et al. 2010). Indeed, a previous study from our lab using a much smaller subset of these data did find a statistical association between the *Saccharomyces* and *Quercus* genera (Sylvester et al. 2015). This significant association was a result of direct hypothesis testing as opposed to the presently applied query of all possible associations. To ensure that these different results were due to increased sampling and not to differences in statistical methods, we applied the same statistics used to detect a *Quercus* - *Saccharomyces* association in Sylvester et al. 2015 and failed to recover a significant association in our larger dataset. Further, if Sylvester et al. 2015 had queried all possible associations using our permutation analysis, the association would have been non-significant. Because of the extensive literature focused on isolating *Saccharomyces* from oak trees, the assumption that oak exudate is a primary habitat of *Saccharomyces* species has percolated pervasively into the broader yeast literature. While it is obvious that *Saccharomyces* species can indeed be isolated from from oak trees and surrounding soils and leaf litter, it is also clear from our data that *S. cerevisiae*, *S. paradoxus*, and other *Saccharomyces* can be found on a wide array of substrates. The abundance of *S. cerevisiae* isolated outside oak-associated habitats has been noted before, leading to the suggestion that *S. cerevisiae* may be a widely distributed generalist, as opposed to occupying narrower sugar-rich niches, such a fruit and oak exudate (Goddard and Greig, 2015). Our findings provide further evidence that *S. cerevisiae*, and *Saccharomyces* spp. more broadly, occupy diverse habitats and are not confined to the bark of oak and related trees.

Studies wielding sufficient power to statistically associate yeast OTUs with habitat-types in a similar manner to our analyses are still rare, and therefore most of what is known about substrate preferences of yeasts come from a few isolations or anecdotal evidence. Still, many of the significant substrate associations we found

in our data confirmed previously reported associations based on fewer isolations. A novel species previously described by our laboratory, *Kw. betulae*, was named for the birch genus from which it was isolated (Sylvester et al. 2015). The presently analyzed data, which contains the same isolates presented in the original description paper and several new isolates, found a statistical association with the *Betula* genus. *Sy. polyspora* is a known conifer pathogen, and we found it was statistically associated with the pine genus *Picea*, as well as with needles (Guertin, Zitouni, Targuay, Hogue and Beaulieu 2018; Talgø et al. 2010). The association we found between the opportunistic human pathogen *P. kudriavzevii* and fruit has been reported several times before (Douglass et al. 2018; Kurtzman, Fell and Boekhout 2011; Opulente et al. 2019). Additional previously described associations that we replicated included *T. delbrueckii* and soil (Kurtzman, Fell and Boekhout 2011), the xylose-consuming yeast *Sc. ergatensis* and bark (Kurtzman, Fell and Boekhout 2011), *H. uvarum* and fruit (Hamby, Hernández, Boundy-Mills and Zalom 2012; Vadkertiová et al. 2012), and *Tr. porosum* and soil (Middelhoven, Scorzetti and Fell 2001). We also found logical indirect connections between the associations we found and previously reported isolation substrates. For example, *Su. bolitotheri* and *Te. cretensis/Te. kruisii* were originally isolated from the guts of basidiocarp-feeding beetles (Suh, McHugh and Blackwell 2004), and we found them both associated with fungi. *Scl. sphagnicola* is considered a moss pathogen (Tsuneda, Chen and Currah 2001), and we found it associated with lichens, an association that may be due to the similar habitats of mosses and lichens in the boreal forest from which it was isolated. Conversely, some associations found in these data were novel. *H. uvarum* has not been previously described as being a common yeast of macrofungi, but we found a robust association between the two. *Cu. cygneicollum* has been isolated from diverse substrates across the globe (Sampaio et al. 2004), but, to our knowledge, it has not been associated with fruit before.

Isolation temperature – higher taxonomic group associations

Isolations in this dataset were performed at one of four temperatures: 4°C, 10°C, 22°C, or 30°C. In total, isolation temperature was confidently assigned for 1,749 isolations. We have previously shown that isolation temperature dramatically affects taxonomic representation among isolates (Sylvester et al. 2015), and we found the same trends in the current data. Significant enrichment for Basidiomycota was observed at cooler isolation

temperatures of 4°C ($P_{\text{adj}}= 1.29 \times 10^{-35}$) and 10°C ($P_{\text{adj}}= 1.71 \times 10^{-54}$, **Fig. 2A, Suppl. Table 10**). Similarly, an enrichment for Saccharomycotina was observed at 22°C ($P_{\text{adj}}= 2.18 \times 10^{-23}$) and 30°C ($P_{\text{adj}} = 8.0 \times 10^{-41}$).

Individual OTU – isolation temperature associations

To drill down on the specific yeast OTUs affected by isolation temperature, we examined our data for specific OTU – isolation temperature associations. Out of 423 yeast OTU – isolation temperature combinations (**Fig. S9C**), 17 significant positive associations were found (**Fig. 2B, Suppl. Table 11**). As expected, all OTUs associated with isolation at 4°C were Basidiomycota, and all of the genus *Mrakia* (*Mrakia blollopis*, $P_{\text{adj}}= 0.009$; *Mrakia aquatica*, $P_{\text{adj}}= 0.037$; *Mr. gelida*, $P_{\text{adj}} < 0.0001$; *Mrakia* spp. complex, $P_{\text{adj}}= 0.013$; *Mrakia cryoconiti*, $P_{\text{adj}}= 0.02$). Five OTUs were associated with 10°C isolation, four of which were Basidiomycota (*C. capitatum*, *Mr. gelida*, *Mrakia* spp. complex, $P_{\text{adj}} < 0.0001$; *Cu. cygneicollum*, $P_{\text{adj}}= 0.019$), and one of which was a Saccharomycotina OTU (*C. sake*, $P_{\text{adj}} < 0.0001$). There were no significant associations at 22°C. Six total OTUs were found to be significantly associated with 30°C isolation, all of which were Saccharomycotina (*Kluyveromyces marxianus*, *L. kluyveri*, $P_{\text{adj}} < 0.0001$; *M. pulcherrima* spp. complex, $P_{\text{adj}}= 0.005$; *S. cerevisiae*, $P_{\text{adj}}= 0.02$; *Candida pseudolambica*, $P_{\text{adj}}= 0.037$; *T. delbrueckii*, $P_{\text{adj}}= 0.037$).

