**Survey Paper Manuscript.**

***Possible titles***

* *Spatial patterns and substrate specificity across yeast taxa revealed through analysis of 1,962 natural isolates*
* *Substrate, temperature, and geographical patterns across 1,962 natural yeast isolates*
* *Collection of 1,962 wild yeast isolates reveals substrate specificity, temperature dependence, and geospatial patterns in isolation of yeasts*

**Keywords**: yeast isolation, biodiversity, yeast distribution, yeast-substrate associations

W. J. Spurley

wspurley\_at\_wisc.edu

Laboratory of Genetics

University of Wisconsin-Madison, Madison, WI 53726, USA

K. J. Fisher

kaitlinfisher19\_at\_gmail.com

Laboratory of Genetics

University of Wisconsin-Madison, Madison, WI 53726, USA

ORCID: 0000-0002-7536-0508

Middle authors – D. A. Opulente, Q. K. Langdon, *….*

C. T. Hittinger

cthittinger\_at\_wisc.edu

Laboratory of Genetics

University of Wisconsin-Madison, Madison, WI 53726, USA

ORCID: 0000-0001-5088-7461

**Abstract**

**Introduction**

Yeasts - fungi that spend at least part of their life cycle in a unicellular state - are globally distributed, speciose, diverse, and have important relationships with humans. Yeasts are polyphyletic and are represented in Ascomycota (Saccharomycotina, Pezizomycotina, and Taphrinomycotina subphyla) and Basidiomycota (Nagy et al. 2014). The phylogenetic and genomic diversity within yeasts is suggestive 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 yeast flora; typical methods for environmental detection of bacterial microbes are less viable for the more sparsely distributed eukaryotic yeasts (Periera-Marques et al. 2019). As a result, most research in yeast ecology is reliant on environmental sampling and isolation in synthetic media.

Isolation-dependent approaches are admittedly not ideal for studying ecology due to most analyses being performed away from the sample collection site, the presence of bottlenecks in the isolation process, and the fact that most fungal diversity is likely unculturable (James et al. Annu. Rev. Microbiol. 2020). However, given broad and dense enough sampling, it is nonetheless possible to begin to understand the cryptic ecology of yeasts. For example, exhaustive environmental sampling revealed that *S. cerevisiae* is not a purely domestic species once thought (Vaughan-Martini, Ann, and Alessandro Martini 1995; Török et al. 1996) and has led to a much better understanding of the global distribution of *S. cerevisiae* (Peter et al. 2018). Similar insights have been gained into the distributions of other members of the *sensu stricto* through isolation-based surveys (Charron et al. 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 learned about the natural habitats and ecology of even these well-studied 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 habitat-types (Sylvester et al. 2015; Opulente et al. 2019; additional?). Connecting yeast species to biotic or abiotic substrates is the first step in constructing species-specific niches. Hypotheses regarding the ecological roles of yeast 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 significant to humans, both those that pose threats as emerging pathogens (Friedman & Schwartz 2019; Geddes‐McAlister & Shapiro 2019, Opulente et al. 2019) and those that are potentially valuable for the development of bio-industrial capabilities (Cordente et al. 2019; ).

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 can 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 has historically been isolated mostly from soil and bark, but there have been successful isolations from many other substrates including aquatic environments, flowers, leaves, and insects (Roth Jr et al., 1964; Han et al., 2015; Sláviková et al., 2007; Stefanini et al. 2016). Coupling diversity of substrates with regional distribution of yeasts can lead to more precision in isolation.

In an effort to better understand the distribution and natural ecology of yeasts, our lab has led an ongoing initiative to sample yeast diversity from natural substrates. Previously, our lab has identified eight new species of yeasts and began to investigate preferences and associations of yeasts (Sylvester et al., 2015). This growing collection has also described associations of pathogenic yeasts and their geographic distribution in the United States (Opulente et al., 2019). Here, we used this expansive collection to curate a dataset of 1,926 isolates collected over a period of five years, the largest dataset of natural yeast isolates analyzed to date, to examine the spatial distribution patterns of diverse yeast species, look for associations between specific taxa and habitat-types, and examine overall yeast diversity amongst different types of sampled substrates. Using spatial and temporal isolation patterns we identify those yeasts exhibiting cosmopolitan distribution within our dataset. Due to the statistical power afforded by such a large dataset, we are able to confirm previously substrate associations as well as identify new ones. Finally, we examine the diversity of yeasts associated with different substrate types to better understand where yeast communities are expected to be complex.

