**Survey Paper Manuscript. List possible titles below.**

1. **Spatial patterns and substrate specificity across yeast taxa**

W. J. Spurley

wspurley\_at\_wisc.edu

Laboratory of Genetics

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

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

K. J. Fisher

Kaitlinfisher19\_at\_gmail.com

Laboratory of Genetics

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

C. T. Hittinger

Laboratory of Genetics

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

**Abstract**

To be written last.

**Introduction**

1. Budding yeasts are globally distributed, speciose, diverse, and have important relationships with humans. The habitat and ecology of yeasts have been much speculated on but remain poorly understood.
   1. Challenges to understanding the ecology of wild yeasts.
   2. Current approaches.
2. What do we currently know about the distribution and ecological roles of yeasts?
   1. Basidiomycota? Pezizos? Saccharomycotina?
3. In an effort to better understand the distribution and natural niches of yeasts, our lab has led an ongoing initiative to sample yeast diversity from natural substrates.
   1. What we have previously reported (Sylvester 2016, Opulente et al. 2019).
   2. How our collection is different now.
   3. What questions will we use this collection to ask in this paper?

**METHODS**

Often times, multiple isolates of the same OTU were isolated from the same exact substrate sample. To eliminate false positive associations due to this, prior to each analysis 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. SONE**).

***Sample Collection and Isolation***

Samples of natural substrates were collected across the continental United States as well as Alaska. They 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. The enrichment of these samples was performed as described in previously published literature (Sylvester et al., 2015) and was done so in order to increase the diversity of isolates. The enrichment included a sample inoculation in 9 mL of Wild Yeast Medium (list components?) in a conical tube that is then stored at 10C, 22C, and 30C. For a select number of samples, an inoculation was also stored at 4C The inoculations remained at their respective temperatures for 1-2 days or until there was evidence of fermentation. The tubes were then vortexed and 10 μL were added to 4 mL of fresh Wild Yeast Medium in a new tube. These tubes were then incubated at the same temperature of the original inoculation for another 1-2 days.

***Sample identification***

After there was evidence of yeast growth in the second enrichment, a 1:10000 serial dilution was performed, and that dilution was then plated onto YPD plates (list components) and stored at the corresponding isolation temperature until colonies were visible. Morphotypes were then identified and streaked out in order to obtain a single colony. From that single colony, a culture was created using YPD media (components). A small amount of culture was frozen down in glycerol for preservation while 200 μL was prepped for identification using a PCR of the ITS region. The PCR product was verified using gel electrophoresis and then sequenced using Sanger sequencing methods.

***Database creation***

Strains that were able to be identified were then included in a curated Wild Yeast Database. This database included strains identified from 20\_\_ to 20\_\_ that were collected by a variety of different collectors. ….

***Association analyses***

For each association analysis, the first step was to generate the necessary data frame for each condition. This was done by first removing OTUs that did not have sufficient metadata for the association analysis. We then removed OTUs with identical Set ID’s in order to eliminate any replicates in the data set. Once the data frame was prepared, it was permuted 10,000 times. From those permutations, an initial p-value was generated by taking the number of times the permuted value was greater than the observed value and dividing by the total number of permutations. The p-value was then adjusted using a Benjamini-Hochberg adjustment, and a threshold for significance of 0.05 was set.

***Fisher’s Exact Tests***

We performed Fisher’s exact t-tests for subphylum and isolation temperature, subphylum and substrate, as well as singleton and cosmopolitan OTUs with both substrate and subphylum. (How much do I need to explain Fisher’s exact?).

***Diversity Analyses***

In order to determine the diversity among isolation temperature and substrates, we generated a Shannon-Wiener index for each condition to control for sampling disparities. This was performed using the R package vegan. Data frames were trimmed in similarly to as described in the association analyses to only include unique isolates and remove OTUs without sufficient metadata. Once the indicies were determined, they were plotted and sorted by the level of diversity. The isolation temperature diversity index was further sorted by subphylum using the same methods. Since all of the substrates were collected from at varying frequencies, the sampling density was also included in that plot.