We have now repeatedly found taxonomic differences in isolation temperature, with Basidiomycota dominating cool isolation temperatures and Saccharomycotina dominating warm isolation temperatures (Sylvester et al. 2015). Here we found that taxonomic group effects emerge at the level of individual OTU – temperature associations as well. Eleven of the twelve significant associations with cold temperatures belonged to Basidiomycota. Conversely, all six yeasts associated with a warm isolation temperature belonged to Saccharomycotina and none to Basidiomycota. One significant exception to the higher taxonomic divisions among isolation temperatures was *C. sake*, a Saccharomycotina yeast significantly associated with 10°C isolations. *C. sake* is known as a cold-tolerant species and has been isolated from Arctic environments (Ballester-Tomás et al. 2017). Neither of the two known cold-tolerant *Saccharomyces* species in our dataset had significant isolation temperature associations, but both species were more successfully isolated at room

temperature or slightly lower. *S. eubayanus* was isolated at about the same rate at 10°C (6 isolations) and 22°C (5 isolations) and never isolated at 30°C or 4°C, while *S. uvarum* was isolated only once at 10°C.

Germination, growth rates, and microbial competition are but a few of the processes likely to affect species-specific isolation success at different temperatures. The strong phylogenetic effects we found suggest that intrinsic biological properties drive most of the variation in isolation success. In confirmation of our power to detect biologically real temperature preferences, we recovered an association between *S. cerevisiae* and its known optimal growth temperature of 30°C (Salvadó, Arroyo-López, Barrio, Querol, and Guillamón 2011). These differences are useful for informing isolation approaches for different species and may point to differing ecology of these groups, as well as. For example, we did not recover the same warm-temperature association for *Saccharomyces paradoxus* that we did for *Saccharomyces cerevisiae*, and this result corresponds well with the inferred different temperature-driven distributions of these two sister species (Robinson, Pinharanda and Bensasson 2016).

Substrate and temperature diversity

Variation in diversity by substrate sampled

Maximizing taxonomic diversity recovered when sampling would be advantageous for yeast ecologists and taxonomists. We compared substrate types using two estimated measures of species diversity: species richness, which reflects the total number of species isolated per substrate, and Shannon entropy, which reflects both the total number of species and the evenness with which they are distributed amongst samples. Because substrates were sampled unevenly across our data, we first estimated sample coverage as in Chao et al. 2014 and compared only substrates with estimated sample coverage greater than 75%. Soil, bark, leaves, and fruit were sufficiently sampled to estimate alpha diversity (**Fig. 3A-C, Suppl. Table 12**). Bark samples exhibited a higher estimated Shannon entropy than all other substrate types (**Fig. 3B**), while soil was estimated to have the highest species richness (**Fig. 3C**). These data indicate that soil is highly speciose; however, OTUs isolated from soil samples are dominated by common taxa. Isolates from bark are somewhat less speciose than soil but

are more evenly distributed amongst taxa. This may be reflective of the diversity of plant species from which bark was sampled, the homogeneity of soil ecosystems, or both. Both explanations are also supported by our correlative finding that soil samples are enriched for cosmopolitan taxa. Higher diversity on bark relative to soil may be an ecological property unique to yeasts among fungi, as a study examining metagenome datasets recently found soil to have higher total fungal diversity than shoots or wood (Baldrian, Větrovský, Lepinay and Kohout 2021). Of the four substrates considered, fruit had both the lowest estimated Shannon entropy and estimated species richness. While many yeasts familiar to laboratory and food scientists are found frequently on fruit, rotting fruits represent simple and ephemeral habitats that may be unlikely to support a high diversity of yeasts.

Variation in diversity by isolation temperature

Diversity isolated from samples may also depend on enrichment temperatures, so we also estimated the Shannon entropy index of each enrichment temperature. We found 22°C to have the highest estimated level of diversity and 4°C to have the lowest (**Fig. 4A&B, Suppl. Table 13**). Because isolation temperature had a drastic effect on the subphyla of resultant isolates, we used the same approach to analyze diversity individually within Saccharomycotina and Basidiomycota (**Fig. 4C, Suppl. Table 13**). Pezizomycota was too sparsely sampled to yield useful estimates. Among OTUs within Saccharomycotina, Shannon entropy indices were highest at 30°C and decreased with temperature, ending with a sharp drop off at 4°C. A reciprocal pattern was seen in Basidiomycota, which exhibited high diversity at lower temperatures and a sharp drop off at 30°C. When considering higher taxonomic units separately, maximum diversity is achieved using isolation temperatures at opposite temperature extremes, while collective diversity is highest at 22°C. This result, coupled with the finding that no individual taxonomic unit was preferentially isolated at 22°C, suggests that intermediate temperatures are ideal for inclusively querying the yeast diversity present in a given sample.

Both the yeast taxon-isolation temperature associations and the alpha diversity analyses reported above are indirect evidence that different isolation temperatures yield very different OTU compositions. To directly test this hypothesis, we examined beta diversity amongst isolation temperatures using Jaccard distance (d_j) (**Fig.**

4D, Suppl. Table 14). As expected, there was very little overlap between the assemblages of taxa recovered from isolations at 4°C and 30°C ($d_J = 0.95$). d_J values overall correlated roughly, but not precisely, with the difference between isolation temperatures. For example, d_J between 10°C and 4°C (0.83) was higher than between 22°C and 30°C (0.61). Therefore, a drop from 10°C to 4°C isolations had a greater effect on the taxa recovered than an increase from 22°C to 30°C, despite the former having a difference of 6° and the latter having a difference of 8°. The low beta diversity observed between 22°C and other isolation temperatures (excluding 4°C) further supports the use of isolation temperatures near 22°C to maximize the recovery of yeast taxonomic diversity.

Conclusions: A window into yeast ecology reveals geospatial variation, habitat specificity, and diverse thermal niches

Inferences regarding the distributions, habitat preferences, and broader ecological roles of yeasts have proven difficult to generate due to the low abundance of yeasts in nature and the tedium associated with detecting them. Here, we leveraged a dataset of nearly 2,000 natural yeast isolates, the largest such collection accrued using the same standardized isolation protocols and conditions, to gain insight into patterns of yeast distribution, habitat specificity, and temperature associations. While many yeast taxa were repeatedly isolated in our dataset, we found that only 12 such taxa were always isolated in regions where sampling should have been dense enough to detect them. Our analysis does not exclude any OTU from being cosmopolitan; rather, we present a simple method for objectively detecting those taxa that are broadly distributed across regions. We similarly employed an objective statistical method of detecting associations between yeast OTUs and sampled substrates and found some yeast OTUs to exhibit strong patterns of substrate specificity. When the same approach was applied to isolation temperature, we found a significant effect of incubation temperature on the phylogenetic groups ultimately isolated, a result we had reported before (Sylvester et al. 2015) that is firmly reinforced here. Finally, we examined the overall yeast diversity achieved from different substrates and using different incubation temperatures, and found bark samples and room temperature incubations to yield the

highest yeast diversity. Taken together, these results paint a picture of ecological structure in the distribution, habitat, and temperature preferences of yeasts.