**METHODS**

***Sample Collection and Isolation***

Samples of natural substrates were collected across the continental United States as well as Alaska. Data recording the GPS location, substrate types, and, if applicable, plant species accompanied all collections. Samples were collected using a sterile bag and without any human contact in order 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 detail in Sylvester et al., 2015. Briefly, samples were inoculated into 9ml of Wild Yeast Medium (see Sylvester et al. 2015) in a 15 ml conical tube and incubated at 4C, 10C, 22C, or 30C until there was evidence of fermentation (1-2 days at warm temperatures, 1-2 weeks at cold temperatures). Cultures were then passaged once more through 4 ml Wild Yeast Media and incubated at respective temperatures until growth was evident. Enriched cultures were plated to YPD agar plates. Single colonies were selected for identification, restruck to purify, and frozen down in glycerol stocks.

Taxonomy was assigned through amplification and sequencing of the *ITS* region. *ITS2* sequences were queried against the nucleotide database of BLAST and query sequences that matched a database hit by 97% or more were positively identified based on that search. Queries that did not return matches of at least 97% were resequenced at the D1/D2 domain of the *RDN25* region to achieve a positive identification. No novel species are described in the dataset of isolates in this analysis. Some taxa can only be identified at the level of species complex, and therefore we use the term operational taxonomic unit in lieu of species throughout this analysis.

***Dataset curation***

*Need Dana to write a few sentences about why/how dataset was curated.*

Oftentimes, multiple isolates of the same OTU isolated from the same exact substrate sample appear in the dataset. To eliminate false positive associations due to this, prior to each analysis below identical OTUs isolated from the same processed sample were eliminated. Exclusion of identical OTUs isolated from the same processed sample results in 1,518 unique isolations (**Fig. S1**).

***Detection of Cosmopolitan Taxa***

We sought to use our extensive dataset of isolation locations to determine which taxa in our data could be considered cosmopolitan across the expansive sampled region. Instead of simply determining the taxa that were isolated temporally or spatially frequently, we considered differences in sampling depth across the US to identify those taxa that were always isolated in regions that were sufficiently sampled to detect them. We began by dividing the country into the ten climatic regions described by the National Oceanic and Atmospheric Association (NOAA)(see **Fig. S3**). For each taxonomic category we then calculated an expected isolation rate based on the most densely sampled climatic region in which that taxa was found. For example, *Torulaspora delbrueckii* was isolated 60 independent times out of 382 samples collected in the Upper Midwest, giving it an isolation rate of once every 6.3 samples. 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 each taxon is expected to have been isolated from and compared these to the regions each taxon was actually found in. We eliminated from consideration all yeasts that had only been isolated only once (116 taxa) and all those yeasts that were only expected in a single region (80 taxa, all expected only in the Upper Midwest). We identified cosmopolitan yeasts by examining the remaining data for taxa that were isolated either in all regions where expected or all but one region where expected. For example, *T. delbrueckii* was expected in all regions except the Northern Rockies and Southwest and was found in all expected regions with the exception of the Arctic. Because we allowed for one instance of not being detected when expected, *T. delbrueckii* is categorized as a cosmopolitan yeast taxon.

***Association analyses***

Enrichments of yeast subphyla in substrate-type categories and isolation temperatures were examined with a one-tailed Fisher’s exact test followed by a Benjamini Hochberg post-hoc correction. Positive associations at the level of individual yeast taxa were detected by independently permuting yeast-substrate, yeast-substrate phylum, or yeast-temperature combinations to find those combinations present more often than expected by chance. We did not examine our data for depleted combinations as these would likely simply be artifacts of sampling. P-values and expected rates were calculated based on 10,000 permutations for each association analysis. A Benjamini-Hochberg adjustment was applied to P-values. Permutations were performed by completely re-sampling yeast taxa assigned to each substrate or temperature category without replacement using custom R scripts.

***Diversity Analyses***

We used H’ indices to compare isolate diversity amongst substrate types and isolation temperatures using the R vegan package (Oksanen et al. 2019). The Shannon-Wiener (H’) index, which quantifies the uncertainty of predicting the identity of an individual from a given sample by considering both the species richness and species abundance, is a frequently used method of measuring taxonomic diversity (Tuomisto 2012). To remove the effect of phylum on isolation temperature estimates, we further compared diversity amongst isolation temperatures independently in each subphylum. Like all diversity metrics, comparison of H’ indices are complicated by uneven sampling. We used a linear regression to show that sampling density does affect H’ estimates (*p*= 7.757e-05)*.* However, because sampling density alone only explains a small part of H’ variance (*R2*adj=0.323), we reasoned that H’ indices could still provide useful comparisons of diversity, particularly amongst those substrates that are sampled at similar depths.