***Determining Cosmopolitan OTUs***

In order to determine which OTUs in our data set were cosmopolitan, we performed a geospatial analysis of the United States. We began by dividing the country into the ten climactic regions described by the National Oceanic and Atmospheric Association (NOAA). From there we determined how many times an OTU was isolated in each region. Using the region with the highest density of this OTU, we then determined the expected isolation frequency of that OTU by taking the total number of that isolate divided by the total number of isolations in that region. We were then able to determine if that OTU was expected to be isolated from another region if its isolation frequency was lower than the total number of isolations in that region. Using this expected value we then determined cosmopolitan OTUs by the criteria that it was isolated when expected more than one time and was not found when expected one or fewer times.

**Results & Discussion**

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

The final dataset curated for this study consisted of 1,962 isolates of 262 unique operational taxonomical 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, followed by144 in the Ohio Valley, 140 in the Northwest, 136 in the Northwest, 105 in the Southeast, 40 in the West, 32 in the South, and 6 in the Northern Rockies. Zero isolates were collected from the Southwest. An additional 44 isolations were sampled from Alaska.

***116 singleton isolations***

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**).

***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 species-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 (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 climactic regions (*Candida railensis, Candida sake, Cryptococcus flavescens, Cyberlindnera saturnus, Debaryomyces hansenii, Leucosporidium scotti, Rhodoturula fujisanensis, Saccharomyces paradoxus, Scheffersomyces ergatensis, Torulaspora delbrueckii,* and *Wikerhamomyces anomalus* (**Suppl. Table 3**). Some taxa identified as cosmopolitan were isolated frequently (e.g. *T. delbruckii* isolated 107 times) while other cosmopolitan taxa were isolated less frequently but nonetheless were broadly distributed across climactic regions (**Fig. S5, Fig. S6C**). Cosmopolitan species 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 certainly affected by sampling biases in our data. The approach also only considers total climactic 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 may 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 4**), 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 taxanomic 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 5**). Soil, which was the most heavily sampled substrate, was significantly associated with two Basidiomycota and three Saccharomycotina taxanomic 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 Basidomycota 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 comprised of 209 OTUs (**Fig.** **S8**, **Suppl. Table 6**). 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* - perrenial 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 *Saccahroymces* speciesand the oak genus *Quercus*. We isolated *S. cerevisiae* 21 independent times with only five isolations from *Quercus*-associated substrates (**Suppl. Table 7**). *S. paradoxus* was isolated 43 independent times with 7 isolations from *Quercus* substrates (**Suppl. Table 7**).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). 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. 2016). 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. paraxodoxus* can be found on a wide array of substrates and oak trees are unlikely to be the sole major reseviour 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. 2016). 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 confier pathogen and we find it statistically associated with the pine genus *Picea* as well as with needles (Talgo et al. Fungal biology 2010; Guertin et al. 2018). The opportunistic human pathogen *Pichia kudriazevii* 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* *delbreckii* and soil (Kurtzman, Fell and Boekhout 2011), (Talgo et al. Fungal biology 2010; Guertin et al., 2018), 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 (Suneda, 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 are novel. *Hanseniaspora uvuarm* has not been previously described as being 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, and 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 8**). 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 9**). 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 species – 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 taxanomic 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 10**). 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. For example, moss and sand are sampled at similar depth to nuts and wood, but have much higher diversity indicies. 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 innoculautions may also depend on laboratory processing 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 11**). 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**), 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 taxanomic 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***

Wrap up paragraph of our main conclusions.

**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’ indicies 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 taxanomic units (lower, italics) are shown. Climate regions correspond to the nine climactically 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 Peizozmycotina subphyla. C.) Isolation locations of singletons.

**Figure S5)** The isolation count for each cosmopolitan OTU.

**Figure S6)** Cosmopolitan species were defined as those species 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’ indicies 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 appear to have a narrower optimal isolation temperature window.