All data is provided in the form of supplemental tables. Our isolation protocol is provided as a Supple. Note.

Suppl. Table 1 – Raw metadata for 1,962 isolates.

Suppl. Table 2 – OTUs isolated only once.

Suppl. Table 3 – OTUs isolated greater than 20 times.

Suppl. Table 4 – Discovery rates and regional expectations for 66 OTUs.

Suppl. Table 5 – Cosmopolitan OTUs.

Suppl. Table 6 – Contingency table for higher taxonomic group representation across substrate types.

Suppl. Table 7 – Permutation results for specific substrate – yeast OTU associations.

Suppl. Table 8 – Permutation results for substrate genus – yeast OTU associations.

Suppl. Table 9 – All isolates of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*.

Suppl. Table 10 – Contingency table for higher taxonomic group representation across isolation temperatures.

Suppl. Table 11 – Permutation results for isolation temperature – yeast OTU associations.

Suppl. Table 12 – Diversity indices for 4 substrate types.

Suppl. Table 13 – Diversity indices for 4 isolation temperatures.

Suppl. Table 14 - Jaccard distance matrix of isolation temperatures.

Suppl. Table 15 - ITS Regions of 57 isolates failing to meet GenBank submission criteria.

Suppl. Note 1 - Detailed yeast enrichment, isolation, and identification protocol.

References

- Baldrian, P., Větrovský, T., Lepinay, C., & Kohout, P. (2021). High-throughput sequencing view on the magnitude of global fungal diversity. *Fungal Diversity*, 1–9.
- Ballester-Tomás, L., Prieto, J. A., Gil, J. V., Baeza, M., & Randez-Gil, F. (2017). The Antarctic yeast *Candida sake*: Understanding cold metabolism impact on wine. *International Journal of Food Microbiology*, 245, 59–65.
- Benito, Á., Calderón, F., & Benito, S. (2018). *Schizosaccharomyces pombe* isolation protocol. In *Schizosaccharomyces pombe* (pp. 227–234). Springer.
- 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.
- Boynton, P. J., Kowallik, V., Landermann, D., & Stukenbrock, E. H. (2019). Quantifying the efficiency and biases of forest *Saccharomyces* sampling strategies. *Yeast*, 36(11), 657–668.
- Chao, A., Gotelli, N. J., Hsieh, T. C., Sander, E. L., Ma, K. H., Colwell, R. K., & Ellison, A. M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs*, 84(1), 45–67.
- Charron, G., Leducq, J. B., Bertin, C., Dubé, A. K., & Landry, C. R. (2014). Exploring the northern limit of the distribution of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* in North America. *FEMS Yeast Research*, 14(2), 281–288.
- Cordente, A. G., Schmidt, S., Beltran, G., Torija, M. J., & Curtin, C. D. (2019, June 4). Harnessing yeast metabolism of aromatic amino acids for fermented beverage bioflavouring and bioproduction. *Applied Microbiology and Biotechnology*, Vol. 103, pp. 4325–4336.
- Douglass, A. P., Offei, B., Braun-Galleani, S., Coughlan, A. Y., Martos, A. A. R., Ortiz-Merino, R. A., ... Wolfe, K. H. (2018). Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names. *PLoS Pathogens*, 14(7), e1007138.
- Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C., & Huttenhower, C. (2015). Sequencing and beyond: Integrating molecular “omics” for microbial community profiling. *Nature Reviews Microbiology*, 13(6), 360–372.
- Friedman, D. Z. P., & Schwartz, I. S. (2019). Emerging Fungal Infections: New Patients, New Patterns, and New Pathogens. *Mdpi.Com*.
- Geddes-McAlister, J., & Shapiro, R. S. (2019). New pathogens, new tricks: emerging, drug-resistant fungal pathogens and future prospects for antifungal therapeutics. *Annals of the New York Academy of Sciences*, 1435(1), 57–78.
- Goddard, M. R., & Greig, D. (2015). *Saccharomyces cerevisiae*: A nomadic yeast with no niche? *FEMS Yeast Research*, 15(3), 1–6.
- Guertin, J. F., Zitouni, M., Tanguay, P., Hogue, R., & Beaulieu, C. (2019). Detection of *Delphinella* shoot blight in plantations of balsam fir (*Abies balsamea*) Christmas trees in Quebec, Canada. *Canadian Journal of Plant Pathology*, 41(1), 87–97.

- Hamby, K. A., Hernández, A., Boundy-Mills, K., & Zalom, F. G. (2012). Associations of yeasts with spotted-wing *Drosophila* (*Drosophila suzukii*; Diptera: Drosophilidae) in cherries and raspberries. *Applied and Environmental Microbiology*, 78(14), 4869–4873.
- Han, S.-M., Hyun, S.-H., Lee, H. B., Lee, H. W., Kim, H.-K., & Lee, J.-S. (2015). Isolation and identification of yeasts from wild flowers collected around Jangseong lake in Jeollanam-do, Republic of Korea, and characterization of the unrecorded yeast *Bullera coprosmaensis*. *Mycobiology*, 43(3), 266–271.
- 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: Polz M., Rajora O. (eds) *Population Genomics: Microorganisms*. Population Genomics. Springer, Cham.
- Hittinger, C. T., Steele, J. L., & Ryder, D. S. (2018). Diverse yeasts for diverse fermented beverages and foods. *Current Opinion in Biotechnology*, 49, 199–206.
- Hsieh TC, Ma KH, Chao A (2020). iNEXT: Interpolation and Extrapolation for Species Diversity. R package version 2.0.20.
- James, T. Y., Stajich, J. E., Hittinger, C. T., & Rokas, A. (2020). Toward a Fully Resolved Fungal Tree of Life. *Annual Review of Microbiology*, 74, 291–313.
- Knight, S., & Goddard, M. R. (2015). Quantifying separation and similarity in a *Saccharomyces cerevisiae* metapopulation. *ISME Journal*, 9(2), 361–370.
- Kowallik, V., & Greig, D. (2016). A systematic forest survey showing an association of *Saccharomyces paradoxus* with oak leaf litter. *Environmental Microbiology Reports*, 8(5), 833–841.
- Krause, D. J., Kominek, J., Opulente, D. A., Shen, X.-X., Zhou, X., Langdon, Q. K., ... Rokas, A. (2018). Functional and evolutionary characterization of a secondary metabolite gene cluster in budding yeasts. *Proceedings of the National Academy of Sciences*, 115(43), 11030–11035.
- Kurtzman, C., Fell, J. W., & Boekhout, T. (2011). *The yeasts: a taxonomic study*. Elsevier.
- Langdon, Q. K., Peris, D., Eizaguirre, J. I., Opulente, D. A., Buh, K. V., Sylvester, K., ... Hittinger, C. T. (2020). Postglacial migration shaped the genomic diversity and global distribution of the wild ancestor of lager-brewing hybrids. *PLoS Genetics*, 16(4).
- Leducq, J. B., Charron, G., Samani, P., Dubé, A. K., Sylvester, K., James, B., ... & Landry, C. R. (2014). Local climatic adaptation in a widespread microorganism. *Proceedings of the Royal Society B: Biological Sciences*, 281(1777), 20132472.
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., ... Koufopanou, V. (2009). Population genomics of domestic and wild yeasts. *Nature*, 458(7236), 337–341.
- Liti, G., Warringer, J., & Blomberg, A. (2017). Isolation and laboratory domestication of natural yeast strains. *Cold Spring Harbor Protocols*, 2017(8), pdb-prot089052.
- McCullough, M. J., Clemons, K. V, McCusker, J. H., & Stevens, D. A. (1998). Intergenic transcribed spacer PCR ribotyping for differentiation of *Saccharomyces species* and interspecific hybrids. *Journal of Clinical Microbiology*, 36(4), 1035–1038.