**Results & Discussion**

***Curated dataset of isolates (Needs review by Dana)***

The final dataset curated for this study consisted of 1,962 isolations of 262 unique operational taxonomic units (OTUs) (Suppl. Table 1). 475 strains in this study were also included in a previously published survey of wild yeast isolates (Sylvester et al. 2015, Suppl. Table 1). 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 *ITS2* sequence used to identify the species. Represented in the data are yeasts from both major divisions of Dikarya, Basidiomycota (90 OTUs) and Ascomycota (172 OTUs). Within Ascomycota, isolates belong to the subphyla Saccharomycotina (153 OTUs), Pezizomycotina (13 OTUs), and Taphrinomycotina (1 OTU) (**Fig. S2**).

***Geospatial variation among isolations***

***Continental United States is unevenly sampled***

Isolates were obtained from all major climate regions of the contiguous United States with the exception of the Southwest, as well as Alaska (**Fig. S3**). The Upper Midwest was the most densely sampled region with 868 unique isolations from 382 samples, followed by 144 isolations from 53 samples in the Ohio Valley, 140 isolations from 58 samples in the Northeast, 136 isolations from 87 samples in the Northwest, 105 isolations from 55 samples in the Southeast, 40 isolations from 14 samples in the West, 32 isolations from 16 samples in the South, and 6 isolations from 3 samples in the Northern Rockies. Zero isolates were collected from the Southwest. An additional 44 isolations came from 31 Alaskan samples.

***116 singleton isolations and 16 taxa isolated over 20 times***

Of the 262 unique OTUs, 116 are singletons that were found in only one isolation (**Suppl. Table 2**), while the remaining 146 generally have fewer than thirty isolations. The singleton taxa isolated in this study were primarily isolated from bark and soil, however singletons were not enriched for any substrate category (**Fig. S4A**). Singletons were, however, enriched for yeasts belonging to the Basidiomycota (*P*adj=0.003) and Pezizomycotina (*P*adj= 0.0005) subphyla (**Fig. S4B**), reflecting an isolation protocol designed to capture Saccharomycotina species. Singletons appear evenly distributed across sampled regions (**Fig. S4C**).

Sixteen yeast taxa were isolated over 20 independent times in this dataset (**Suppl. Table 3)**. Amongst these are two *Saccharomyces* species, *S. paradoxus* and *S. cerevisiae,* for which considerable population genetic work has already been done (Xia et al. 2017; Peter et al. 2018, Hénault et al. 2017). Other frequently isolated taxa are of broad interest but have been subject to relatively little population genetic analysis. These include *Kluyveromyces lactis*, *Hanseniaspora uvarum*, and the *Metschnikowia pulcherrima* sp. complex. *Do we have anything published that describes the intraspecific phenotypic heterogeneity we find amongst isolates?* The collection described herein is teeming with potential for population-level studies for these and more taxa.

***11 cosmopolitan isolates***

Although many yeast species are frequently referred to as cosmopolitan, there is no clear criteria by which yeasts are or are not classified as such. We examined our dataset for yeasts that are widely distributed across the United States by considering both the number and spatial distribution of isolations for each taxa. Briefly, we defined a OTU-specific isolation rate and then looked for taxa that are generally able to be isolated in all regions where sampling is dense enough to detect them (**Suppl. Table 4**, see Methods). While this approach cannot exclude yeasts from being broadly distributed, it can identify those taxa in our dataset that are most likely cosmopolitan in distribution. We determined that eleven OTUs in our dataset are generally found when expected across climatic regions (*Candida railensis, Candida sake, Cryptococcus flavescens, Cyberlindnera saturnus, Debaryomyces hansenii, Leucosporidium scottii, Rhodotorula fujisanensis, Saccharomyces paradoxus, Scheffersomyces ergatensis, Torulaspora delbrueckii,* and *Wickerhamomyces anomalus*, **Suppl. Table 5**). Only 5 of these taxa were also among the 16 OTUs isolated over 20 times - indicating that frequent isolation in individual regions may not indicate otherwise broad geographic distributions. Three of the taxa we identified as cosmopolitan were isolated less than 10 times (*Debaryomyces hansenii*, *Leucosporidium scottii*, *Rhodotorula fujisanensis*), but were nonetheless always detected in regions with sufficient sampling (**Fig. S5A**, **Fig. S6C**). Cosmopolitan taxa range from being isolated in just two regions (due to requiring dense sampling to detect) to being isolated in a maximum of 7 regions (**Fig. S5B**, **Fig. S6C**). No yeast taxon was isolated from all nine sampled US regions. Cosmopolitan yeasts were not isolated from equal rates amongst substrates, but rather are significantly enriched for soil substrate samples (*P*adj==3.09E-05, **Fig. S6A**).

Our approach to detecting cosmopolitan isolates is imperfect and affected by sampling biases in our data. The approach also only considers total climatic region sampling density and not the sampling densities of any specific substrates. Most substrates are not uniformly sampled across regions and yeasts with high substrate-specificities are likely to be missed. Unfortunately, powerful approaches to determining distributions for microbial fungal taxa remain elusive. They are 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’s possible that many more yeast taxa in our collection have wide distributions, we can say for certain that the eleven taxa we identify as cosmopolitan show patterns of ubiquitous distributions. The strong statistical enrichment for cosmopolitan yeasts among soil samples may implicate soil as a reservoir of ubiquitous yeast taxa.

