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

Draft authored by W. J. Spurley & K. J. Fisher

**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?

**Results & Discussion**

***Curated dataset of isolates – DANA NEEDED TO READ/AUGMENT ON HOW DATA WERE CURATED***

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. Isolates were obtained from all major climate regions of the contiguous United States with the exception of the Southwest, as well as Alaska (**Fig. S1**). 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**).

***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. Ma). Contingency tables were used to identify associations between yeast subphyla and substrate. When subphylum representation across substrate types was examined (**Suppl. Table 2**), 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. 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 (**Figure S3**).Of the 715 observed combinations (**Fig. S4**), we found 22 yeast OTU by substrate associations that occur more frequently than expected by chance (*P*adj<0.05, **Fig. 1**, **Suppl. Table 3**). 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. S5)**. 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.** **S4**, **Suppl. Table 4**). Nineteen of 657 observed substrate genera – yeast associations were found more often than expected by chance (**Fig. 2**). *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. S4**), 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 5**). *S. paradoxus* was isolated 43 independent times with 7 isolations from *Quercus* substrates (**Suppl. Table 5**).Further, an absence of associations cannot be explained by lack of sampling, as *Quercus* was the most deeply sampled genus in our dataset (**Fig. S5**). 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.

Many of the significant substrate associations we find here have been previously reported in the literature. 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. Additional previously described associations include *Torulaspora* *delbreckii*/soil (Kurtzman, Fell and Boekhout 2011), *Pichia kudrazevii/*fruit (Kurtzman, Fell and Boekhout 2011; Douglass et al. 2018; Opulente et al. 2019), *Sydowia polyspora*/*Picea* and needles (Talgo et al. Fungal biology 2010; Guertin et al., 2018), *Scheffersomyces ergatensis/*soil (Kurtzman, Fell and Boekhout 2011), and *Trichosporon* *porosum/*soil (Middlehoven et al., 2001).

***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=1.75E-35) and 10°C (*P*adj=1.00E-54, **Fig.** **S6**, Suppl. Table 5). Similarly, an enrichment for Saccharomycotina spp. is observed at 22°C (*P*adj=2.45E-23) and 30°C(*P*adj =8.90E-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 (Figure Fc), 16 significant positive associations were found (Figure I, Suppl. Table 6). 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. Ten of 13 30°C isolation– associated yeasts 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 yeasts contain one Pezizomycotina taxa (*Lecythophora* sp.; *P*adj<0.0001).

Among the two isolation temperature associations, OTU and subphylum, corroboration is evident. The subphylum analyses revealed that Basiodmycota yeasts were generally associated with cold temperature isolations and Saccharomycotina yeasts were generally associated with warmer isolation temperatures. This is reflected in the OTU associations with isolation temperature as well. Eight of the nine significant associations with cold temperatures were isolates in the subphylum Basidomycota. Conversely, Saccharomycotina yeasts were the most present OTUs among 30°C associations. Of those Saccharomycotina yeasts, we observed a significant association of *S. cerevisiae* at 30°C. This association has been observed on multiple instances and further strengthens our results (reference).

The results of the isolation temperature associations were expected to an extent. The range of isolation temperatures used in isolation protocols is broad and it would be difficult for an isolate to successfully grow and reproduce at each level. Because of this, we observed a clustering of associations of OTUs and more extreme isolation temperatures of 4°C, 10°C, and 30°C. These trends can be applied to isolation protocols. If one was to isolate a yeast from the subphylum Basidiomycota, it would be beneficial to incubate at a colder temperature than if the desired yeast was part of the subphylum Saccharomycotina.

***Substrate and temperature diversity***

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

Maximizing the isolate diversity 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 (Figure J\*, Suppl. Table 7). 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. A number of substrates with lower sampling density showed high relative diversity, including moss, sand, pinecones, twigs, and needles. Substrates exhibiting low levels of diversity include water, compost, insects, and needles/berries.

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

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. K\*). Because isolation temperature has a drastic effect on the subphyla of resultant isolates, we used the same approach to analyze diversity by 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. L). The isolation protocol used in this analysis was created with the intention of avoiding non-yeast isolates like those in the subphylum Pezizomycotina which is reflected in a lower H’ index for all isolation temperatures among Pezizomycotina.

Again – need to address some specifics about the findings. Discuss that the H’ index is useful when comparing similarly sampled substrates. Are there any stark differences there? Is there a reason to expect higher or lower diversity (maybe check out the work Chris mentions – should we expect lower diversity in insects?) The sentence below is good, but the idea can be refined and built on. The maximum diversity for each subphylum is 22°C but there are no associations with this isolation temperature at the isolate level. This shows that an isolation protocol seeking to isolate the highest number of OTUs would be best suited to incubate at 22°C whereas an isolation protocol seeking to identify a specific yeast would want to incubate at a more extreme temperature.