- Middelhoven, W. J., Scorzetti, G., & Fell, J. W. (2001). *Trichosporon porosum* comb. nov., an anamorphic basidiomycetous yeast inhabiting soil, related to the *loubieri/laibachii* group of species that assimilate hemicelluloses and phenolic compounds. *FEMS Yeast Research*, 1(1), 15-22.
- Nagy, L. G., Ohm, R. A., Kovács, G. M., Floudas, D., Riley, R., Gácsér, A., ... Hibbett, D. S. (2014). Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nature Communications*, 5(1), 1-8.
- Naumov, G. L. (1996). Genetic identification of biological species in the *Saccharomyces sensu stricto* complex. *Journal of Industrial Microbiology*, 17(3), 295–302.
- Naumov, G. I., Naumova, E. S., & Sniegowski, P. D. (1998). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* are associated with exudates of North American oaks. *Canadian Journal of Microbiology*, 44(11), 1045–1050.
- Nespolo, R. F., Villarroel, C. A., Oporto, C. I., Tapia, S. M., Vega-Macaya, F., Urbina, K., ... Cubillos, F. A. (2020). An Out-of-Patagonia migration explains the worldwide diversity and distribution of *Saccharomyces eubayanus* lineages. *PLoS Genetics*, 16(5), 1–23.
- O'Donnell K. (1993). *Fusarium* and its near relatives. In Reynolds, DR and Taylor, JW (Ed.), *The Fungal Holomorph: Mitotic, Meiotic and Pleomorphic Speciation in Fungal Systematics* (pp. 225–233). Wallingford: C A B International.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Solymos, P. (2019). *vegan: Community Ecology Package*. R package version 2.5-6. 2019.
- Opulente, D. A., Langdon, Q. K., Buh, K. V., Haase, M. A. B., Sylvester, K., Moriarty, R. V., ... Hittinger, C. T. (2019). Pathogenic budding yeasts isolated outside of clinical settings. *FEMS Yeast Research*, 19(3), 1–6.
- Pereira-Marques, J., Hout, A., Ferreira, R. M., Weber, M., Pinto-Ribeiro, I., Van Doorn, L. J., ... Figueiredo, C. (2019). Impact of host DNA and sequencing depth on the taxonomic resolution of whole metagenome sequencing for microbiome analysis. *Frontiers in Microbiology*, 10(Jun), 1–9.
- Peris, D., Langdon, Q. K., Moriarty, R. V., Sylvester, K., Bontrager, M., Charron, G., ... & Hittinger, C. T. (2016). Complex ancestries of lager-brewing hybrids were shaped by standing variation in the wild yeast *Saccharomyces eubayanus*. *PLoS genetics*, 12(7), e1006155.
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., ... Llored, A. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*, 556(7701), 339–344.
- 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.
- Rose MD, Winston F, H. P. M. in Y. G. (1990). *Methods in Yeast Genetics*. New York: Cold Spring Harbor.
- Roth Jr, F. J., Orpurt, P. A., & Ahearn, D. G. (1964). Occurrence and distribution of fungi in a subtropical marine environment. *Canadian Journal of Botany*, 42(4), 375–383.

- Salvadó, Z., Arroyo-López, F. N., Barrio, E., Querol, A., & Guillamón, J. M. (2011). Quantifying the individual effects of ethanol and temperature on the fitness advantage of *Saccharomyces cerevisiae*. *Food microbiology*, 28(6), 1155-1161.
- Sampaio, J. P., Golubev, W. I., Fell, J. W., Gadanho, M., & Golubev, N. W. (2004). *Curvibasidium cygneicollum* gen. nov., sp. nov. and *Curvibasidium pallidicorallinum* sp. nov., novel taxa in the microbotryomycetidae (Urediniomycetes), and their relationship with *Rhodotorula fujisanensis* and *Rhodotorula nothofagi*. *International Journal of Systematic and Evolutionary Microbiology*, 54(4), 1401–1407.
- Sláviková, E., Vadkertiová, R., & Vránová, D. (2009). Yeasts colonizing the leaves of fruit trees. *Annals of Microbiology*, 59(3), 419–424.
- Sniegowski, P. D., Dombrowski, P. G., & Fingerman, E. (2002). *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS Yeast Research*, 1(4), 299–306.
- Spagnuolo, M., Yaguchi, A., & Blenner, M. (2019). Oleaginous yeast for biofuel and oleochemical production. *Current Opinion in Biotechnology*, 57, 73–81.
- Stefanini, I., Dapporto, L., Legras, J.-L., Calabretta, A., Di Paola, M., De Filippo, C., ... Turillazzi, S. (2012). Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. *Proceedings of the National Academy of Sciences*, 109(33), 13398–13403.
- Suh, S. O., McHugh, J. V., & Blackwell, M. (2004). Expansion of the *Candida tanzawaensis* yeast clade: 16 novel *Candida* species from basidiocarp-feeding beetles. *International Journal of Systematic and Evolutionary Microbiology*, 54(6), 2409–2429.
- 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), 1–16.
- Talgø, V., Chastagner, G., Thomsen, I. M., Cech, T., Riley, K., Lange, K., ... Stensvand, A. (2010). *Sydowia polyspora* associated with current season needle necrosis (CSNN) on true fir (*Abies spp.*). *Fungal Biology*, 114(7), 545–554.
- Török, T., Mortimer, R. K., Romano, P., Suzzi, G., & Polsinelli, M. (1996). Quest for wine yeasts - An old story revisited. *Journal of Industrial Microbiology and Biotechnology*, Vol. 17, pp. 303–313.
- Tsuneda, A., Chen, M. H., & Currah, R. S. (2001). Characteristics of a disease of *Sphagnum fuscum* caused by *Scleroconidioma sphagnicola*. *Canadian Journal of Botany*, 79(10), 1217–1224.
- Vadkertiová, R., Molnárová, J., Vránová, D., & Sláviková, E. (2012). Yeasts and yeast-like organisms associated with fruits and blossoms of different fruit trees. *Canadian journal of microbiology*. 58(12), 1344-1352.
- Vaughan-Martini, A., & Martini, A. (1995). Facts, myths and legends on the prime industrial microorganism. *Journal of Industrial Microbiology*, 14(6), 514–522.