***Substrate and temperature associations***

***Substrate – subphylum associations***

We were able to assign one of 40 specific substrate categories to 1,522 isolations comprised of 161 unique yeast OTUs (**Fig. S7**). Contingency tables were used to identify associations between yeast subphyla and substrate. When subphylum representation across substrate types was examined (**Suppl. Table 6**), we found that leaves are enriched for Basidiomycota yeasts (*P*adj=0.017) but found no other significant enrichment of subphyla across substrate types.

***Specific yeast - 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 715 observed combinations (**Fig. S8**), we found 22 yeast OTU by substrate associations that occur more frequently than expected by chance (*P*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 taxonomic units (*Mrakia* sp*.*, *Trichosporon porosum, Torulaspora delbrueckii*, *Cyberlindnera saturnus*, and *Saccharomyces paradoxus*, *P*adj= 0.01, 0.046, 0.016, 0.024, 0.041). Bark was also heavily sampled but yielded only two significant associated taxa, both of which are Saccharomycotina (*Lachancea kluyveri* & *Scheffersomyces ergatensis*, *P*adj= 0.029, 0.032). Leaf samples were also associated with two taxa, both of which are Basidiomycota yeasts (*Rhodotorula nothofagi* & *Mrakia gelida*, *P*ajd = 0.024, 0.032). Fungal samples were significantly associated with three Saccharomycotina yeast taxa (*Hanseniaspora uvarum, Suhomyces bolitotheri*, *P*adj<0.0001, & *Teunomyces cretensis/kruisii complex*, *P*adj= 0.010); fruit with one Basidiomycota and one Saccharomycotina taxon (*Curvibasidium cygneicollum* & *Pichia kudriavzevii*, *P*ajd= 0.016, 0.024); sand with two Saccharomycotina taxa (*Kazachstania serrabonitensis*, *P*adj<0.0001& *Pichia scaptomyzae*, 0.024). Feathers, flowers, insects, lichens, and needles were associated with a single taxon each (*Peterozyma toletana*, *P*adj=0.024; *Zygowilliopsis californica*, *P*adj<0.0001; *Kwoniella newhampshirensis*, *P*adj=0.027; *Scleroconidioma sphagnicola*, Padj<0.0001; *Sydowia polyspora*, P*adj=*0.032). Plant matter samples (excluding matter that falls into other plant-related categories) were associated with a single taxon (*Candida mycetangii*, *P*adj = 0.024).

When possible, substrate genera were identified for plant and fungal substrates directly sampled (**Fig. S7)**. Substrate genera were also assigned samples indirectly associated with plant genera (e.g. soil sampled from base of tree). In total, 66 substrate genera were identified across 1,026 isolations comprising 209 OTUs (**Fig.** **S8**, **Suppl. Table 8**). Nineteen of 657 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. Three Saccharomycotina yeasts were significantly associated with Quercus spp., (*Lachancea fermentati*, *Kluyveromyces lactis*, and *Lachancea thermotolerans, P*adj< 0.0001, *P*adj = 0.008, 0.035). Notably, none of the Quercus-associated taxa belong to the genus *Saccharomyces*, although all three do belong to the Saccharomycetaceae, the same family as *Saccharomyces*. Spruce trees of the genus *Picea* were associated with the Pezizomycotina yeast *Sydowia polyspora* (*P*adj*=*0.008) and the Saccharomycotina yeast *Kazachstania servazzii* (*P*adj=0.035). *Ceris*, a genus of large flowering shrubs, is associated with the Saccharomycotina *Metschnikowia pulcherrima* sp. complex (*P*adj<0.0001) and *Lachancea kluyveri* (*P*adj=0.048). *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 yeast *Vanrija humicola* (*P*adj= 0.03), and the Saccharomycotina *Teunomyces cretensis/kruisii* sp. complex (*P*adj=0.035). Ten additional plant genera were associated with a single taxon each (*Vaccinium* – berry-producing shrubs, *Curvibasidium cygneicollum*, *P*adj =0.035; *Thuja* - cypress trees, *Debaryomyces hansenii, P*adj=0.033; *Taxus* - yew trees, *Candida coipomoensis*, *P*adj<0.0001; *Taraxacum* – dandelions, *Zygowilliopsis californica*, *P*adj<0.001; *Rubus* – berry-producing bushes, *Hanseniaspora uvarum*, *P*adj=0.048; *Pinus* – pine trees, *Schwanniomyces polymorphus*, *P*adj<0.0001; *Morus* – mulberries, *Papiliotrema flavescens*, *P*adj=0.035; *Festuca* - perennial tufted grasses, *Wickerhamomyces onychis*, *P*adj= *0*.015; *Betula* - birch trees, *Kwoniella betulae*, *P*adj=0.018; *Alnus* - alder trees, *Nakazawaea anatomiae/populi* sp. complex *P*adj<0.0001).