***Geospatial variation among isolations***

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

Majority of yeast isolates in this study were collected from the continental United States. 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 (Figure A). 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, while the remaining 146 generally have fewer than thirty isolations (Figure C, Suppl. Table 8). The singletons isolated in this study were primarily isolated from bark and soil, however singletons were not enriched for any substrate category (Figure Da). Singletons were, however, enriched for yeasts belonging to the Basidiomycota (*P*adj=0.003) and Pezizomycotina (*P*adj= 0.0005) subphyla (Fig. Db). Singletons appear evenly distributed across sampled regions (Fig. Dc).

***11 cosmopolitan isolates***

Our cosmopolitan analysis determined that eleven OTUs in the dataset are found when expected all but one time while being in more than two regions (*Candida railensis, Candida sake, Cryptococcus flavescens, Cyberlindnera saturnus, Debaryomyces hansenii, Leucosporidium scotti, Rhodoturula fujisanensis, Saccharomyces paradoxus, Scheffersomyces ergatensis, Torulaspora delbrueckii,* and *Wikerhamomyces anomalus,* Fig. E*).* Of these OTUs, four are also associated with a specific substrate that exhibits high diversity (*T. delbrueckii, S. paradoxus,* and *C. saturnus* with soil; *S. ergatensis* with bark). Two cosmopolitan OTUs are found to be associated with 10°C (*C. sake* and *R. fujisanensis)*, and *D. hansenii* is associated with the coniferous genus Thuja. The isolation counts for the OTUs showed varying results with *T. delbruckii* being isolated 107 times and three OTUs being isolated fewer than ten times (Fig. O).

This regional analysis was able to identify cosmopolitan OTUs even with an unevenly sampled United States. Even in the regions that were less densely sampled, each OTU was able to be isolated where it was expected. Since there is a substantial amount of variation among these regions, we can conclude that the eleven OTUs that were found to be cosmopolitan exhibit sustainable levels of fitness in a wide variety of conditions. The associations found with these cosmopolitan OTUs demonstrate that while these OTUs are able to survive in multiple climates, they may still have ideal substates or temperatures. It is probable that there are more cosmopolitan OTUs in the United States and further sampling of the sparce regions may help to reveal them.

***Little evidence of endemism***

We examined our dataset for yeast taxa that demonstrate range restriction. We looked for any yeast taxa that were isolated over 5 independent times from a single geographic region and never isolated in any other region. The eight OTUs that met these criteria were all isolated from the most densely sampled region, the Upper Midwest. This strongly suggests that any signal of range restriction in these data is likely a product of sampling bias. …Most of the yeasts have also been described on other continents… need a table or figure or something if we’re keeping this …

**Figure legends**

**Figure 1)** Permutations identified 22 yeast taxa – substrate category associations observed more often than expected (*P*adj<0.05). Blue points indicate the number of times each combination was observed, yellow points indicate the number of times the yeast taxa was observed in the permuted dataset, orange points indicate the expected values of each association, and green points indicate the number of times the substrate category was observed in the permuted dataset.

**Figure 2)** Permutations identified nineteen yeast taxa – substrate genus associations observed more often than expected (*P*adj<0.05). Blue points indicate the number of times each combination was observed, yellow points indicate the number of times the yeast taxa was observed in the permuted dataset, orange 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 S1)** 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 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**) 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 S4)** 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 S5)** 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).

**Fig. S6)** Distribution of isolations for the 1,750 isolations for which isolation temperature is known. Colors correspond to the subphylum of the corresponding isolate.

**Figure 1**



**Figure 2**

****

**Figure D**



Figure D. Distribution of singletons among substrate, subphyla, and geographic location in the United States.

**Fig. D)** 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 E**

****

**Fig. E)** 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 I**



**Fig. I)** Permutations identified 16 yeast taxa – isolation temperature associations observed more often than expected (*P*adj<0.05). Blue points indicate the number of times each combination was observed, yellow points indicate the number of times the yeast taxa were observed in the permuted dataset, and red points indicate the expected value of each association.

**Figure J**

****

**Figure J)** 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.

**Figure K**



**Fig. K)** 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 L**



**Fig. L)** 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.

**Figure N**

**Figure O**

****

**Fig. O)** The isolation count for each cosmopolitan OTU.