- Wang, Q., Liu, W., Liti, G., Wang, S., & Bai, F. (2012). Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular Ecology*, 21(22), 5404–5417.
- Wohlbach, D. J., Kuo, A., Sato, T. K., Potts, K. M., Salamov, A. A., LaButti, K. M., ... & Gasch, A. P. (2011). Comparative genomics of xylose-fermenting fungi for enhanced biofuel production. *Proceedings of the National Academy of Sciences*, 108(32), 13212-13217.
- 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.
- Zhang, H., Skelton, A., Gardner, R. C., & Goddard, M. R. (2010). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: evidence for migration from Europe and interspecies hybrids. *FEMS Yeast Research*, 10(7), 941–947.

Figure Legends

Figure 1) A) Permutations identified 23 yeast OTU – substrate category associations observed more often than expected ($P_{adj}<0.05$). Blue triangles indicate the number of times each combination was observed, red squares indicate the number of times the yeast OTU was observed in the permuted dataset, yellow circles indicate the expected rates of each association, and pink plus signs indicate the number of times the substrate category was observed in the permuted dataset. Associations are ordered by the frequency of the observed combinations. **B)** Permutations identified 19 yeast OTU – substrate genus associations observed more often than expected ($P_{adj}<0.05$). Blue triangles indicate the number of times each combination was observed, red squares indicate the number of times the yeast OTU was observed in the permuted dataset, yellow circles indicate the expected values of each association, and pink plus signs indicate the number of times the substrate genus was observed in the permuted dataset. Associations are ordered by the frequency of the observed combinations.

Figure 2) A) Distribution for the 1,749 isolations for which isolation temperature was known. Colors correspond to the taxonomic group of the corresponding isolate. * Basidiomycota OTUs were enriched at 4°C ($P_{adj}=1.29\times 10^{-35}$) and 10°C ($P_{adj}=1.71\times 10^{-54}$), and Saccharomycotina OTUs were enriched at 22°C ($P_{adj}=2.18\times 10^{-23}$) and 30°C ($P_{adj}=8.0\times 10^{-41}$). **B)** Permutations identified 16 yeast OTU – isolation temperature associations observed more often than expected ($P_{adj}<0.05$). Blue triangles indicate the number of times each combination was observed, red squares indicate the number of times the yeast OTU was observed in the permuted dataset, and yellow circles indicate the expected value of each association. The order of the results are by isolation temperature, followed by the number of isolations for each taxon-temperature combination.

Figure 3) A) Sample-based rarefaction curves for soil, bark, leaves, and fruit samples. Solid lines are curves based on observed data, and dashed curves are extrapolated with shaded 95% confidence intervals. **B)** Asymptotic estimates of Shannon diversity \pm standard error. **C)** Asymptotic estimates of species richness \pm standard error.

Figure 4) A) Rarefaction curves of each isolation temperature. Solid lines are curves based on observed data, and dashed curves are extrapolated with shaded 95% confidence intervals. **B)** Asymptotic estimates of Shannon diversity of all isolated yeast OTUs \pm standard error. **C)** Asymptotic estimates of Shannon diversity of Saccharomycotina OTUs (top) and Basidiomycota OTUs (bottom) \pm standard error. **D)** Heatmap of Jaccard distances for pairwise isolation temperature comparisons. Darker green reflects high dissimilarity between OTU assemblages, and lighter green reflects high similarity between OTU assemblages. Blue to red color gradient in **A-C** corresponds to increasing temperature.

Figure S1) A) Histogram of isolates in the complete dataset. **B)** When the 1,962 isolates were filtered to remove duplicate OTUs derived from the same processed sample, 1,516 unique isolations remained (bottom).

Figure S2) The distribution of isolations in the dataset by climate region. The number of unique isolations (upper, bold) and the number of unique OTUs (lower, italics) are shown. Climate regions correspond to the 9 climatically consistent regions of the contiguous US identified by NOAA (<https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php>). We included the additional category of Arctic for Alaskan samples.

Figure S3) A) Rarefaction curve for 688 unique samples (black line) with 95% confidence intervals. Solid curve based on observed data and dashed curve based on extrapolated data. **B)** Rarefaction of 688 unique samples by the higher taxonomic units Basidiomycota (yellow), Pezizomycotina (blue), and Saccharomycotina (green). Solid curve based on observed data; dashed curve based on extrapolated data with shaded 95% confidence intervals.

Figure S4) Taxonomic representation across unique isolations (light grey, top count) and unique OTUs (dark grey, bottom count). The dearth of Pezizomycotina OTUs is likely reflective of our isolation protocol, in which the selection of mold-like colonies was deliberately avoided.