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 *Hanseniaspora uvarum* and the *Rubus* genus of fruit-bearing bushes would have been missed at the substrate level, as 3 of the five independent *H. uvarum* isolations came from fruit and 2 came from leaves. Nonetheless we do find corroboration between the two analyses. This is most obvious among the four OTUs are that are associated with a biotic substrate and also with a corresponding genus into which that substrate falls; *Zygowilliopsis californica* (*Taraxcum* and flower)*, Sydowia polyspora* (*Picea* and needles)*, Lachancea kluyveri* (*Cercis* and bark)*,* and *Teunomyces cretensis/kruisii* complex (*Amanita* and fungus). These analyses were performed independently on different datasets (**Fig. S8**), and so these corroborative findings lend high confidence in these associations. Alternatively, in the case of *Hanseniaspora uvarum*, significant associations at the substrate and substrate genus level are seemingly unrelated. As mentioned, *H. uvarum* is associated with the *Rubus* plants at the genus level, but also with fungi at the substrate level. Thus *H. uvarum* potentially has a broad niche that extends beyond its well-known role in rotting fruit fermentation (Spencer et al. 1992, Albertin et al. 2016).

We do not find associations between *Saccharomyces* speciesand 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**).Further, an absence of associations cannot be explained by lack of sampling, as *Quercus* was the most deeply sampled genus in our dataset (**Fig. S8**). The absence of associations are noteworthy as oak trees have long been thought of as the natural habitat of *Saccharomyces* species, particularly *S. cerevisiae* and *S*. *paradoxus* (Sneigowski et al. 2002; Zhang et al. 2010; Wang et al. 2012; Kowallik & Greig 2016). In fact, a previous study from our lab using a much less extensive version of these data did find an association between *Saccharomyces* and *Quercus* species (Sylvester et al. 2015). Because of the extent of literature focused on isolating *Saccharomyces* from oak trees, this assumption has percolated effectively into the broader yeast literature. While it is obvious from our data and from others’ 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* and *S. paradoxus* can be found on a wide array of substrates and oak trees are unlikely to be the sole major reservoir of wild *Saccharomyces* yeasts.

Studies wielding power to statistically associate yeast taxa 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, we do find that many of the significant substrate associations we find in these data confirm previously reported associations based on fewer isolations. A novel species previously described by our laboratory, *Kwoniella betulae*, was named for the birch genus from which it was isolated (Sylvester et al. 2015). The presently analyzed data contains the same isolates presented in the original description paper and further identifies a statistical association with the *Betula* genus. *Sydowia polyspora* is a known conifer pathogen and we find it statistically associated with the pine genus *Picea* as well as with needles (Talgo et al. 2010; Guertin et al. 2018). The opportunistic human pathogen *Pichia kudriavzevii* has been repeatedly associated with fruit before (Kurtzman, Fell and Boekhout 2011; Douglass et al. 2018; Opulente et al. 2019). Additional previously described associations include *Torulaspora* *delbrueckii* and soil (Kurtzman, Fell and Boekhout 2011), the xylose-consuming yeast *Scheffersomyces ergatensis* and bark (Kurtzman, Fell and Boekhout 2011), and *Trichosporon* *porosum* and soil (Middlehoven et al. 2001). We also find logical indirect connections between the associations we find and previously reported isolation substrates. For example, *Suhomyces bolitotheri* and *Teunomyces cretensis/kruisii* were originally isolated from the guts of basidiocarp-feeding beetles (Suh, McHugh and Blackwell 2004) and we find them both associated with fungi. *Scleroconidioma sphagnicola* is considered a moss pathogen (Tsuneda, Chen and Currah 2001) which we find 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 are novel. *Hanseniaspora uvarum* has not been previously described as being a common flora of fungi, yet we find a robust association between the two. *Curvibasidium cygneicollum* has been isolated from diverse stubstrates across the globe (Sampaio et al. 2004), but to our knowledge it has not been associated with fruit before.

***Isolation temperature – subphylum associations***

Isolations in this dataset were performed at one of four temperatures: 4°C, 10°C, 22°C, or 30°C. On occasion the same environmental sample was subjected to multiple different temperatures for isolation, and these were considered separate isolations for temperature analyses. In total, isolation temperature was confidently assigned for 1,750 isolations. We have previously shown that isolation temperature drastically affects subphylum representation amongst isolates (Sylvester et al. 2015), and we find the same trends repeated in the presently described data. Significant enrichment for Basidiomycota taxa is observed at cooler isolation temperatures of 4°C (*P*adj=1.75x10-35) and 10°C (*P*adj=1.00x10-54, **Fig.** **2A**, **Suppl. Table 10**). Similarly, an enrichment for Saccharomycotina spp. is observed at 22°C (*P*adj=2.45x10-23) and 30°C(*P*adj =8.90x10-41).

***Specific yeast – isolation temperature associations***

To drill down on the specific yeast taxa that drive isolation temperature difference between subphyla, we examined our data for specific taxa – isolation temperature associations. Out of 443 yeast taxa – isolation temperature combinations (**Fig. S8**), 16 significant positive associations were found (**Fig. 2B**, **Suppl. Table 11**). As expected, all OTUs associated with isolation at 4°C are Basidiomycota yeasts (*Mrakia blollopis*, *Mrakia gelida*, *Mrakiella cryoconite*, *Mrakia* sp., and *Cystofilobasidium capitatum.*; *P*adj= 0.004, 0.004, 0.008, 0.017, 0.05). Five taxa were associated with 10°C isolation, four of which are Basidiomycota yeasts (*Cystofilobasidium capitatum*, *Mrakia gelida*, *Mrakia* sp.; *P*adj<0.0001 and *Rhodotorula fujisanensis*; *P*adj=0.022), and one of which is a Saccharomycotina OTU (*Candida sake, P*adj<0.0001). There were no significant associations at 22°C. Six total taxa were found to be significantly associated with 30°C isolation, five of which are Saccharomycotina (*Kluyveromyces marxianus*, *Lachancea kluyveri*, *Metschnikowia pulcherrima* sp. complex; *P*adj<0.0001, *Saccharomyces cerevisiae*, *Candida pseudolambica*; *P*adj=0.017, 0.05). The remaining 30°C isolation– associated yeast is a Pezizomycotina taxa (*Lecythophora* sp.; *P*adj<0.0001).

We have now repeatedly found subphylum differences in isolation temperature, with Basidiomycota dominating cool isolation temperatures and Saccharomycotina dominating warm isolation temperatures (Sylvester et al. 2015). Here we find that subphylum effects emerge at the level of individual taxon – temperature associations as well. Eight of the nine significant associations with cold temperatures belong to Basidiomycota. Conversely, five of the six yeasts associated with a warm isolation temperature belong to Saccharomycotina and none to Basidiomycota. One possible exception to the subphyla divisions amongst isolation temperatures is *Candida 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).

Germination, growth rates, and microbial competition are but a few of the processes likely affect species-specific isolation success at different temperatures. The strong phylogenetic affects we find suggest that the intrinsic biological properties drive most of the variation in isolation success. In confirmation of our power to detect biologically real temperature preferences, we recover *S. cerevisiae* as a 30°C-associated taxa. These differences are useful in both understanding the differing biology and ecology of these groups, but also in informing isolation approaches for different species.

***Substrate and temperature diversity***

***Variation in diversity by substrate sampled***

Maximizing taxonomic diversity recovered when sampling would be advantageous for yeast ecologists. To determine which substrates yield the highest diversity of yeasts, we estimated relative diversity of each substrate using the Shannon-Wiener index (**Fig. 3A**, **Suppl. Table 12**). To make sure our estimates of diversity are not overly influenced by sampling bias, we used a linear regression to show that, while H’ indices are affected by sampling density (*p*= 7.757e-05)*,* sampling density alone only explains a small part of H’ variance (*R2*adj=0.323). Still, the substrates with the highest H’ were bark and soil, the most densely sampled substrates. Useful comparison can be made among substrates with similar sampling densities, nonetheless. For example, moss and sand are sampled at similar depth to nuts and wood but have much higher diversity indices. Some diversity differences were surprising, e.g. lichens have a much higher diversity of yeasts than mold.

***Variation in diversity by isolation temperature***

Diversity deriving from sample inoculations may also depend on incubation temperatures. We used a similar approach to examine the diversity among each isolation temperature. We found 22°C to have the highest level of diversity with an index value of 4.25 and 4°C to have the lowest level of diversity with an index value of 3.21 (**Fig. 3B, Suppl. Table 13**). Because isolation temperature has a drastic effect on the subphyla of resultant isolates, we used the same approach to analyze diversity individually within each subphylum. Among OTUs within Saccharomycotina, H’ indices were similarly high at 30°C, 22°C, and 10°C, but dropped off sharply at 4°C. The reciprocal pattern was seen in Basidiomycota, which exhibited a sharp drop off at 30°C. Pezizomycotina seems to have the narrowest range for capturing maximum diversity at 22°C (**Fig. S9**), and this may simply reflect our protocol’s intentional avoidance of hyphal-appearing colonies, a morphology common in Pezizomycotina. Even when considering subphyla separately maximum diversity is always achieved at 22°C. This, coupled with the finding that no individual taxonomic unit is preferentially isolated at 22°C, would indicate that 22°C is the ideal incubation temperature for generally querying the yeast diversity present in a given sample.