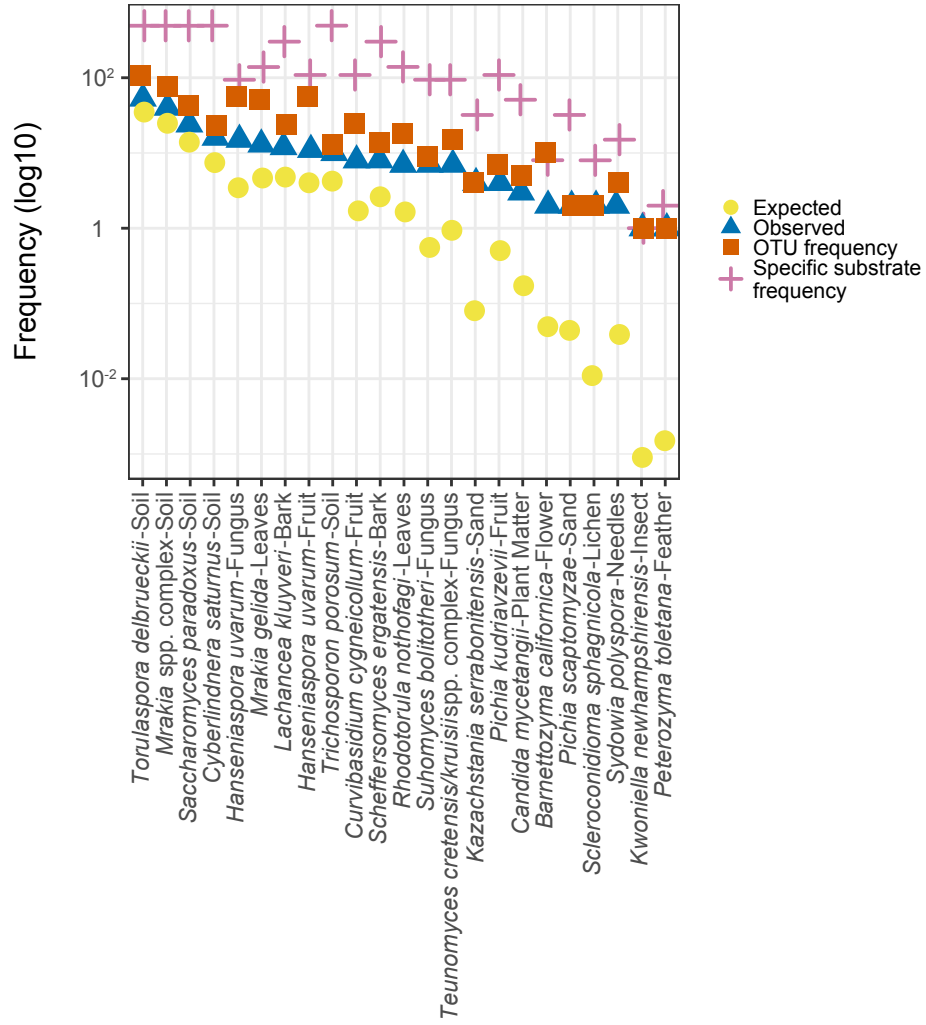
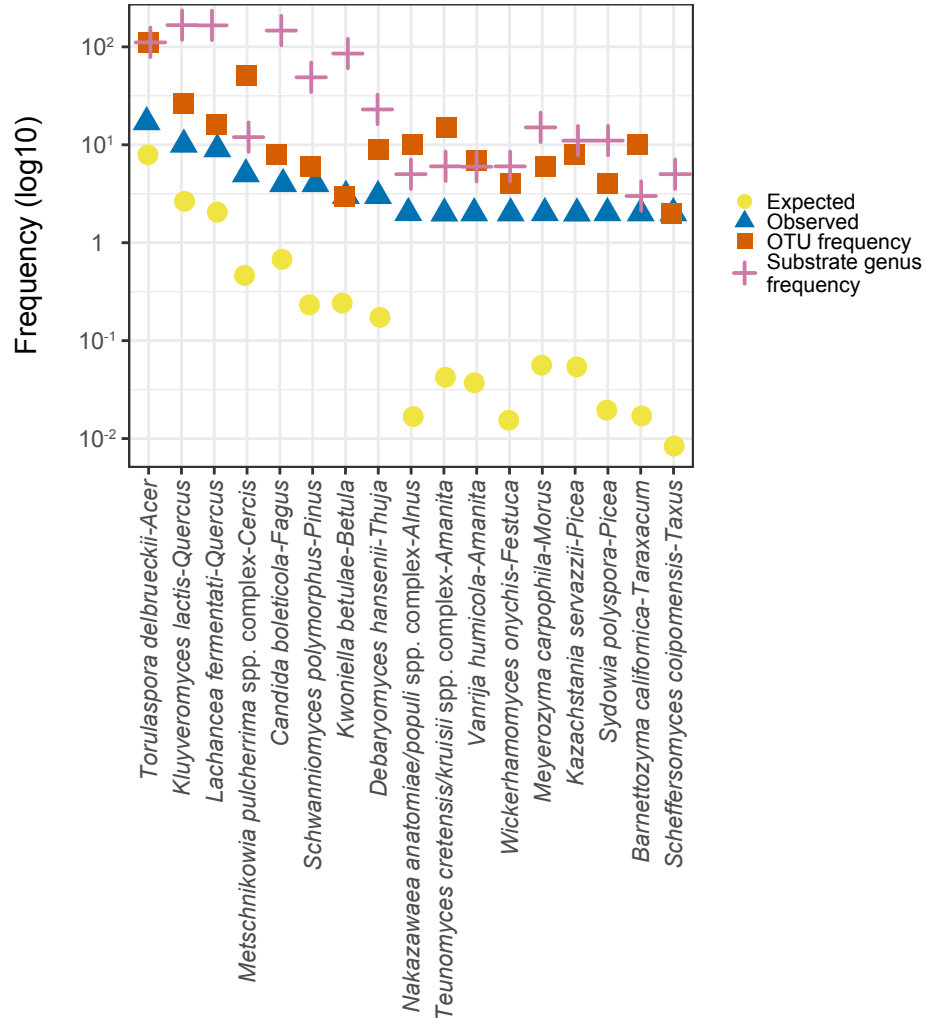
Figure S5) 103 singleton OTUs were isolated just once. **A)** Distribution of substrate categories among singletons. Singleton OTUs were not enriched for any substrate type. **B)** Distribution of singletons among higher taxonomic ranks. Singletons were not enriched for any taxonomic group. **C)** Isolation locations of singletons.

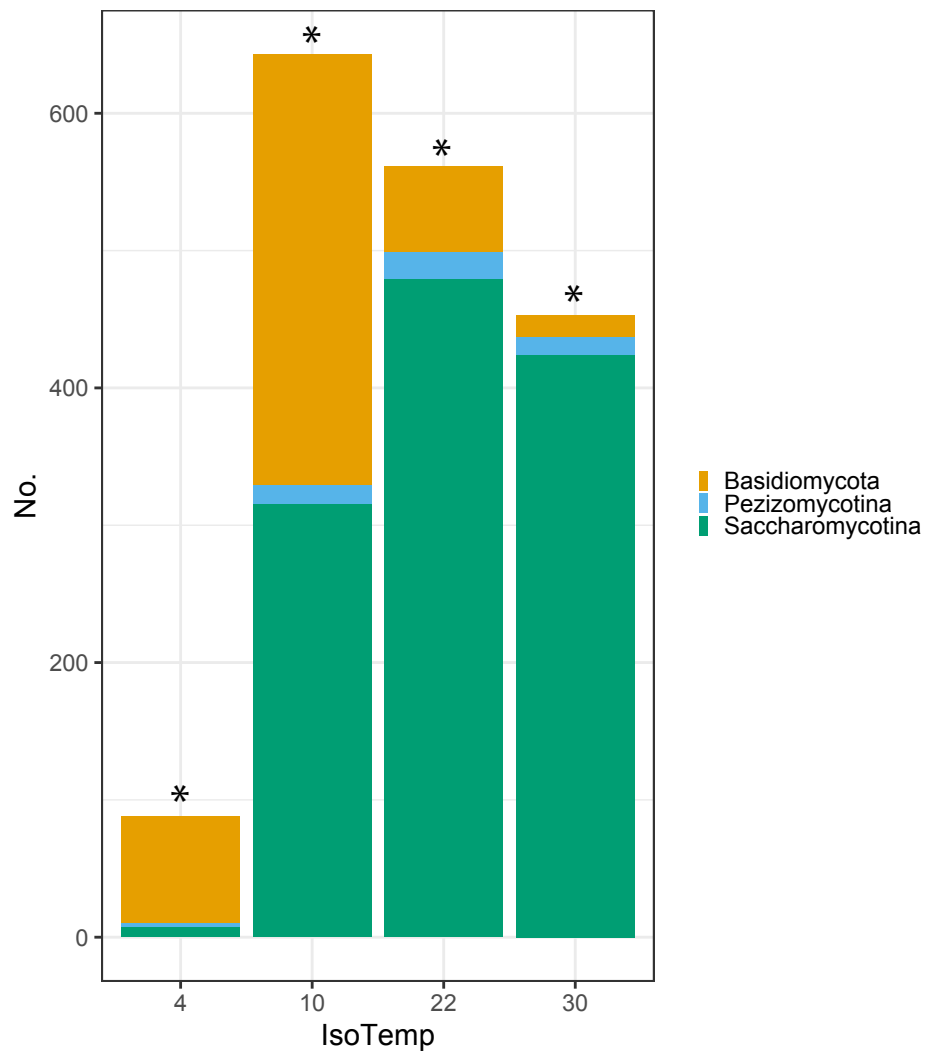
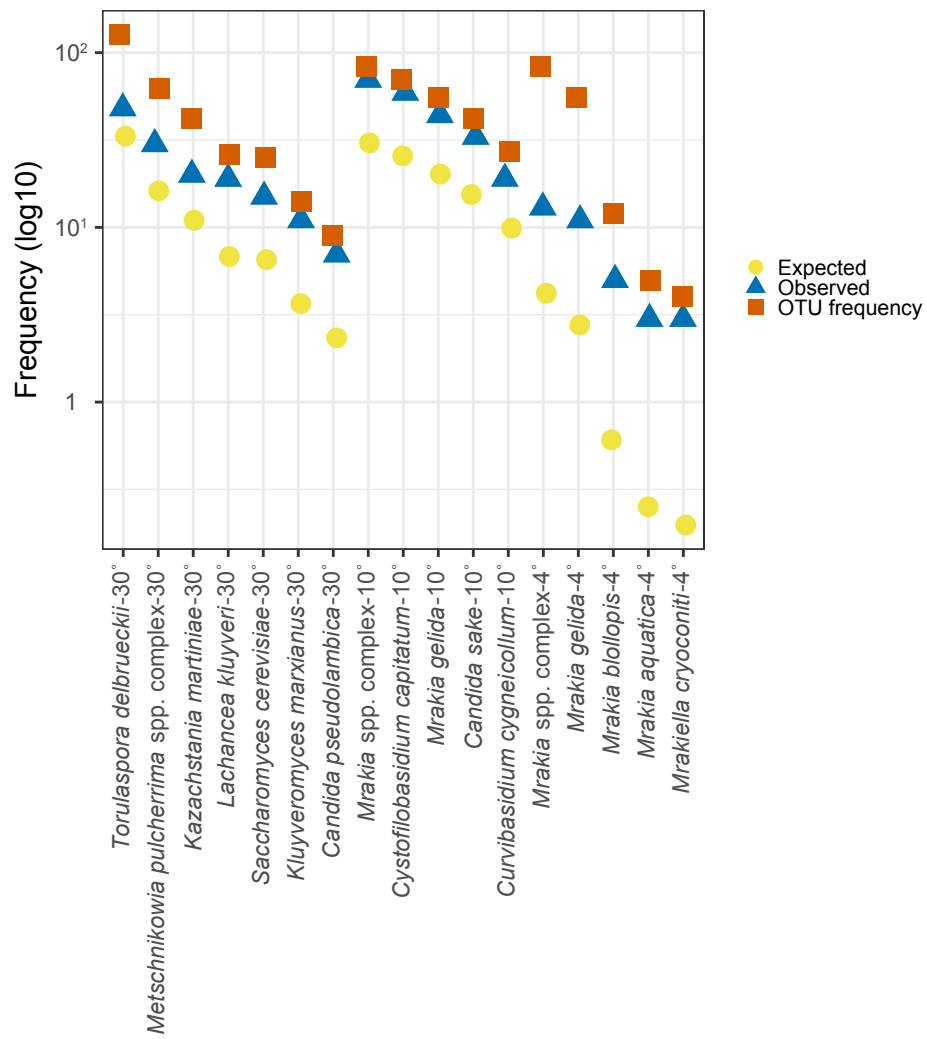
Figure S6) **A)** The number of independent isolations of each cosmopolitan OTU. **B)** The number of climatic regions in which each cosmopolitan OTU was detected.