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

Final paragraph to be written last, along with abstract.

**Acknowledgements**

**Competing interests**

We routinely short oak exudate futures and will make millions of dollars.

**Data availability**

Code can be archived (this may or may not be necessary).

All data is provided in the form of supplemental tables.

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

Suppl. Table 2 – Metadata for isolates of taxa isolated only once.

Suppl. Table 3 – Metadata for isolates of taxa isolated greater than 20 times.

Suppl. Table 4 – Discovery rates and regional expectations for 27 OTUs.

Suppl. Table 5 – Metadata for isolates of cosmopolitan taxa.

Suppl. Table 6 – Contingency table for subphyla representation across substrate types.

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

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

Suppl. Table 9 – Metadata for all isolates of *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*.

Suppl. Table 10 – Contingency table for subphyla representation across isolation temperatures.

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

Suppl. Table 12 – Shannon-Weiner indices for 40 substrate types.

Suppl. Table 13 – Shannon-Weiner indices for 4 isolation temperatures.

**References**

Albertin, W., Setati, M. E., Miot-Sertier, C., Mostert, T. T., Colonna-Ceccaldi, B., Coulon, J., … Masneuf-Pomarede, I. (2016). Hanseniaspora uvarum from winemaking environments show spatial and temporal genetic clustering. *Frontiers in Microbiology*, *6*(JAN), 1569. https://doi.org/10.3389/fmicb.2015.01569

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. https://doi.org/10.1016/j.ijfoodmicro.2017.01.009

Charron, G., Leducq, J., Bertin, C., … A. D.-F. yeast, & 2014, undefined. (n.d.). Exploring the northern limit of the distribution of Saccharomyces cerevisiae and Saccharomyces paradoxus in North America. *Academic.Oup.Com*. Retrieved from https://academic.oup.com/femsyr/article-abstract/14/2/281/532657

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. https://doi.org/10.1007/s00253-019-09840-w

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. https://doi.org/10.1038/nrmicro3451

Friedman, D. Z. P., & Schwartz, I. S. (2019). Emerging Fungal Infections: New Patients, New Patterns, and New Pathogens. *Mdpi.Com*. https://doi.org/10.3390/jof5030067

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. https://doi.org/10.1111/nyas.13739

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. https://doi.org/10.1080/07060661.2018.1547791

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*. https://doi.org/10.1007/13836\_2017\_4

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. https://doi.org/10.1146/annurev-micro-022020-051835

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. https://doi.org/10.1111/1758-2229.12446

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., … Todd Hittinger, C. (2020). Postglacial migration shaped the genomic diversity and global distribution of the wild ancestor of lager-brewing hybrids. *PLoS Genetics*, *16*(4). https://doi.org/10.1371/journal.pgen.1008680

Nagy, L. G., Ohm, R. A., Kovács, G. M., Floudas, D., Riley, R., Gácser, A., … Hibbett, D. S. (2014). Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms5471

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. https://doi.org/10.1371/journal.pgen.1008777

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. https://doi.org/10.1093/femsyr/foz032

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. https://doi.org/10.3389/fmicb.2019.01277

Peter, J., Chiara, M. De, Friedrich, A., Yue, J., Nature, D. P.-, & 2018, undefined. (n.d.). Genome evolution across 1,011 Saccharomyces cerevisiae isolates. *Nature.Com*. Retrieved from https://www.nature.com/articles/s41586-018-0030-5

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.

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. https://doi.org/10.1099/ijs.0.03037-0

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.

Spencer, D. M., Spencer, J. F. T., De Figueroa, L., & Heluane, H. (1992). Yeasts associated with rotting citrus fruits in Tucumán, Argentina. *Mycological Research*, *96*(10), 891–892. https://doi.org/10.1016/S0953-7562(09)81038-1

Stefanini, I., Dapporto, L., Berń, L., Polsinelli, M., Turillazzi, S., & Cavalieri, D. (2016). Social wasps are a saccharomyces mating nest. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1516453113

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. https://doi.org/10.1099/ijs.0.63246-0

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. https://doi.org/10.1093/femsyr/fov002

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. https://doi.org/10.1007/bf01574705

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. https://doi.org/10.1139/b01-102

Tuomisto, H. (2012). An updated consumer’s guide to evenness and related indices. *Oikos*, *121*(8), 1203–1218. https://doi.org/10.1111/j.1600-0706.2011.19897.x

Vaughan-Martini, A., & Martini, A. (1995). Facts, myths and legends on the prime industrial microorganism. *Journal of Industrial Microbiology*, *14*(6), 514–522. https://doi.org/10.1007/BF01573967

Wang, Q., Liu, W., Liti, G., Wang, S., & Bai, F. (2012). Surprisingly diverged populations of s accharomyces cerevisiae in natural environments remote from human activity. *Molecular Ecology*, *21*(22), 5404–5417.