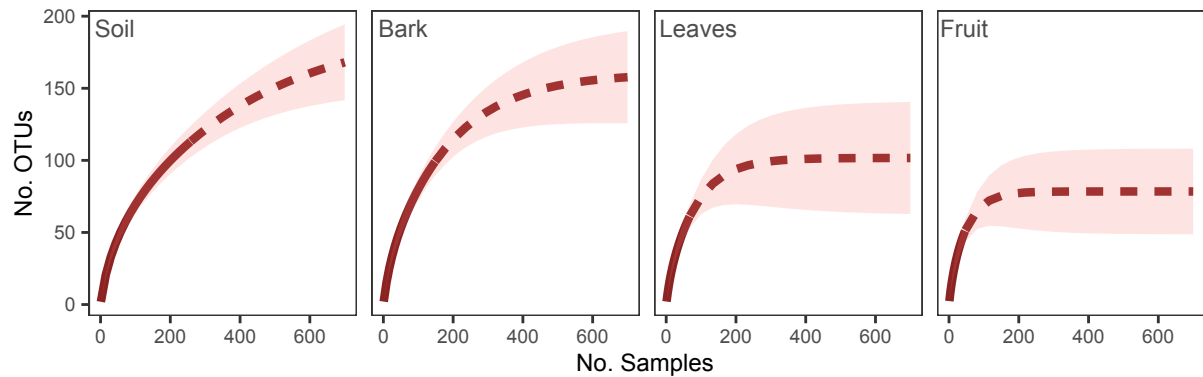
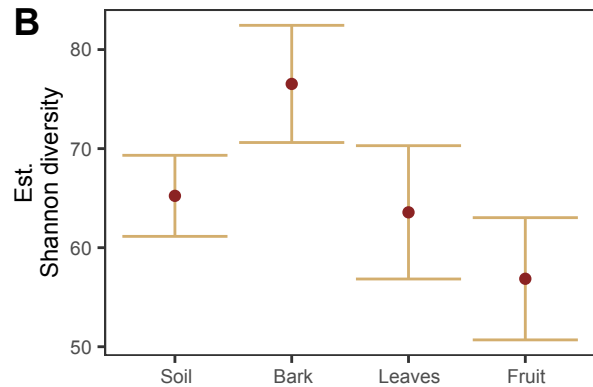
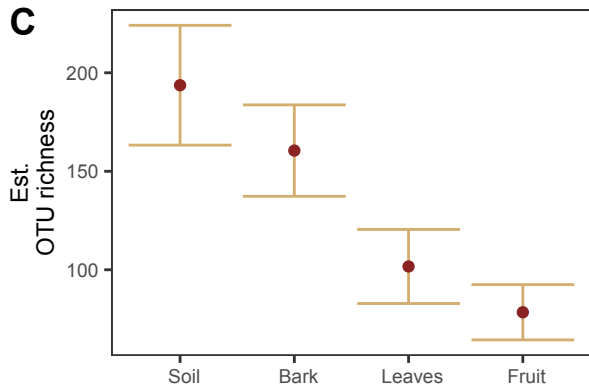
Figure S7) Cosmopolitan OTUs were defined as those OTUs that were isolated from regions where they were expected based on sampling density (allowing for one missed region). Twelve cosmopolitan OTUs were identified using this approach. **A)** Cosmopolitan yeasts were enriched in soil samples ($p= 0.003$). **B)** Cosmopolitan yeasts all belong to Saccharomycotina or Basidiomycota, but there was no enrichment for taxonomic groups among cosmopolitan OTUs. **C)** Map of cosmopolitan isolation locations. OTUs are differentiated by color.

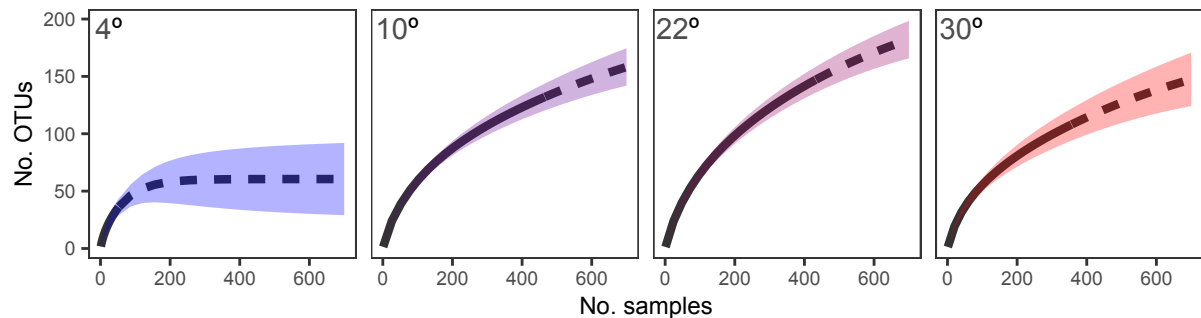
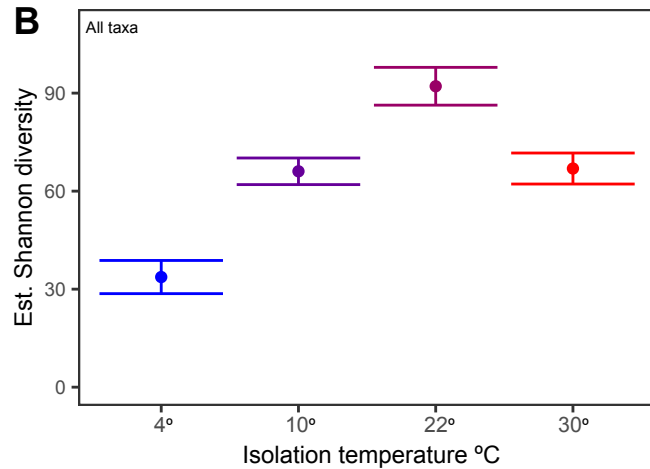
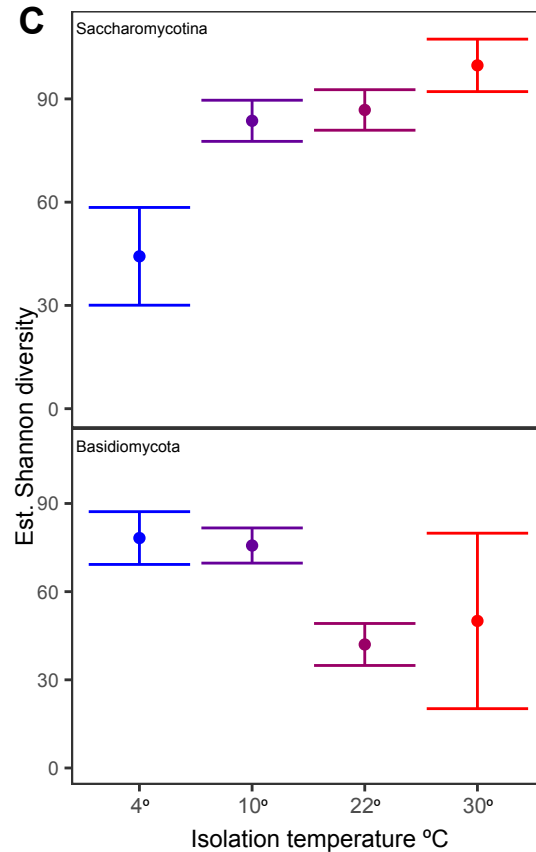
Figure S8) Forty discrete substrate categories were annotated for 1,520 isolations. **A)** Distribution of unique isolations among substrate categories. Category sampling was uneven. **B)** Substrate genera could be assigned to 1,061 isolations of diverse substrate categories. For each genus, the type of substrate collected is shown with different color bars. Substrate categories were either directly harvested from the substrate (e.g. pine needles from pine trees) or indirectly associated with the substrate (e.g. soil from the base of a pine tree).

Figure S9) **A)** Observed data fed into substrate association permutations. **B)** Observed data fed into substrate plant genus association permutations. **C)** Observed data fed into isolation temperature associations.

A**B**

A**B**

A**B****C**

A**B****C****D**