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

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 22 yeast taxa – substrate category associations observed more often than expected (*P*adj<0.05). Yellow points indicate the number of times each combination was observed, red points indicate the number of times the yeast taxa was observed in the permuted dataset, blue points indicate the expected rates of each association, and green points indicate the number of times the substrate category was observed in the permuted dataset. **B)** Permutations identified nineteen yeast taxa – substrate genus associations observed more often than expected (*P*adj<0.05). Yellow points indicate the number of times each combination was observed, red points indicate the number of times the yeast taxa was observed in the permuted dataset, blue points indicate the expected values of each association, and green points indicate the number of times the substrate genus was observed in the permuted dataset.

**Figure 2) A)** Distribution of isolations for the 1,750 isolations for which isolation temperature is known. Colors correspond to the subphylum of the corresponding isolate. \* Basidiomycetes are enriched 4°C (*P*adj=1.75x10-35) and 10°C (*P*adj=1.00x10-54) and Saccharomycotina spp. are enriched 22°C (*P*adj=2.45x10-23) and 30°C(*P*adj =8.90x10-41). **B)** Permutations identified 16 yeast taxa – isolation temperature associations observed more often than expected (*P*adj<0.05). Yellow points indicate the number of times each combination was observed, red points indicate the number of times the yeast taxa were observed in the permuted dataset, and blue points indicate the expected value of each association.

**Figure 3) A)** Shannon-Wiener (H’) indices were used to examine diversity amongst substrate categories. Closed points indicate the value of the H’ index (left axis) and open points indicate how many times the substrate category appears in the dataset (right axis). Sampling density does affect the H’ index estimate, however, some substrate categories have similar H’ values with very different sampling densities.  **B)** H’ indices were used to examine diversity by temperature across the dataset. Isolations at 22°C are optimal for maximizing yeast diversity. 4°C isolations have dramatically reduced diversity.

**Figure S1**) Histogram of isolates in the complete dataset (top). When the 1,962 isolations are filtered to remove duplicate OTUs derived from the same processed sample, 1,518 unique isolation events remain (bottom)

**Figure S2**) Subphylum representation across unique isolations (light grey, top count) and unique OTUs (dark grey, bottom count). A lack of Pezizomycotina taxa is likely reflective of our isolation protocol in which the selection of hyphal colonies is actively avoided.

**Figure S3)** The distribution of isolations in the dataset by climate region. The number of unique isolations (upper, bold) and the number of unique taxonomic units (lower, italics) are shown. Climate regions correspond to the nine climatically consistent regions identified by NOAA(<https://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-regions.php>).

**Figure S4)** 116 singleton OTUs were isolated just once. A) Distribution of substrate categories amongst singletons. Singleton taxa not enriched for any substrate type. B) Singletons are enriched for OTUs belonging to the Basidiomycota and Pezizomycotina subphyla. C.) Isolation locations of singletons.

**Figure S5)** **A**) The number of independent isolations of each cosmopolitan OTU. **B**) The number of climatic regions each OTU was detected in.

**Figure S6)** Cosmopolitan taxa were defined as those OTUs that are isolated from most regions where they are expected based on sampling density (see methods). Eleven cosmopolitan OTUs were identified using this approach. A) Cosmopolitan yeasts are enriched for soil associations (*p*= 3.09e-05). B) Cosmopolitan yeasts all belong to Saccharomycotina or Basidiomycota, however there is no enrichment for subphylum among cosmopolitan taxa. C) Map of cosmopolitan isolation locations. Taxa are differentiated by color.

**Figure S7)** Forty discrete substrate categories were annotated for 1,522 isolations. A) Distribution of unique isolations amongst substrate categories. Categories sampling was extremely uneven. B) Substrate genera could be assigned to 1,026 isolations of diverse substrate categories. Substrate categories were either directly harvested from the substrate (e.g. pine needles from pine tree) or indirectly associated with the substrate (e.g. soil from the base of a pine tree).

**Figure S8)** Top) Observed data fed into substrate association permutations. Middle) Observed data fed into substrate genus association permutations. Bottom) Observed data fed into isolation temperature associations.

**Figure S9)** H’ indices broken down by temperature and subphylum correspond well with temperature enrichment analyses. Basidiomycete diversity is maximized at lower temperatures, Saccharomycotina diversity is maximized at higher temperatures. Pezizomycotina appears to have a narrower optimal isolation temperature window.